The Bayesian and the Abductivist

Mattias Skipper · Olav Benjamin Vassend

October 19, 2024

Abstract

A major open question in the borderlands between epistemology and philosophy of science concerns whether Bayesian updating and abductive inference are compatible. Some philosophers—most influentially Bas van Fraassen—have argued that they are not. Others have disagreed, arguing that abduction, properly understood, is indeed compatible with Bayesianism. Here we present two formal results that allow us to tackle this question from a new angle. We start by formulating what we take to be a minimal version of the claim that abduction is a rational pattern of reasoning. We then show that this minimal abductivist principle, when combined with Bayesian updating by conditionalization, places surprisingly strong and controversial constraints on how we must measure explanatory power. The lesson is not that Bayesianism is definitely incompatible with abduction, but that both compatibilism and incompatibilism have hitherto unrecognized consequences. We end the paper by formulating these consequences in the form of a trilemma.

1 Introduction

We often judge the credibility of a hypothesis based on considerations about how well or badly the hypothesis, if true, would explain the available evidence. Sometimes these considerations are highly sophisticated, such as when theoretical physicists debate the explanatory virtues or vices of different quantum mechanical theories. Other times our reliance on explanatory considerations is more mundane, such as in the following scenario:

> **Knocking on Evan's Door**: Whilst preparing dessert for tonight's dinner party, Brad discovers that he is out of sugar. Determined not to disappoint

> his guests, Brad rushes over to his neighbor, Evan, who is usually stocked up on dry goods. "Knock, knock." No one is answering. "Darn it," Brad mutters to himself, "it looks like Evan isn't home." Disappointed he walks away, empty-handed.

It is easy to imagine in this scenario that Brad takes the fact that no one is answering the door as evidence for the hypothesis that Evan isn't home, because the hypothesis that Evan isn't home, if true, would provide a good explanation for why no one is answering the door. There are, of course, other possible explanations: Evan could be napping, or he could be wearing noise-cancelling headphones, or he might simply not feel like answering the door. But we can imagine that Brad doesn't take these alternative explanations to be as good, and hence doesn't take them to be as credible.

The kind of reasoning in which Evan is engaged here is often referred to as "abductive" reasoning. It is also common to use the term "Inference to the Best Explanation" (IBE), famously coined by Harman (1965). However, we do not want to presume, as some articulations of IBE do, that the reasoning process in question results in anything like a firm conclusion to the effect that this-or-that hypothesis is true. The phenomenon we want to concentrate on here may simply involve apportioning one's degrees of belief in a way that reflects how well or badly a given hypothesis, if true, would explain one's evidence. It is then a further question under which circumstances, if any, we are licensed to outright infer the truth of a putative explanation. So to avoid any potentially misleading connotations, we will stick to the term "abduction" throughout.

Abduction is widely regarded as a rational pattern of reasoning. Of course, abduction can go wrong in any number of ways: a desperate parent might concoct a far-fetched explanation to convince himself that his child didn't commit the crime; a politically motivated conspiracy theorist might pay undue attention to certain subtle patterns that an unbiased person would assign little probative value; and so on. But when done right, abduction is usually seen as a rational way of apportioning one's beliefs to the evidence.

If abduction is indeed rational when done right, we should expect a complete theory of how to rationally revise one's beliefs in light of new evidence to make adequate room for abductive reasoning. However, the perhaps most prominent candidate for such a theory, Bayesian updating by conditionalization, does not appear to say anything to the effect that one should be more confident in better explanations. What it says is that one should accommodate new evidence

by conditionalizing one's prior degrees of belief (or "credences") on that evidence. More precisely, if "$P(H|E)$" denotes one's prior conditional credence in a hypothesis H given evidence E, and "$P_E(H)$" denotes one's posterior unconditional credence in H upon having updated on E:

Conditionalization: $P_E(H) = P(H|E)$

As many writers have observed, this updating rule does not—at least not *prima facie*—say that one's posterior credence in H should in any way depend on how well H explains one's evidence. Yet, if abduction is rational, one's posterior credence in H *should* depend on how well H explains one's evidence.

What to make of this apparent tension? If we look at the existing literature on this question, we can isolate two broad classes of answers. On the one hand, there are those who maintain that abduction is indeed incompatible with Bayesianism. Bas van Fraassen (1989) is a classic example. He influentially argued that abduction effectively amounts to giving "bonus points" (literally: extra credence) to good explanations beyond what is licensed by conditionalization. He took this to show that abduction is ultimately irrational, since any deviation from conditionalization leaves one vulnerable to a Diachronic Dutch Book (cf. Teller (1973)).

Many subsequent writers have agreed with van Fraassen that Bayesianism is incompatible with abduction, but they have typically drawn different conclusions from this incompatibility. For example, Douven (2022) has argued that abductive reasoning enjoys certain advantages over conditionalization which can, under the right circumstances, outweigh the advantages that conditionalization enjoys over abduction. Several others have suggested that, although there may be an ideal sense in which one should always comply with the dictates of conditionalization, abduction can nonetheless serve as a useful heuristic to approximate Bayesian reasoning for cognitively limited agents like ourselves (Okasha, 2000; Lipton, 2004; Dellsén, 2018).[1]

On the other hand, there are those who maintain that abduction, properly understood, is fully compatible with Bayesianism. For example, Weisberg (2009)

[1]The heuristic conception of abduction is sometimes presented as a version of compatibilism, because it assigns a substantive role to both conditionalization and abduction (see, e.g., Dellsén (2018)). For the same reason, one may think of Douven's view as a form of compatibilism, because it makes room for both conditionalization and abduction. However, for present purposes, we find it more useful to place both views in the incompatibilist camp, because they both concede that abduction involves at least a slight deviation from conditionalization.

has argued that explanatory considerations play a vital role in determining the objectively correct prior probabilities. On this view, abduction may be said to impose "external" constraints on the Bayesian framework. An alternative view has been proposed by Henderson (2014), who argues that, even if Bayesian agents are left to pick their prior probabilities without any explicit guidance by explanatory considerations, they will nonetheless end up assigning higher credence to explanatorily powerful hypotheses. On this view, abduction may be said to "emerge" from the Bayesian framework itself.

Our aim here is not to address these existing proposals head-on.[2] Instead, we want to present two formal results that allow us to tackle the debate between compatibilists and incompatibilists from a new angle. We will begin by formulating what we take to be a minimal version of the claim that abduction is a rational pattern of reasoning (§2). We will then show that this minimal abductivist principle, when combined with Bayesian conditionalization, places surprisingly strong and controversial constraints on how we must measure explanatory power (§§3-4). The lesson is not that Bayesianism is definitely incompatible with abduction, but rather that both compatibilism and incompatibilism have hitherto unrecognized consequences. We end by formulating these consequences in the form of a trilemma (§5).

