International Journal of Undergraduate Research and Creative Activities

Volume 4
Issue 2 *Res Cogitans 2012*

Article 1

October 2023

Counterfactuals and the Problem of Old Evidence

Philip Yaure
info@ubiquitypress.com

Counterfactuals and the Problem of Old Evidence

Abstract

In this paper, I consider the Problem of Old Evidence, which is meant to undermine the theory of confirmation Bayesianism uses to explain the role of evidence in science. The problem maintains that the Bayesian definition of evidence cannot include facts known before a theory is introduced (but whose relation to the theory is unknown at the moment of introduction). I argue that this problem can be diffused by the introduction of counterfactuals, which specify conceivable scenarios in which the fact is discovered after the theory is introduced. I consider several sorts of objections to this view, and contend that we have good reason to reject them in their own right, and that the other alternative solution in the literature does not offer a sufficient solution to the problem, further compelling us to face the objections, if we are to maintain a Bayesian confirmation theory.

Res Cogitans (2012) 3:2-8 2155-4838 | commons.pacificu.edu/rescogitans

Counterfactuals and the Problem of Old Evidence

Philip Yaure
University of Chicago

Published online: 07 June 2012

Abstract

In this paper, I consider the Problem of Old Evidence, which is meant to undermine the theory of confirmation Bayesianism uses to explain the role of evidence in science. The problem maintains that the Bayesian definition of evidence cannot include facts known before a theory is introduced (but whose relation to the theory is unknown at the moment of introduction). I argue that this problem can be diffused by the introduction of counterfactuals, which specify conceivable scenarios in which the fact is discovered after the theory is introduced. I consider several sorts of objections to this view, and contend that we have good reason to reject them in their own right, and that the other alternative solution in the literature does not offer a sufficient solution to the problem, further compelling us to face the objections, if we are to maintain a Bayesian confirmation theory.

Broadly speaking, Bayesianism is a position in the philosophy of science that attempts to explain the nature of scientific belief (that is, belief which depends primarily on the evidence available) in terms of probability functions. The notation of these probability functions is such: Pr(X|Y), where 'X|Y' is read as X given (or in virtue of) Y. The function will produce a value between 0 and 1, where 0 means that there is no confidence in the proposition X (in virtue of Y), and 1 means that there is complete confidence in the proposition X (in virtue of Y). Values between 0 and 1 reflect lesser and greater degrees of belief in X.

Bayesianism gets its name from Bayes Law, which is used to define the probability functions Bayesians apply to scientific beliefs. The law takes the following form:

$$\Pr(T|E\&B) = \frac{\Pr(E|T\&B)}{\Pr(E|B)} \Pr(T|B) \quad (1)$$

The left term is the posterior probability—the degree of belief one has in a theory T in virtue of some (new) event E and the background knowledge B. This value can be calculated by multiplying the prior probability, Pr(T|B) by the ratio of the 'likelihood,' Pr(E|T&B), and the 'expectedness,' Pr(E|B).

One of the merits of Bayesianism is that it seems to provide an explication of evidence—that is, it tells us when an event is to be taken as a confirmation of a theory given some background knowledge. Confirmation in a Bayesian framework is generally sketched out as such: E counts as evidence for T iff

$$Pr(T|E\&B) > Pr(T|B) \quad (2)$$

That is, the probability that T is true (which will be interpreted in this paper as the degree of belief an individual has in T) is greater in virtue of knowledge of E than in virtue of its absence. This is meant to hold for all instances of evidence.

However, Clark Glymour, in his paper "Why I am not a Bayesian," presents an issue for this Bayesian confirmation theory, the Problem of Old Evidence. He observes that in instances in which the event is known before the theory is introduced, Bayesianism cannot accommodate the event as evidence for the theory, because the posterior probability, Pr(T|E&B), cannot be greater than the prior probability, Pr(T|B). After noting a few historical examples in which old evidence apparently ought to confirm a theory, such as the advance of the perihelion of Mercury for Einstein's theory of special relativity (Glymour, 262), it is clear that this is a counter-intuitive consequence of Bayesianism.

Thus, the problem of old evidence presents a serious issue to Bayesianism, insofar as it is meant to serve as a theory of confirmation. In this paper, I will primarily consider one line of defense to the problem, the introduction of counterfactuals. I will argue that two lines of criticism against counterfactuals, the practical and the formal, ought to be rejected, and that a third, the interpretive, should not be seen as deeply troubling, given the rejection of the practical and the formal. I will conclude by briefly considering an alternative response to the problem, the restriction on logical omniscience, arguing that this response undercuts Bayesian confirmation theory by pointing to a more fundamental logical structure governing our degrees of belief.

