The Old Evidence Problem and the Inference to the Best Explanation

Abstract

The Problem of Old Evidence (POE) states that Bayesian confirmation theory cannot explain why a theory H can be confirmed by a piece of evidence E already known. Different dimensions of POE have been highlighted. Here, I consider the dynamic and static dimension. In the former, we want to explain how the discovery that H accounts for E confirms H. In the latter, we want to understand why E is and will be a reason to prefer H over its competitors. The aim of the paper is twofold. Firstly, I stress that two recent solutions to the dynamic dimension, recently proposed by Eva and Hartmann, can be read in terms of Inference to the Best Explanation (IBE). On this basis, I gauge the weaknesses and strengths of the two models by showing that the two authors endorse a particular formulation of IBE, and that it is still unsure if it is the one descriptively used. Moreover, I contend that, while one condition of their first model is not expression of this formulation, the only condition of their second model is. Secondly, I focus on the static dimension of POE which, now, has to be expressed in IBE terms. To solve it, I rely on the counterfactual approach, and on a version of IBE in which explanatory considerations help to evaluate the terms in Bayes' theorem. However, it turns out that the problems of the counterfactual approach recur even when it is used to solve the static POE in IBE terms.

Keywords:
Bayesian Confirmation Theory; Problem of Old Evidence; Dynamic and Static Problem of Old Evidence; Inference to the Best Explanation; Bayesian Inference to the Best Explanation.

1. Introduction

Since Glymour (1980) first described it, the Problem of Old Evidence (POE) has been bedevilling the Bayesian theory of confirmation. It has been seen as a major descriptive flaw of such a theory, which struggles to account for a shared intuition of scientific reasoning according to which a theory H can be confirmed by a piece of evidence already known, i.e. by an old piece of evidence.

Different dimensions of POE have been highlighted (Eells, 1985; 1990), and, accordingly, different solutions have been proposed for them. Here, I will consider only two dimensions, i.e. the dynamic and static one. In the former, we want to explain how the discovery that H accounts for the old evidence confirms hypothesis H in the sense that it raises the subject's degree of belief in H. In the latter, instead, we want to understand why the old evidence is and will be a reason to prefer H over its competitors.

The aim of the paper is twofold. Firstly, I will highlight that two recent solutions to the dynamic dimension of POE, proposed by Eva and Hartmann (2020), can be read in terms of Inference to the Best Explanation (IBE). In fact, by using a different formal apparatus, both models show that learning that H is the only available hypothesis that adequately explains the old evidence confirms H.

Moreover, by making such a reading explicit, I will gauge the weaknesses and strengths of the two models. More precisely, I will show that Eva and Hartmann have in mind a specific understanding of the core IBE idea that explanatory considerations have a confirmational import. On this basis, pending the question if this is indeed the formulation of IBE descriptively used, I will point out that, while one condition of their first model is not an expression of such a formulation, the only condition employed in their second model is.

As for the second aim, I will highlight that the explicit realization that real cases of confirmation by old evidence are instances of IBE, sheds some light also on the static dimension of POE which now has to be expressed in IBE terms. To solve the static dimension of POE so expressed, I will rely on the counterfactual approach (Howson, 1984; 1985; 1991), and on the Bayesian IBE (Okasha, 2000; Lipton, 2001). The latter is a probabilistic version of IBE, according to which explanatory considerations help to evaluate the terms in Bayes' theorem. However, we will see that the problems that haunt the counterfactual approach recur even when it is used to solve the static POE in IBE terms.

To show my claims, I will follow this structure. In section 2, I will introduce the old evidence problem for Bayesian confirmation theory, its dynamic and static dimension, and the solutions proposed for them. In section 3, I will present Eva and Hartmann's two novel solutions to the dynamic POE. And, after having made their connection with IBE explicit, I will assess these two models in light of such a connection. Finally, section 4 will focus on the static POE in IBE terms. After having introduced the Bayesian IBE, I will point out how the latter can be employed along with the counterfactual approach to solve the IBE's reading of the static POE. Finally, I will explain how the problems of the counterfactual approach can also undermine the solution at hand.

2. The old evidence problem, and its dimensions

As said in the introduction, the old evidence problem states that the Bayesian confirmation theory cannot account for the intuition according to which a theory can be confirmed by an old piece of evidence.

The most famous instance of confirmation by old evidence is the one of Mercury's perihelion shift (Glymour, 1980). The advance of Mercury's perihelion was an anomalous piece of evidence, as it was not explained by the available scientific theories, like Newtonian mechanics. Then, Einstein realized that his General Theory of Relativity (GTR) accounted for this phenomenon. When such a relationship between the theory and the evidence was discovered, the evidence was already known: the nature of Mercury's perihelion had been the object of intense study by astronomers for many decades. However, according to many physicists, Mercury's perihelion shift confirms GTR because the latter resolves that observational anomaly.

From this example, we can derive a general pattern (Hartmann & Sprenger, 2019, pp. 131-132):

1. we start with an anomalous piece of evidence E.
2. At some point, it is discovered that theory H can account for E.
3. E is an old piece of evidence: at the time in which the relationship between H and E is developed, the scientist is already certain or almost certain that E is real.
4. E confirms H because the latter resolves the observational anomaly E.