2 Minimal Abductivism

The first task is to get more precise about what is minimally involved in saying that abduction is a rational pattern of reasoning. The qualifier "minimally" is important here, because we want to give compatibilism the best possible chances of succeeding. If it proves difficult to reconcile Bayesianism and abduction even under these favorable circumstances, so much the worse for compatibilism.

The idea that abduction is a rational pattern of reasoning is sometimes captured by saying that *explanatory power is a mark of truth*, or that *explanation is a guide to confirmation*, or simply that *good explanations are more credible than bad ones*. While these slogans are on the right track, some clarifications are required.

First, when advocates of abduction say about a hypothesis that it constitutes a good explanation, they do not simply mean that the hypothesis is a *likely*

[2] Interested readers may consult Dellsén (2024), who offers an extensive overview and discussion of various philosophical issues concerning abduction, including its relation to Bayesian inference.

explanation, in the sense of having a high probability of being correct. Otherwise there would be no substantive question as to whether good explanations are more credible than bad ones (trivially, likely explanations are more credible than unlikely ones). The term "explanatory goodness" is rather supposed to track certain properties of a hypothesis that can be assessed independently of its likeliness. What those properties are, exactly, is itself a substantive question, which we need not take a stance on here, but some commonly invoked explanatory virtues include simplicity, informativeness, non-ad hocness, and the ability to unify diverse phenomena.[3]

Second, in saying that a hypothesis constitutes a good explanation, advocates of abduction do not intend to presume that the hypothesis does in fact explain why the evidence obtains. The hypothesis may turn out not even to be true, and falsehoods arguably explain nothing. It is well known, of course, that scientists often rely on idealized models which are known to be at most approximately true. Nonetheless, there is a distinction to be made between being a good *putative* explanation and being a *genuine* explanation, and the term "explanatory goodness" is here used to track the former notion: how well a hypothesis would, *if true*, explain the evidence.

Third, we don't want to presume on behalf of abductivists that the credibility of a hypothesis depends solely on its explanatory power. Other kinds of considerations might also factor into an overall assessment of the hypothesis' plausibility. As Marc Lange (2022) puts it, abduction "permits explanatory considerations to be overridden by other considerations so that the 'best explanation' of one fact need not be the most plausible hypothesis all things considered" (Lange 2022, p. 87). One way this can happen is if two hypotheses explain a body of evidence equally well, but one hypothesis is more plausible than the other prior to receiving the evidence. Here is a toy example:

> **Squares N' Primes**: Sheena is about to roll a fair six-sided die. She fancies both square numbers and prime numbers, so you know that she will be happy if she rolls a square (1 or 4) or a prime (2, 3, or 5), and that she will be sad otherwise. After having rolled the die, you see a smile grow on her face, but you don't see what she rolled.

[3]See Thagard (1978) and McMullin (1983) for classic discussions of explanatory virtues and their role in theory choice, and see Glymour (2015) and Cabrera (2017) for critical discussions of how explanatory virtues relate to Bayesianism.

Let H_s be the hypothesis that Sheena rolled a square, and let H_p be the hypothesis that she rolled a prime. Given that all you learn is that Sheena is happy about what she rolled, H_p and H_s presumably explain the evidence equally well (again, this is not to say that H_p and H_s are equally *probable* explanations, but rather that H_p, *if true*, would explain the evidence no better or worse than H_s). Yet, you should end up being more confident in H_p than H_s, since there are three ways of rolling a prime and only two ways of rolling a square, which means that the prior probability of H_p is greater than that of H_s. Examples like this suggest that the credibility of a hypothesis depends not only on its explanatory power, but also its prior probability. As we will see below, there may be other factors as well that help to determine a hypothesis' overall credibility. For now, the important point is just that, to avoid attributing too much to the abductivist, all we assume is that explanatory power is *a* guide—not necessarily the only guide—to confirmation.

Fourth, in saying, as we have just done, that the credibility of a hypothesis may depend on its prior probability in addition to its explanatory power, we are assuming that a hypothesis' explanatory power does not itself depend on its prior probability. This may seem doubtful given that we often seem to evaluate the goodness of an explanation based in part on its prior probability. For example, we might sensibly say in the Squares N' Primes scenario that H_p is a better explanation than H_s, precisely because H_p has a higher prior probability than H_s and therefore is more likely to be the correct explanation of Sheena's facial expression. However, although this is a perfectly sensible way of talking, it invokes a different notion of explanatory goodness than the one which is operative in the present context. As mentioned, when abductivists say about a hypothesis that it constitutes a good explanation, they are not simply saying that the hypothesis is likely to be the correct explanation. Rather, the operative notion of explanatory goodness is supposed to capture how well a given hypothesis, *if true*, would explain the evidence.[4] Given this way of understanding explanatory goodness, we take it to be plausible that H_p and H_s are equally good explanations of Sheena's facial expression, and, more generally, that a hypothesis' explanatory power does not depend on its prior probability.[5]

[4]Glass (forthcoming, 2023), inspired by remarks in Good (1968), refers to this counterfactual notion of explanatory power as "weak explanatory power" in contrast to "strong explanatory power" which incorporates a hypothesis' prior probability. On this terminology, we can say about the Squares N' Primes scenario that H_p and H_s have the same weak explanatory power, but that H_p has a higher strong explanatory power than H_s due to its higher prior probability.

[5]There is a subtle qualification to be added here, which we defer until section 3 when it

Fifth and finally, we take it to be uncontroversial that explanatory power comes in degrees: explanations are not just good or bad, but better or worse. Accordingly, whatever else it might mean to say that explanation is a guide to confirmation, it must at least mean that, other things being equal, the better a hypothesis explains the available evidence, the more credible the hypothesis is. In other words, increasing the explanatory power of a hypothesis must, other things being equal, have the effect of making the hypothesis more credible. We are not hereby saying anything about *by how much* a hypothesis' credibility should increase as a result of increasing its explanatory power by a certain amount. This is something abductivists might disagree about. All we are attributing to the abductivist is the minimal claim that if we increase the explanatory power of a hypothesis while holding everything else fixed, this should have a positive impact, however small, on the hypothesis' credibility.[6]

To state the central results of the paper, we will need to formulate this minimal abductivist claim in a more mathematically precise way. To this end, we will assume that a rational agent's credences in various hypotheses can be represented by a probability function, P. We do not take this assumption to be an integral part of abductivism itself, but it can safely be added for the purposes of investigating whether abductivism can be reconciled with Bayesianism.