First, it is important to provide a clear sketch of the problem itself. Often, scientists seem to construct theories that explain phenomena that predate the theories. Glymour notes several examples: Copernicus argued for his heliocentric theory of the solar system by citing data across millennia, Newton argued for his theory of universal gravitation via Kepler's second and third laws, and Einstein's special theory of relativity explained the previously observed advance of Mercury's perihelion (Glymour, 262). Eells further draws out this general phenomenon, decomposing it into instances in which the theory was originally designed to explain the event, and instances in which the theory was not originally designed to explain the event (207). In the latter case in particular, because the theory was not 'tailor-made' to explain the event, we would want to say that the theory's ability to explain or predict the event would constitute the event as evidence for the theory.

However, given the Bayesian definition of confirmation (2), it is unclear that the event counts as evidence for the theory. Take Bayes' theorem at the moment of the theory's introduction:

$$\Pr(T|E\&B) = \Pr(T|B)\frac{\Pr(E|T\&B)}{\Pr(E|B)} \quad (3)$$

The problem is with the ratio. The bottom term, the expectedness, Pr(E|B), is the degree of belief the agent has of E independent of T. Because E has already been observed, this is 1. And the top term, the likelihood, is the degree of belief the agent has in E given that T is true. Since the agent's degree of belief in E is 1 independent of T's truth, this is also 1.Thus, the ratio as a whole is 1, and so Pr(T|E&B) = Pr(T|B), and so E is not, by (2), a confirmation of T.[1,2] Either our intuitions about evidence are radically off base or it appears that Bayesian confirmation theory is severely lacking. And given that confirmation theory is meant to explain a concept that we already have intuitively, the balance tips towards the latter disjunct.

One of the standard responses to the problem is to introduce counterfactuals scenarios in which the expectedness is not 1. The intuition of this response is that the temporality of the situation that the problem hinges on is inessential, and so massaging it might provide a solution. The goal is to manufacture a counterfactual set of background knowledge, B', from which E and any related knowledge has been removed. Because E (and any related information) is not contained within B', Pr(E|B') << 1, enabling our posterior probability to be greater than our prior probability (which constitutes E as evidence for T). This method is usually taken to translate into imagining a situation in which we do not know E (or any related information) when T is introduced, and subsequently discover E. E can thus count as evidence for T in the actual world because it would have fit into the Bayesian definition of evidence if it hadn't been discovered before T was introduced.

There are a number of objections to this response. We can group them into three categories: the practical, the formal, and the interpretive. I will focus below on the former two, under the assumption that if these can be overcome, we will have enough ground to justify the stretched interpretation that counterfactuals may require.

Practical objections doubt whether it would actually be possible to calculate Pr(E|B'). Glymour observes: "It is not obvious that there are, for each of us, degrees of belief we personally would have had in some historical period" (265).[3] There seem to be two sorts of practical concerns: 1. It is unclear that we could manufacture a B', given the intricate nature of our actual background knowledge; and 2. It is unclear that we could calculate Pr(E|B') in virtue of the fact that it is a counterfactual belief.

Proposition 2 has some intuitive force, but I believe that proposition 1 is ultimately doing the work for 2. We may want to say it is hard for me to conceive of degrees of beliefs other than those I actually have, but it is clear that I can, in fact, consider at least some counterfactuals degrees of belief. For example, if I were asked to imagine that there would be a football game tomorrow night between the Eagles and the Cowboys, I could certainly offer personal betting ratios (which correspond to degrees of belief) about who would win this counterfactual game. And probability textbooks often construct counterfactual examples to display concepts (see Hacking, 177-79 for an example). The mere introduction of counterfactuals doesn't seem to prevent calculations.

Rather, when we want to say that we can't calculate a counterfactual degree of belief, it seems that the problem lies in our inability to fully flesh out the situation; that is, in this context, to full conceive of B'. That is to say, most instances of proposition 2 are grounded in supposed instances of proposition 1. It certainly seems challenging to remove E, and, in particular, to determine all the related facts which must be removed from our body of knowledge to properly manufacture B'. And even if we were able to do so, as Glymour notes, it seems that there would be a large set of possible counterfactuals to select from, and it is unclear that we could select from them on non-arbitrary grounds (265).

We may ask, however, if the practical objection holds the counterfactual response to too high of a standard. The objection requires that we manufacture a 'unique' (non-arbitrary) counterfactual for which a definite value for Pr(E|B') can be calculated. Certainly, the non-arbitrary aspect of this requirement is too strong. While the B' that we manufacture must have some constraints on it (this will be considered below), it is unclear that we need constraints that permit only one B'. The counterfactual response is simply meant to show that there is at least one B', similar enough to a one's actual epistemic situation, such that Pr(T|E&B') > Pr(T|B')—having multiple B' that satisfy this certainly doesn't undermine the response.[4]