Now, if we formalize this schema in a Bayesian fashion, we obtain:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} = P(H) \tag{1}$$

Indeed, since E is already known, the scientist's degree of confidence in it is maximal, i.e. P(E) = 1. From this, it follows that also P(E|H) is equal to 1. To see why, note that the law of total probability implies that:

$$P(E) = P(E|H)P(H) + P(E|\neg H)P(\neg H) \tag{2}$$

Thus $P(E) = 1$, for all $P(H) \in (0; 1)$, implies that $P(E|H) = P(E|\neg H) = 1$.

Given that the posterior probability of the theory, $P(H|E)$, is equal to its prior probability, $P(H)$, E does not confirm H, according to the notion of 'confirmation as increase in firmness' (Hartmann & Sprenger, 2019, pp. 50-55). The latter, in fact, states that E confirms H if and only if $P(H|E) > P(H)$. This is the relevant sense of confirmation in this example, as confirming evidence raises the rational agent's degree of belief in the theory. Thus, as announced, Bayesian confirmation theory cannot explain the scientific intuition according to which a theory can be confirmed by a piece of evidence already known.

Different dimensions of POE have been highlighted (Eells, 1985; 1990). Here, I will take into consideration only two of them, i.e. the dynamic and static dimension.

2.1. The dynamic dimension of POE

In the dynamic POE, we find ourselves in a moment in time in which H and its relationship with E are discovered. What we want to know is how the discovery that H accounts for E raises the subject's degree of belief in H.

To understand this better, let us take into consideration again the Mercury perihelion shift's example. It took Einstein some time to find out that GTR (H) entailed the anomalous shift (E), i.e. $X = H \vdash E$ (Brush, 1989; Earman, 1992). By learning X, Einstein increased his degree of belief in H, as X was a surprising fact, in line with the scientific intuition that surprising evidence has more confirmational value. So, we can take the inequality $P(H|X, E) > P(H|E)$ to represent Einstein's actual degrees of belief respectively after and before learning X. How this inequality between actual degrees of belief is reached is precisely what we want to understand in the dynamic POE.

Notice that the confirming evidence is not E itself, but learning the *logical* fact X. Modelling this situation in a Bayesian fashion implies an abandonment of the logical omniscience of the Bayesian agent who otherwise cannot learn the logical fact X, always part of her background knowledge. Indeed, the latter is the starting point of the 'classical' approaches to the dynamic POE, due to Garber (1983), Niiniluoto (1983), Jeffrey (1983), and Earman (1992). In fact, despite their differences, they all rely on the following:

1. they abandon the logical omniscience of the Bayesian agent.

2. They add atomic sentences of the form X to the set of propositions about which the agent can have degrees of belief.
3. They show that, under suitable assumptions, we have that $P(H|X, E) > P(H|E)$.

However, all in all, the classical approaches struggle to work. Garber and Niiniluoto's ones only show the *existence* of probability functions that solve dynamic POE. But they do not say what we need to know, i.e. what conditions capture *real* cases of confirmation by old evidence and lead the scientist to conclude $P(H|X, E) > P(H|E)$. This gap is filled by Jeffrey and Earman's approaches, which, however, are based on very problematic assumptions (for a complete overview of the classical approaches, and their problems, see Hartmann & Sprenger, 2019, pp. 133-137).

In addition to what we have just said, Hartmann and Fitelson (2015, p. 714) highlight another reason why the classical solutions are not adequate. That is, they do not allow for the natural possibility that explanatory facts that can also be non-deductive can provide the basis for confirmation by old evidence. Thus, they propose a new interpretation of X and Y. Namely:

- X: H *adequately explains* E.
- Y: H's best competitor (H') *adequately explains* E.

Then, they derive $P(H|X) > P(H)$ from the following qualitative constrains (assuming that $P(E) = 1$):

HF1 $P(H|X, \neg Y) > P(H|\neg X, \neg Y)$.

HF2 $P(H|X, \neg Y) > P(H|\neg X, Y)$.

HF3 $P(H|X, Y) > P(H|\neg X, Y)$.

HF4 $P(H|X, Y) \geq P(H|\neg X, \neg Y)$.

Prima facie, these conditions seem plausible. According to **HF1** and **HF2**, H's probability is higher supposing that it adequately explains E and H' does not, than supposing that H does not adequately explain E, along with H' (**HF1**), or when H' adequately explains E (**HF2**). **HF3** says that if H' adequately explains E, then H is more probable if it adequately explains E than if it does not. Finally, **HF4** is an exclusive conjunction of two claims: H's probability is strictly higher, given that both H and H' adequately explain E, than supposing that neither H nor H' adequately explain E; H's probability, given the supposition that both H and H' adequately explain E, is equal to the one H would have supposing that neither H nor H' adequately explain E. Both of these claims seem to be compelling: one may be willing to rank P(H|X, Y) strictly higher than P(H|¬X, ¬Y), as X ∧ Y implies that H adequately explains the old evidence E, whereas ¬X ∧ ¬Y implies that H does not adequately explain E; on the other hand, it could be argued that in both suppositions there is no difference between H and H' with respect to explaining E, and so there should not be a difference between the two in terms of probability.