We will also assume that the explanatory power of various hypotheses can be represented by a function, $\mathcal{E}$, which takes a hypothesis, H, and a body of evidence, E, as input, and outputs a real number, $\mathcal{E}(H, E)$, representing how well or badly H explains E. We do not assume that $\mathcal{E}(H, E)$ is itself probabilistic or depends on the probability of H, E, or indeed any other probabilities. The only assumption we make is that the range of $\mathcal{E}(H, E)$ is some subinterval of the real numbers. Again, we do not take this assumption to be an integral part of abductivism itself, since abductivists might deny that degrees of explanatory power are as sharp or fine-grained as the real numbers, but it allows us to capture, in a simple way, the idea (which *is* integral to abductivism) that explanatory power comes in degrees.

Here, then, is what we take to be minimally involved in saying that abduction is a rational pattern of reasoning:

becomes relevant in connection with the principle called "Irrelevance of Priors."

[6]Here and henceforth, we are presuming that the hypothesis is not already assigned an extreme probability of 0 or 1, since in that case it is already decisively settled whether the hypothesis is true or false.

Minimal Abductivism: $P_E(H)$ is an increasing function of $\mathcal{E}(H, E)$.

A couple of things are worth noting about this principle, which connect it to what has been said above.

First, Minimal Abductivism is a purely ordinal claim: it says that an increase in $\mathcal{E}(H, E)$ should lead to an increase in $P_E(H)$, but it says nothing about how big the increase should be.

Second, Minimal Abductivism leaves open whether the posterior probability of H depends on factors other than its explanatory power: it says that $P_E(H)$ is an increasing function of $\mathcal{E}(H, E)$, but it doesn't say that $P_E(H)$ is a function of nothing else. In the next section, we give an example of what such other factors may be.

Third, Minimal Abductivism is not a comparative claim: it doesn't say that $P_E(H_1) > P_E(H_2)$ if $\mathcal{E}(H_1, E) > \mathcal{E}(H_2, E)$, precisely because the posterior probability of a hypothesis might, for all Minimal Abductivism says, depend on factors other than its explanatory power. In this respect, Minimal Abductivism is weaker than most articulations of IBE, which tend to be comparative.

Finally, it is worth emphasizing that Minimal Abductivism doesn't take a stance on whether conditionalization is the correct updating rule: for all Minimal Abductivism says, the posterior probability $P_E(H)$ may or may not be identical to $P(H|E)$. So, in particular, Minimal Abductivism doesn't commit us to thinking of abduction as the practice of assigning "bonus points" to hypotheses that are explanatorily powerful, although it is compatible with this view. The question we will be interested in is whether this minimal form of abductivism can be reconciled with conditionalization.

Needless to say, abductivists might want to endorse something stronger than Minimal Abductivism. But as we will see in the next section, it turns out that even this minimal abductivist principle, when combined with Bayesian updating by conditionalization, places strong and controversial constraints on which form the explanatory power measure can take.

On the other hand, one can also imagine weakening Minimal Abductivism even further. For example, those who think of abduction as a heuristic to approximate Bayesian reasoning might hold that $P_E(H)$ doesn't *always* increase with $\mathcal{E}(H, E)$, although it does so in a wide range of cases.[7] We will not pursue such a view in detail here, since the heuristic conception of abduction already

[7] We are indebted to Bob Beddor on this point.

concedes to the incompatibilist that abduction involves at least a small deviation from conditionalization. However, we suspect that the same challenges raised for compatibilism below can be made to apply to such heuristic conceptions of abduction as well, by restricting the scope of the discussion to whatever subclass of cases $P_E(H)$ is claimed to increase with $\mathcal{E}(H, E)$.

3 Minimal Abductivism and Conditionalization

In this section, we will present two formal results about the relationship between Minimal Abductivism, Conditionalization, and the explanatory power measure. Proofs of the results are provided in the appendix.

The first result provides a sufficient condition on jointly satisfying Minimal Abductivism and Conditionalization:

Result 1. *Minimal Abductivism and Conditionalization are jointly satisfiable if the explanatory power measure is ordinally equivalent to the following "ratio" measure:*

$$\mathcal{E}_r(H, E) = \frac{P(E|H)}{P(E)}.$$

This result shows that it is indeed possible to satisfy Minimal Abductivism and Conditionalization at the same time by adopting $\mathcal{E}_r$, or an ordinal equivalent of $\mathcal{E}_r$, as our explanatory power measure. To say that two explanatory power measures are "ordinally equivalent" is to say that they rank all hypothesis-evidence pairs in the same way: that is, $\mathcal{E}$ and $\mathcal{E}'$ are ordinally equivalent if and only if $\mathcal{E}(H, E) \geq \mathcal{E}(H', E') \Leftrightarrow \mathcal{E}'(H, E) \geq \mathcal{E}'(H', E')$. So, for example, $\mathcal{E}_r$ is ordinally equivalent to the following explanatory power measure, which was first proposed by I. J. Good (1960):

$$\mathcal{E}_G(H, E) = \log \frac{P(E|H)}{P(E)}.$$

Ordinally equivalent measures of explanatory power can arguably be treated as equivalent *simpliciter*, because the absolute scale on which explanatory power is measured arguably does not substantively matter (just as it does not substantively matter whether temperature is measured on a Fahrenheit or Celsius scale).[8] We will not be relying on this assumption for anything in this paper,

[8] Although see Vassend (2019) for a dissenting perspective.

but for simplicity we will work with $\mathcal{E}_r$ rather than $\mathcal{E}_G$ throughout.[9]

Before we look at whether $\mathcal{E}_r$ is a plausible measure of explanatory power, we want to present our second result. To state this result, some additional bookkeeping is needed. As we remarked earlier, the posterior probability of a hypothesis may depend on factors other than its explanatory power, such as its prior probability (as illustrated by the Squares N' Primes example). More generally, we want to allow that $P_E(H)$ may depend on $\mathcal{E}(H, E)$, $P(H)$, as well as other factors that reflect non-explanatory ways in which E may influence the posterior plausibility of H. We will assume that these other factors can be quantified in terms of real numbers, X_1, X_2, ..., X_n, just like we assume that prior plausibility and explanatory power may be quantified in terms of $P(H)$ and $\mathcal{E}(H, E)$, respectively. We do not claim that $P_E(H)$ will always, or even typically, depend on factors other than $\mathcal{E}(H, E)$ and $P(H)$, but we think it is plausible that it sometimes will.