It is also unclear whether we need to actually need to perform the calculation for an arbitrary B' to show that the counterfactual response holds. Certainly, one would be hard-pressed to deny that a B', which is similar to an agent's actual epistemic situation and provides a Pr(E|B') sufficiently less than 1, is possible. Proposition 1 essentially maintains that B' is inconceivable: since we are unable to manufacture a fine-grained B', we are unable to produce a definite calculation. But it is unclear that we actually need conceivability in this 'fine-grained' sense. Rather, we can invoke a weaker sense of conceivability to establish the possibility of the desired B'. For example, regardless of whether we could etch out a body of knowledge available to Einstein when he posited special relativity in which we have specified precisely which facts related to the perihelion of Mercury have been eliminated, we can certainly grasp that such a situation is possible (for whatever reason, nobody bothered to look closely at Mercury's orbit, perhaps). In virtue of this weaker sense of conceivability, we have a counterfactual situation in which Pr(T|E&B') > Pr(T|B').[5]

Nevertheless, one may attempt to push around this claim through a formal objection, noting that while the counterfactual response may not need to offer non-arbitrary, definite counterfactual degrees of belief, it is unclear that any counterfactual degrees of belief can be defined, for it is unclear that any counterfactual expectedness can be evaluated. Glymour observes (264) that a typical method of manufacturing counterfactual degrees of belief calculates counterfactual expectedness as such:

$$Pr(T_1)Pr(E|T_1)+Pr(T_2)Pr(E|T_2)+\ldots+Pr(T_k)Pr(E|T_k)+Pr(\sim(T_1 \text{ v}\ldots\text{v } T_k))Pr(E|\sim(T_1 \text{ v}\ldots\text{v}T_k))$$

The problems is with the last term, the 'catch-all.' This term is meant to reflect the possibility that all of the existing theories are false—in other words, that the true theory is not contained within the set of considered theories. It is unclear how we could produce any value for this term. In particular, the term asks us to calculate the likelihood of E given that all of the theories we are considering (all of the current possible theories of science) are false. It seems that we cannot provide any value for such a probability function—a fact which is meant to be deeply troubling, for if this term is not defined, Pr(E|B') is also undefined, and so the above defense to the practical objection is irrelevant, as it rests on the assumption that there is at least some B' for which the expectedness is defined. While this probably applies generally to Bayesianism, it is a severe danger for counterfactuals in particular: in actual cases, the Bayesian might be able to wave her hands by saying that this term is so small that it can be neglected or that it can be estimated by a scientists raw natural intuitions about phenomena (untainted by theory) and lack of confidence in the currently live theories in the scientific community. These responses are implausible in actual cases, but are completely useless in the weak-conceivability scenarios I have argued for above, as we have no concrete values or intuitions to work from.

It seems, however, that we may be able to draw a solution from Salmon's paper, "Rationality and Objectivity in Science." Salmon offers a solution generally to this formal concern, regardless of whether we are dealing which an actual expectedness or counterfactual (188). He observes that we can eliminate the expectedness from calculation by evaluating the ratios of particular theories (192):

https://digitalcommons.cwu.edu/ijurca/vol4/iss2/1
DOI: -

$$\frac{\Pr(T_1|E\&B')}{\Pr(T_2|E\&B')} = \frac{\Pr(T_1|B')\Pr(E|T_1\&B')}{\Pr(T_2|B')\Pr(E|T_2\&B')} \quad (4)$$

This equation contains no mention of the expectedness, and thus avoids the formal concern outlined above. Instead of calculating raw degrees of belief for a theory, we can compare a theory in question to another, say, the commonly accepted theory of the time, and formulate the Bayesian definition of confirmation as such: E counts as evidence for T iff

$$\frac{\Pr(T_1|E\&B')}{\Pr(T_2|E\&B')} > \frac{\Pr(T_1|B')}{\Pr(T_2|B')} \quad (5)$$

We can ground such a move in a Kuhnian intuition, that we never assess the merits of a theory in a vacuum, but only in relation to other theories live in the scientific community (Salmon, 191). E is thus evidence for T_1 if it improves its standing in the agent's mind relative to T_2. There seems to be no reason to prevent this from applying to counterfactual cases—certainly if the advance of Mercury's perihelion was discovered after special relativity was introduced, the ratio between special relativity and the previous paradigmatic theory would tip in the former's favor.

Nevertheless, there are still serious concerns with the counterfactual method. In particular, it seems that we may be letting too much in by allowing the problem old of evidence to be resolved by an argument from the possibility of B', as outlined above (especially if our intuitions about instances of evidence are to constitute this possibility).