As an example of what such other factors may look like, consider a case where we are trying to explain why some event happened. In many contexts, though perhaps not all, giving an explanation of an event arguably involves identifying its *causes*, as well as the role each cause played in bringing about the event. For example, if Tom has lung cancer and we seek to explain why, then a causal account is plausibly what we are after. Now, we know that smoking is an important cause of lung cancer, so the hypothesis that Tom is a smoker would, if true, be a good explanation of why he has cancer. More formally, it is plausible that $\mathcal{E}(Smoking, Cancer)$ is high. Suppose that it is also the case that Tom has a high prior probability of being a smoker (perhaps because of his family background or work environment) so that $P(Smoking)$ is also high. It is easy to see that these factors do not imply that $P_{Cancer}(Smoking)$ must also be high. To take a simple (if somewhat contrived) example, suppose that having lung cancer is strongly correlated with having a certain gene, and that having this gene in turn strongly reduces the chance that one will be a smoker (figure 1 shows a possible causal diagram depicting this situation). Under these circumstances, the evidence that Tom has cancer may in fact reduce the probability that he is a smoker, even though his prior probability of smoking is high and smoking is a cause of cancer. This is because having cancer will increase the probability that Tom has the gene, which in turn will decrease the probability that he

[9]For those who wonder why anyone might prefer to work with $\mathcal{E}_G$ rather than $\mathcal{E}_r$, one benefit of adding the log-factor is that it creates a natural zero point, since $\mathcal{E}_G(H, E) = 0$ when E is probabilistically independent of H.

is a smoker. In the vernacular of the causal modelling framework of the type developed by e.g. Pearl (2009), there is a (non-causal) "backdoor path" between smoking and cancer. This backdoor path is then an example of an "X factor" that affects the posterior plausibility of the smoking hypothesis.

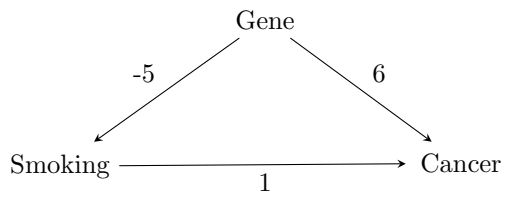

Figure 1: *Causal diagram illustrating how the evidence that Tom has cancer may reduce the probability that he is a smoker, even though his prior probability of smoking is high and smoking is a cause of cancer. The numbers next to the arrows represent degrees of causal strength: positive numbers indicate positive causal influence and negative numbers indicate negative causal influence. For our purposes, it doesn't matter where these numbers come from or what, precisely, they mean. For simplicity, we assume that the causal diagram is complete, so that there are no other relevant variables that act as common causes of any of the variables that occur in the diagram.*

We think examples like the preceding are a good reason to think that factors other than prior probability and explanatory power may affect the overall posterior plausibility of a hypothesis. Hence, we want our framework to allow for the possibility that such factors sometimes exist. We emphasize, however, that our subsequent argument does not depend on the assumption that such factors *must* exist—indeed, the argument could be significantly simplified if we assumed that such factors do not exist. However, given that we want to allow for the possibility that they do, we need to make some additional assumptions about how these other factors relate to both explanatory power and prior probability.

The first assumption we will make is that such other factors, if they exist, are *separable* from $\mathcal{E}(H, E)$ and $P(H)$ in the following sense:

> **Separability**: It is possible to hold fixed at any value any additional factors $X_1, X_2, ..., X_n$ influencing $P_E(H)$ while varying $\mathcal{E}(H, E)$ and $P(H)$.

Separability is arguably not a substantive assumption, for suppose X_i is such that it is *not* possible to hold it fixed at any value while varying $\mathcal{E}(H, E)$ and $P(H)$. Then X_i must be a function of either $\mathcal{E}(H, E)$ or $P(H)$ (or both), and hence we can write X_i as a function of $\mathcal{E}(H, E)$, $P(H)$, and some variable Y_i,

such that Y_i is *not* a function of either $\mathcal{E}(H, E)$ or $P(H)$.[10] Hence, the claim that $P_E(H)$ functionally depends on $\mathcal{E}(H, E)$, $P(H)$, and X_i is equivalent to the claim that $P_E(H)$ functionally depends on $\mathcal{E}(H, E)$, $P(H)$, and Y_i, where Y_i does not functionally depend on either $\mathcal{E}(H, E)$ or $P(H)$. In other words, assuming Separability involves no loss of generality.

A second piece of bookkeeping concerns the relationship between $\mathcal{E}(H, E)$ and $P(H)$. Since $\mathcal{E}(H, E)$ is supposed to represent how well H would explain E, *if H were true*, it is natural to think that $\mathcal{E}(H, E)$ should not depend on $P(H)$. Indeed, conditions to this effect have been posited by Schupbach and Sprenger (2011) and Eva and Stern (2019). However, the claim that $\mathcal{E}(H, E)$ does not depend on $P(H)$ may not be true without qualification. As we will see in the next section, it is sensible to think that $\mathcal{E}(H, E)$ may depend on $P(E)$ and $P(E|H)$, and both $P(E)$ and $P(E|H)$ may in turn be expressed as functions of $P(H)$.[11] Moreover, the additional factors $X_1, X_2, \ldots, X_n$ that we want to allow influencing the posterior probability of H may conceivably also influence $\mathcal{E}(H, E)$, if only indirectly (at least we do not want to foreclose this possibility). Thus, *if* $\mathcal{E}(H, E)$ functionally depends on $P(E|H)$, $P(E)$, or some other factor X_i influencing the posterior probability of H, then it might be possible for $P(H)$ to indirectly influence $\mathcal{E}(H, E)$ via these factors. Nonetheless, we think it remains true that $P(H)$ should not have a *direct* influence on $\mathcal{E}(H, E)$. In other words, if $P(H)$ has an influence on $\mathcal{E}(H, E)$ at all, then it can only be via its influence on either $P(E|H)$, $P(E)$, or possibly one of the other factors X_i that may influence the posterior probability of H. Hence, we think the following condition is plausible:

> **Irrelevance of Priors**: If $P(E|H)$, $P(E)$, and any factors other than $\mathcal{E}(H, E)$ and $P(H)$ that may influence $P_E(H)$ are held fixed, then it is not possible to change $\mathcal{E}(H, E)$ by varying $P(H)$.

To forestall confusion, we emphasize again that the Irrelevance of Priors condition does not assume that $\mathcal{E}(H, E)$ in fact depends on either $P(E|H)$, $P(E)$, or indeed any other X_i factors. We simply want to allow for this possibility for the sake of generality.

The final ingredient we need to state our second result concerns the rela-

[10] If there is no such variable Y_i, then X_i is purely a function of $\mathcal{E}(H, E)$ or $P(H)$, and so is not an additional factor affecting $P_E(H)$ after all.

[11] Since $P(E) = P(E|H)P(H) + P(E|\neg H)P(\neg H)$ and $P(E|H) = \frac{P(H \wedge E)}{P(H)}$.

tionship between $\mathcal{E}(H, E)$ and the other non-explanatory factors X_i that may affect the posterior probability of H. Here, again, we find it plausible that X_i should not have a *direct* influence on $\mathcal{E}(H, E)$. Indeed, if it were the case that X_i had a direct effect on $\mathcal{E}(H, E)$, then arguably that would *ipso facto* make X_i an explanatory variable, which should be captured by $\mathcal{E}(H, E)$ itself.

The idea that X_i should not have a direct effect on $\mathcal{E}(H, E)$ also finds intuitive support if we look at particular cases. Consider our earlier example where we are trying to explain Tom's cancer. We saw that even if the smoking hypothesis is a strong causal explanation and Tom has a high prior probability of being a smoker, that does not mean that Tom has a high posterior probability of being a smoker, since there may be other (non-causal and therefore non-explanatory) factors that influence the posterior probability that Tom is a smoker. In particular, in Figure 1 we noted that the smoking hypothesis may lower the probability that Tom has cancer via a (non-causal) "backdoor path." In this particular case, it is plausible that the strength of this backdoor path should not have a direct influence on the degree to which the smoking hypothesis, if true, would causally explain the fact that Tom has cancer.

However, even though we find it plausible that no non-explanatory factor X_i should have a direct influence on $\mathcal{E}(H, E)$, we still want to leave room for the possibility that it has an indirect effect via its effect on other factors that may affect $\mathcal{E}(H, E)$. For example, non-explanatory factors might conceivably influence the surprisingness of the evidence, i.e., $P(E)$, or the likelihood of the hypothesis on the evidence, i.e., $P(E|H)$. Again, we do not claim or argue that X_i necessarily will have this sort of effect, but we want to leave room for the possibility that it does.

The preceding considerations motivate the following condition, which mirrors closely our earlier Irrelevance of Priors condition:

> **Irrelevance of Non-Explanatory Factors**: If $P(E|H)$, $P(E)$, and $P(H)$ are held fixed, then it is not possible to change $\mathcal{E}(H, E)$ by varying a non-explanatory factor X_i that may influence $P_E(H)$.

We find both Irrelevance of Priors and Irrelevance of Non-Explanatory Factors to be well-motivated and plausible. However, note that even though we are excluding explanatory power measures that directly depend on prior probabilities or non-explanatory factors, we are not thereby committing ourselves to a form of abduction that is more restrictive than it otherwise could have been, because

we are still allowing for these other factors to influence the posterior probability distribution. In other words, the assumption that prior probabilities and non-explanatory factors do not have a direct influence on explanatory power involves no loss of generality.

With the above conditions in place, we can finally state our second result:

Result 2. *Minimal Abductivism and Conditionalization together with the auxiliary assumptions Separability, Irrelevance of Priors, and Irrelevance of Non-Explanatory Factors jointly entail that $\mathcal{E}$ is ordinally equivalent to $\mathcal{E}_r$.*

This result may be contrasted with an earlier result due to Sprenger and Hartmann (2019), which shows that $\mathcal{E}(H, E)$ must be ordinally equivalent to $\mathcal{E}_r(H, E)$ if the following conditions hold:

Confirmatory Value: $\mathcal{E}(H, E_1) > \mathcal{E}(H, E_2)$ iff $P(H|E_1) > P(H|E_2)$.

Difference-Making: $\mathcal{E}(H, E)$ is solely a function of $P(E|H)$ and $P(E)$.

We have derived a similar conclusion from weaker assumptions. Indeed, in our result we have not even assumed that $\mathcal{E}(H, E)$ can be defined in terms of probabilities. Instead, it follows from our result that it must be.[12]

In sum, then, Result 1 shows that we can jointly satisfy Minimal Abductivism and Conditionalization by adopting $\mathcal{E}_r$ as our measure of explanatory power, and Result 2 shows that no other measure of explanatory power allows us to do so (assuming Separability, Irrelevance of Priors, and Irrelevance of Non-Explanatory Factors). The next question is whether $\mathcal{E}_r$ is an acceptable measure of explanatory power.

[12]There is a subtle, but important point here that we feel compelled to address. We have made much of the fact that the posterior probability $P_E(H)$ may, in certain cases, depend on factors $X_1, X_2, \ldots, X_n$ other than $\mathcal{E}(H, E)$ and $P(H)$. In light of Result 2 one might wonder how any other such factors *could* play any role. After all, Conditionalization guarantees that $P_E(H)$ is identical to $P(E|H)P(H)/P(E)$. If $\mathcal{E}(H, E)$ is ordinally equivalent to $P(E|H)/P(E)$, as Result 2 seems to establish, then it follows that $P_E(H)$ can be written purely as a function of $P(H)$ and $\mathcal{E}(H, E)$, and hence it looks like there is no role to play for any other factors after all. The solution to this puzzle is that even though such other factors cannot affect the *form* assumed by the explanatory power measure, which according to Result 2 has to be $P(E|H)/P(E)$, they *can* affect each of $P(E|H)$ and $P(E)$, i.e., the probabilities that occur in this ratio. The proof of Result 2 makes this apparent.