I believe that the answer to these interpretive concerns rest on two elements. First, as noted above, it is not the possibility of any old B', but one sufficiently similar to one's actual epistemic situation that matters. A proper sufficiency condition of this sort should restrict what sort of instances can be let in. While a complete condition is unfortunately beyond the scope of this paper, it does not seem to me that such a condition is indefinable, in principle. We might say that the condition should only alter the status of T and closely related theories (so that my beliefs in evolution are not altered the absence of evidence related to the perihelion of Mercury), where 'closely related' can be defined in terms of a specialized scientific discipline (such as, say, physics, astronomy, or studies in planetary orbits; in the case of special relativity, given its historic influence on physics as a whole, a wide net should be cast). Ultimately, it is the burden of the proponent of the problem to show that such a sufficiency condition cannot, in principle, be established, given the sketch provided here.

Of course, one may be concerned enough with these interpretive concerns to lean towards another alternative in the literature.[6] This alternative can only be glossed in conclusion. It is generally labeled as the placing of a restriction on logical omniscience, the Bayesian idealization that the agent is aware of all logical relations between all propositions at all times. The proponent of this solution argues that there is something genuinely discovered in cases of old evidence: the relation between E and T, that T entails E. Posterior probabilities are to be revised in virtue of this discovery.

However, the introduction of logical relations seems to undermine the very motivation for a Bayesian confirmation theory. One of the essential merits of applying Bayesianism to confirmation theory is that it avoids any explicit reference to logical relations: E's evidentiary nature is constituted simply in terms of its influence on our degree of belief in T. To introduce logical relations into this scheme is to point beyond degrees of belief to "a certain logical or structural connection" which renders Bayesian degrees of belief "at best, epiphenomental" (Glymour, 266). Ultimately, it seems that on the restricted logical omniscience picture, it is no longer Bayesian confirmation theory doing the real explanatory work, but rather the logical relations. If we are to maintain that degrees of belief themselves should serve as the explanation of evidentiary relations, then, it seems, we ought to accept the counterfactual response and push towards solutions to its interpretive concerns.[7]

Works Cited

Eells, Ellery. "Bayesian Problems of Old Evidence." *Scientific Theories*. Ed. C. Wade Savage. University of Minnesota, 1990. 205-23. Print.

Garber, Daniel. "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory." *Testing Scientific Theories*. Ed. John S. Earman. University of Minnesota, 1983. 99-131. Print.

Glymour, Clark. "Why I Am Not a Bayesian." *Philosophy of Probability*. Ed. Antony Eagle. Oxon: Routledge, 2011. 250-67. Print.

Hacking, Ian. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge UP, 2008. Print.

Salmon, Wesley C. "Rationality and Objectivity in Science." 1990. *Scientific Theories*. Ed. C. Wade Savage. University of Minnesota. 175-204. Print.

[1] Another way to see this is that E, because it is already known, is in the set of our background knowledge, B. Thus, 'E&B' is redundant, and so Pr(T|E&B) = Pr(T|B).

[2] One may object that we never actually have a degree of belief 1 in any proposition. Glymour notes that (263), even granting this, Bayesianism does not give us the sort of picture we want about old evidence. For if the likelihood and the expectedness are both nearly 1 (as they would be in the case of skepticism about genuine observations), and the likelihood is slightly greater than the expectedness, we would get a posterior probability only slightly greater than the prior. This clearly does not describe the sorts of historical examples Glymour cites, as the advance of Mercury's perihelion was taken as strong evidence for special relativity, such that Pr(T|E&B) >> Pr(T|B).

[3] Glymour goes on further to note that if we require that the agent operate on degrees of belief of the scientific community at large: "we should have to condemn a great mass of scientific judgments on the grounds that those making them had not studied the history of science" (265). This concern, about whether we work with personal or community-wide degrees of belief, is unfortunately beyond the scope of this paper. Certainly some version of the problem can be fleshed out solely in terms of an individual who knows an event and then discovers that a new theory explains it, and this is the sort of problem I will attempt to use counterfactuals to address.

[4] Nor does having some B' that fails to satisfy (1). We can simply classify any such B' as too similar to the actual situation—we simply didn't remove all of the knowledge related to E from our epistemic situation.

[5] Of course, one may be concerned that we are sacrificing one of the essential merits of Bayesianism, the calculation of precise degrees of belief, for another, its version of confirmation theory. I believe that there is no clear way to resolve this concern, and so it must simply be acknowledged. Ultimately, it seems to me that we

should prize Bayesianism's confirmation theory over its precise degrees of belief, because of the philosophical significance of the former; but we may also note that the Problem of Old Evidence need not be resolved by establishing precise instances in which the posterior probability is greater than the prior — the satisfaction of (2) does not require precise values. The problem of old evidence, in the end, seems simply to require that (2) be established for a set of instances in which it seems that it cannot be established, and the argument from the possibility of B', which entails a Pr(E|B') sufficiently less than 1, establishes this fact.

[6] Garber (1983) is one proponent of this alternative.

[7] My thanks to Dawn Chow for very helpful comments on a previous draft of this paper, and to Professor Kevin Davey for much helpful and patient discussion on the topic.