4 Measuring Explanatory Power

In recent years, Bayesian philosophers have proposed several different measures of explanatory power, which have been investigated more or less independently of the debate concerning the relationship between Bayesianism and abduction. What all of these proposed measures have in common is that they are *purely probabilistic* in the sense that they depend only on the probabilities associated with the hypothesis and evidence under consideration. So, for example, $\mathcal{E}_r$ is purely probabilistic, since $\mathcal{E}_r(H, E)$ depends only on $P(E|H)$ and $P(E)$. The same goes for the following two alternatives to $\mathcal{E}_r$, which have been suggested by Schupbach and Sprenger (2011) and Crupi and Tentori (2012), respectively:

$$\mathcal{E}_{SS}(H, E) = \frac{P(H|E) - P(H|\neg E)}{P(H|E) + P(H|\neg E)}$$

$$\mathcal{E}_{CT}(H, E) = \begin{cases} \dfrac{P(E|H) - P(E)}{1 - P(E)} & \text{if } P(E|H) \geq P(E) \\ \dfrac{P(E|H) - P(E)}{P(E)} & \text{if } P(E|H) < P(E) \end{cases}$$

The motivation that Schupbach and Sprenger offer for analyzing explanatory power in probabilistic terms is that explanatory power seems closely related to a hypothesis' ability to reduce the degree to which the evidence under consideration is *surprising* or *unexpected*. For example, while it may have been surprising to learn that light moves at a certain fixed speed in vacuum (why 300,000 km/s rather than some other speed?), this is precisely what we would expect given Maxwell's theory of electromagnetism. In other words, the probability that light moves at this particular speed is higher conditional on Maxwell's theory of electromagnetism than it is unconditionally. Examples like this may be taken to suggest that an adequate measure of explanatory power should obey the following probabilistic condition:

> **Surprise Reduction**: $\mathcal{E}(H, E)$ is an increasing function of $P(E|H)$, a decreasing function of $P(E)$, and a constant function if E and H are probabilistically independent.[13]

[13] The corresponding condition in Schupbach and Sprenger (2011) is called "positive relevance."

To say that good explanations make their explananda less surprising is not to say, conversely, that *whenever* a hypothesis makes a body of evidence less surprising, the hypothesis is thereby explanatory of the evidence. It may well be that $P(E|H) > P(E)$, even if H is in no way explanatory of E. For example, it may be unsurprising to learn that Jim is a smoker given that he has lung cancer, even if the latter fact is not explanatory of the former.[14] The claim is just supposed to be that, insofar as H is explanatory of E, the degree of explanatory power that H has over E increases with the difference between $P(E|H)$ and $P(E)$. Note that since $\mathcal{E}_r(H, E)$ satisfies Surprise Reduction, our results from the previous section show that the compatibilist is committed to Surprise Reduction.

Even if an adequate measure of explanatory power should obey Surprise Reduction, this does not suffice to show that we can or should analyze explanatory power in purely probabilistic terms. Indeed, Roche and Sober (2023) have recently presented a general set of objections against all extant probabilistic measures of explanatory power (including $\mathcal{E}_r$, $\mathcal{E}_{SS}$, and $\mathcal{E}_{CT}$). However, Result 2 from the previous section shows that the compatibilist is committed to analyzing explanatory power probabilistically. Hence, those who are convinced by Roche and Sober's arguments may take the results from the previous section to show that compatibilism is untenable in virtue of being committed to a purely probabilistic analysis of explanatory power.

However, even among those who are more optimistic about the prospects of analyzing explanatory power in purely probabilistic terms, $\mathcal{E}_r$ has been subject to a great deal of criticism. We will not review all of the details of this debate here, since these can be found elsewhere (Sprenger and Hartmann, 2019, ch. 7), but it is worth briefly considering what has perhaps been the main reason for skepticism about $\mathcal{E}_r$ and its ordinal equivalents.

The alleged problem, which to our knowledge was first articulated by Schupbach and Sprenger (2011), is that $\mathcal{E}_r$ and its ordinal equivalents are *invariant under the addition of irrelevant evidence.* Suppose H is a good explanation of E, and suppose I is an irrelevant body of evidence in the sense that I and H are probabilistically independent of each other given E: $P(I|H \wedge E) = P(I|E)$. For simplicity, assume also that there are no additional factors $X_1, X_2, \ldots, X_n$ aside from $\mathcal{E}(H, E)$ and $P(H)$ that influence the posterior probability of H. A

[14] More generally, since $P(A|B) > P(A) \Leftrightarrow P(B|A) > P(B)$, a bidirectional interpretation of Surprise Reduction would imply that A is explanatory of B if and only if B is explanatory of A, which is clearly false.

simple calculation then shows that, if we adopt $\mathcal{E}_r$ or any of its ordinal equivalents, H comes out as having the same explanatory power over the conjunction $E \wedge I$ as it has over E itself.[15] But, the objection goes, this is an implausible verdict in many cases. For example, consider again the situation where we are trying to explain Tom's cancer, except we now assume that there is no "smoking gene" lurking in the background. The evidence, E, is the observation that Tom has cancer, H is the hypothesis that he is a smoker, and we can take I to be some causally and probabilistically irrelevant proposition such as "Tom has a green toothbrush." The causal relationship between these variables is depicted in figure 2.

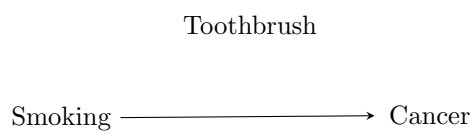

Figure 2: *Causal diagram similar to that in figure 1, except that the "smoking gene" is replaced by a causally irrelevant variable representing the color of Tom's toothbrush.*

Under the specified conditions where there are no additional X_i factors to consider, $\mathcal{E}(Smoking, Cancer)$ and $P(Smoking)$ jointly suffice to determine the posterior probability $P_{Cancer}(Smoking)$. So, if we adopt $\mathcal{E}_r$ or any of its ordinal equivalents as our measure of explanatory power, we commit ourselves to saying that *Smoking* has the same explanatory power over the conjunction $Cancer \wedge Toothbrush$ as it has over *Cancer* itself. But it seems plausible that the hypothesis that Tom smokes should be a better explanation of the observation that he smokes than of the observation that he smokes *and* has a green toothbrush.

Of course, there may be room for debate about what to make of this objection. Perhaps our reluctance to accept that *Cancer* and $Cancer \wedge Toothbrush$ are equally well explained by *Smoking* is due to pragmatic (broadly Gricean) factors rather than any actual difference in how well *Smoking* explains *Cancer* and $Cancer \wedge Toothbrush$, respectively. If so, $\mathcal{E}_r$ may be defensible after all. Our aim here is not to settle the question of whether a purely probabilistic measure of explanatory power is ultimately tenable, and, if so, what such a measure should look like. We simply want to point out that the results from the previous

[15] $\mathcal{E}_r(H, E \wedge I) = \frac{P(E \wedge I|H)}{P(E \wedge I)} = \frac{P(I|H \wedge E)P(E|H)}{P(I|E)P(E)} = \frac{P(E|H)}{P(E)} = \mathcal{E}_r(H, E)$. The same calculation holds, *mutatis mutandis*, for any ordinal equivalent of $\mathcal{E}_r$.

section reveal a highly substantive commitment of compatibilism, one which has not yet been recognized let alone defended.

5 Conclusion: a trilemma

Result 2 shows that even a very weak form of abductivism, which holds that explanatory considerations are just one factor among several that determine the overall plausibility of a hypothesis, combined with the constraint that the overall plausibility assessment be consistent with conditionalization, forces the measure of explanatory power to not only be probabilistic, but to assume a very particular form, namely $\mathcal{E}_r$. On the other hand, as we saw in Section 4 there are good reasons for thinking that $\mathcal{E}_r$ is not an adequate measure of explanatory power. We therefore have the following inconsistent triad (conditional on the assumptions of Result 2 holding):

1. Minimal Abductivism
2. Conditionalization
3. $\mathcal{E}_r$ is not an adequate measure of explanatory power.

Obviously, the only way out for compatibilists is to reject the third proposition and embrace $\mathcal{E}_r$ as their favored measure of explanatory power. Result 1 ensures that they can then accept both Minimal Abductivism and Conditionalization without issue. By contrast, incompatibilists have different options depending on their specific commitments. Incompatibilist Bayesians, such as van Fraassen, will be forced either to reject the third proposition, and go along with the compatibilists in accepting $\mathcal{E}_r$ as the correct measure of explanatory power, or to reject Minimal Abductivism. The costs associated with the first choice have already been discussed. On the other hand, rejecting Minimal Abductivism entails accepting an extremely strong form of incompatibilism, which maintains that explanatory considerations cannot even be one (defeasible) factor among many factors that determine the overall plausibility of hypotheses.

Other incompatibilists, such as Douven (2022), may be comfortable rejecting Conditionalization, in which case they can accept Minimal Abductivism while denying that $\mathcal{E}_r$ is an adequate measure of explanatory power. However, one of the main attractions of Conditionalization—emphasized by Jaynes (2003), Climenhaga (2017), and Pettigrew (2021), among others—is that it has many

desirable properties. For example, to a Bayesian, it does not matter whether you update your probability distribution sequentially on E_1 and then E_2, or on both pieces of evidence at the same time—the final posterior is the same either way. Presumably, even incompatibilists who reject Conditionalization would prefer that abduction obey such basic constraints, if possible. Recent work in statistics and philosophy (e.g., Bissiri, Holmes, and Walker (2016)) and Vassend (2022)) explore generalizations of Bayesianism that replace Conditionalization with alternative updating procedures that retain as many of the desirable properties of Conditionalization as possible. An intriguing possible way out of the trilemma we have posed in this discussion is to replace Conditionalization with one of these alternative updating procedures. We hope to explore this possibility in future work.

Appendix: Proofs of Results

Result 1. *Minimal Abductivism and Conditionalization are jointly satisfiable if the explanatory power measure is ordinally equivalent to the following "ratio" measure:*

$$\mathcal{E}_r(H, E) = \frac{P(E|H)}{P(E)}.$$

Proof. Assuming $\mathcal{E}$ is ordinally equivalent to $\mathcal{E}_r$, there must exist an increasing function, f, such that

$$\mathcal{E}(H, E) = f\left(\frac{P(E|H)}{P(E)}\right).$$

By Bayes' theorem, we can then write Conditionalization as follows:

$$P_E(H) = \frac{P(E|H)}{P(E)}P(H) = f^{-1}(\mathcal{E}(H, E))P(H).$$

Since the inverse of an increasing function is itself an increasing function, it follows that $P_E(H)$ is an increasing function of $\mathcal{E}(H, E)$, as required by Minimal Abductivism. □

Result 2. *Minimal Abductivism and Conditionalization together with the auxiliary assumptions Separability, Irrelevance of Priors, and Irrelevance of Non-Explanatory Factors jointly entail that $\mathcal{E}$ is ordinally equivalent to $\mathcal{E}_r$.*

Proof. Suppose $P_E(H)$ is a function of $\mathcal{E}(H,E)$, $P(H)$, and (possibly) additional factors $X_1, X_2, \ldots, X_n$. For simplicity, we will let $\mathcal{X}$ denote the vector of factors $X_1, X_2, \ldots, X_n$. Hence, we can write:

$$P_E(H) = f(\mathcal{E}(H,E), P(H), \mathcal{X}). \tag{1}$$

Minimal Abductivism tells us that the function f is increasing in its first argument. Furthermore, Conditionalization and Bayes' theorem tell us that:

$$\frac{P(E|H)}{P(E)} P(H) = f(\mathcal{E}(H,E), P(H), \mathcal{X}). \tag{2}$$

Next, Separability assures us that it is possible to hold $\mathcal{X}$ fixed at any arbitrary set of values while $P(H)$ and $\mathcal{E}(H,E)$ vary freely. So, if we suppose that $\mathcal{X}$ is held fixed at some set of values and use g to denote the resulting function, we can write:

$$\frac{P(E|H)}{P(E)} P(H) = g(\mathcal{E}(H,E), P(H)), \tag{3}$$

where again g is increasing in its first argument. Note that Irrelevance of Priors guarantees that $\mathcal{E}(H,E)$ is not solely a function of $P(H)$. Hence, it is possible to hold $P(H)$ fixed while varying $\mathcal{E}(H,E)$, which entails that there must exist an increasing function, h, such that:

$$\mathcal{E}(H,E) = h\left(\frac{P(E|H)}{P(E)}\right), \tag{4}$$

which means that, for any fixed value of $P(H)$, $\mathcal{E}(H,E)$ is ordinally equivalent to $\mathcal{E}_r(H,E)$. This still leaves open the possibility that $\mathcal{E}(H,E)$ depends on both $\mathcal{E}_r(H,E)$ and $P(H)$. In other words, what we have established so far is that there exists a function, k, such that:

$$\mathcal{E}(H,E) = k\left(\frac{P(E|H)}{P(E)}, P(H)\right), \tag{5}$$

where k increases in its first argument. Note that there are two distinct ways in which $P(H)$ may exert an influence on $\mathcal{E}(H,E)$: first, it can affect $\mathcal{E}(H,E)$ *indirectly* by changing either $P(E|H)$ or $P(E)$. Indeed, we know that it can have this sort of effect because both $P(E|H)$ and $P(E)$ may be written as functions of $P(H)$, as we point out in footnote 7. Alternatively, $P(H)$ might

have an independent, direct influence on $\mathcal{E}(H, E)$. However, Irrelevance of Priors guarantees that $\mathcal{E}(H, E)$ cannot depend on $P(H)$ in this latter sense, given that it depends on $P(E|H)$ and $P(E)$ and $\mathcal{X}$ is held fixed. Hence, we can write (5) as follows:

$$\mathcal{E}(H, E) = k\left(\frac{P(E|H)}{P(E)}\right), \tag{6}$$

where $P(H)$ is no longer an independent variable on which $\mathcal{E}(H, E)$ depends. Thus, we conclude that $\mathcal{E}(H, E)$ is ordinally equivalent to $\mathcal{E}_r(H, E)$, even if we let $P(H)$ vary freely.

The proof so far depends on the assumption that $\mathcal{X}$ is held fixed. Thus, what we have established so far is that there exists a function, l, such that:

$$\mathcal{E}(H, E) = l\left(\frac{P(E|H)}{P(E)}, \mathcal{X}\right), \tag{7}$$

where l is increasing in its first argument.

However, in the same way that Irrelevance of Priors entails that $P(H)$ cannot be an independent variable on which k depends, Irrelevance of Non-Explanatory Factors implies that any non-explanatory factor X_i cannot influence $\mathcal{E}(H, E)$ if $P(E|H)$ and $P(E)$ are held fixed, which entails that we can rewrite (7) as follows:

$$\mathcal{E}(H, E) = l\left(\frac{P(E|H)}{P(E)}\right), \tag{8}$$

where the explicit dependence on $\mathcal{X}$ has been removed. This does not mean that $\mathcal{X}$ can have no influence at all on $\mathcal{E}(H, E)$, of course. Indeed, varying $\mathcal{X}$ may influence each of $P(E|H)$ and $P(E)$ individually. However, varying $\mathcal{X}$ will not change the fact that $\mathcal{E}(H, E)$ is ordinally equivalent to the ratio $\frac{P(E|H)}{P(E)}$. Hence, we conclude that $\mathcal{E}(H, E)$ is ordinally equivalent to $\mathcal{E}_r(H, E)$ simpliciter. $\square$

Acknowledgments: For helpful feedback, we are grateful to Zach Barnett, Bob Beddor, two anonymous referees for Noûs, and audience members at the Barcelona Institute for Analytic Philosophy and the 1st Nordic Epistemology Network workshop.

References

Bissiri, P. G, Holmes, C. C. and Walker, S. G. (2016). A General Framework for Updating Belief Distributions. *Journal of the Royal Statistical Society: Series B*, 78, 1103–30. doi: 10.1111/rssb.12158

Cabrera, F. (2017). Can There Be a Bayesian Explanationism? On the Prospects of a Productive Partnership. *Synthese*, 194, 1245–72. doi: 10.1007/s11229-015-0990-z

Climenhaga, N. (2017). Inference to the Best Explanation Made Incoherent. *Journal of Philosophy* 114, 251–73. doi: 10.5840/jphil2017114519

Crupi, V. and Tentori, K. (2012). A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems). *Philosophy of Science*, 79, 365–85. doi: 10.1086/666063

Dellsén, F. (2024). *Abductive Reasoning in Science*. Cambridge University Press.

Dellsén, F. (2018). The Heuristic Conception of Inference to the Best Explanation. *Synthese*, 175, 1745–66. doi: 10.1007/s11098-017-0933-2

Douven, I. (2022). *The Art of Abduction*. MIT Press.

Eva, B. and Stern, R. (2019). Causal Explanatory Power. *The British Journal for the Philosophy of Science*, 70, 1029–50. doi: 10.1093/bjps/axy012

Glass, D. H. (forthcoming). Information and Explanatory Goodness. Fortcoming in *Erkenntnis*. doi: 10.1007/s10670-023-00687-2

Glass, D. H. (2023). How Good is an Explanation? *Synthese*, 201, 1–26. doi: 10.1007/s11229-022-04025-x

Glymour, C. (2015). Probability and the Explanatory Virtues. *British Journal for the Philosophy of Science*, 66, 591–604. doi: 10.1093/bjps/axt051

Good, I. J. (1968). Corroboration, Explanation, Evolving Probability, Simplicity and a Sharpened Razor. *British Journal for the Philosophy of Science*, 19, 123–43. doi: 10.1093/bjps/19.2.123

Good, I. J. (1960). Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments. *Journal of the Royal Statistical Society. Series B (Methodological)*, 22, 319–31. doi: 10.1111/j.2517-6161.1960.tb00378.x

Harman, G. H. (1965). The Inference to the Best Explanation. *The Philosophical Review*, 74, 88–95. doi: 10.2307/2183532

Henderson, L. (2014). Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science*, 65, 687–715. doi: 10.1093/bjps/axt020

Lange, M. (2022). Putting Explanation Back Into 'Inference to the Best Explanation'. *Noûs*, 56, 84–109. doi: 10.1111/nous.12349

Lipton, P. (2004). *Inference to the Best Explanation*. Routledge.

McMullin, E. (1983). Values in Science. In P. Asquith and T. Nickle (Eds.), Proceedings of the 1982 biennial meeting of the philosophy of science association, 2, 3–28. East Lansing: Philosophy of Science Association.

Okasha, S. (2000). van Fraassen's Critique of Inference to the Best Explanation. *Studies in the History and Philosophy of Science Part A*, 31, 691–710. doi: 10.1016/s0039-3681(00)00016-9

Roche, W. and Sober, E. (2023). Purely Probabilistic Measures of Explanatory Power: A Critique. *Philosophy of Science*, 90, 129–49. doi: 10.1017/psa.2021.39

Schupbach, J. N. and Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science*, 78, 105–27. doi: 10.1086/658111

Sprenger, J. and Hartmann, S. (2019). *Bayesian Philosophy of Science*. Oxford University Press.

Thagard, P. (1978). The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy*, 75, 76–92. doi: 10.2307/2025686

Teller, P. (1973). Conditionalization and Observation. *Synthese*, 26, 218–58. doi: 10.1007/bf00873264

Pettigrew, R. (2021). On the Pragmatic and Epistemic Virtues of Inference to the Best Explanation. *Synthese*, 199, 12407–38. doi: 10.1007/s11229-021-03338-7

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. 2nd edition. Cambridge University Press.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Clarendon Press: Oxford.

Vassend, O. B. (2019). Confirmation and the Ordinal Equivalence Thesis. *Synthese*, 196, 1079–95. doi: 10.1007/s11229-017-1500-2

Vassend, O. B. (2022). Justifying the Norms of Inductive Inference. *British Journal for the Philosophy of Science*, 73, 135–60. doi: 10.1093/bjps/axz041

Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese*, 167, 125–43. doi: 10.1007/s11229-008-9305-y