However, Eva and Hartmann (2020) point out that Hartmann and Fitelson's proposal is incomplete, as its conditions (**HF1-HF4**) are jointly sufficient to guarantee P(H|X) > P(H), but they are not sufficient to guarantee P(H|X ∧ ¬Y) > P(H). That is, they allow for the possibility that X ∧ ¬Y can disconfirm H. And this is implausible: learning that H adequately explains E and H's best competitor (H') does not should always make us more confident in the truth of H. In fact, in this situation, I become more confident that H is the only way I can possibly account for the evidence and thereby become more confident that H has to be true.

2.2. The static dimension of POE

In the static dimension of POE, we find ourselves in a moment in time in which belief changes caused by the discovery of H and its relationship to E already happened. However, we want to say why E is and will be a reason to prefer H over its competitors.

The standard approach to the static POE states that the confirmational relation between H and E has to be evaluated relying on a counterfactual degree of belief function where E is not taken for granted (Howson, 1984; 1985; 1991). So, we are giving up the actual degrees of belief in E.

This means that the conditional probabilities, P(E|H) and P(E|¬H), are not equal to 1, as it would turn out if E is taken for granted, as seen before. Thus, P(E|±H) describes the degrees of belief we would have in E if we did not know that E and H were the case. Namely, they describe our degrees of belief in E supposing H and ¬H. This allows a meaningful comparison between the likelihoods to establish confirmation judgements. Let us see why. First of all, notice that the denominator of Bayes' theorem can be neglected as it is the same for all the theories we are considering – remember, we want to say why evidence E confirms H more than other theories. Thus, we have that P(H|E) > P(¬H|E) if and only if P(H)P(E|H)

$> P(\neg H)P(E|\neg H)$. Now, if P is an 'impartial' prior probability distribution, i.e. $P(H) = P(\neg H)$, then $P(H|E) > P(\neg H|E)$ if and only if $P(E|H) > P(E|\neg H)$. Going back to the Mercury perihelion shift's example, we have the latter confirmation judgement. For, $P(E|H) = 1$ because GTR implies Mercury's perihelion shift, whereas $P(E|\neg H) \ll 1$, as Newtonian mechanics and other theories do not make definite predictions about E.

The main flaw of the counterfactual approach is that its novel interpretation of probability raises many philosophical and technical problems. As Eells (1990, pp. 207-208, emphases in the original) says: "Glymour and Garber both point out that there will not necessarily be any *particular* degrees of belief that we can say a person would have had in E, or in H[1] given E, if this person's degree of belief in E had been less than 1. Indeed, it seems plausible that in some cases *H would not even have been formulated had E not been learned*. And surely there also will be cases in which the person's knowledge of E *saved his life* at some time in the past, so that had the individual's degree of belief in E been less than 1, the person would be *dead* now. Also, there are of course the well-known difficulties attending the proper interpretation of counterfactual conditionals that would befall any such modification of Bayesian confirmation theory".

To better understand the two problematic cases Eells mentions to argue that we will not necessarily have any particular degree of belief in E or in H given E, had we not known E, it is useful to make some examples[2].

As for the first case, consider the following example borrowed from Schurz (2008, pp. 209-210). Let us suppose that biologists discover marine fossil records, e.g. fish bones, in the ground of dry land (evidence Ga). To account for this piece of evidence, they introduce hypothesis Ha: 'some geological time span ago, there was a sea here'. To test the hypothesis, they look for further empirical consequences, deducible from Ha and background

[1] Eells refers to hypotheses or, equivalently, theories, by using the capital letter T and not H. However, for consistency, I replaced T with H.

[2] The first problematic case is also considered by Glymour (1980, p. 89), who, however, does not provide any example. Moreover, Glymour (1980, pp. 87-91) makes other arguments in favour of the claim that it is problematic to have *particular* degrees of belief in E, or in H given E, had our degrees of belief in E been less than 1.

assumptions *K*, like, for instance, further geological indications such as calcium deposits, or marine shell fossils, and so on (let us call these pieces of evidence Ea).

As Schurz observes, if the latter findings are observationally verified, the hypothesis is confirmed. Moreover, Schurz points out that if Ea is verified, then both Ga and Ea provide epistemic support to Ha, adding that the initial inference *keeps* its justificatory value. In other words, Ga confirmed Ha from the beginning, and now that Ea is verified, Ga keeps confirming Ha, along with Ea.

Even if Schurz does not provide a Bayesian analysis of the example, from a Bayesian standpoint, we can surely say that: $P(Ha|Ga) > P(Ha)$; and, once Ea is verified, $P(Ha|Ga \wedge Ea) > P(Ha)$. So, in both instances, Ga confirms Ha. However, Ga is old evidence, as it is known before Ha is formulated. That is, we have an old evidence problem.

If we try to solve it by using the counterfactual approach, namely by supposing that we had not known Ga, then we would not even have formulated Ha, because, as the example shows, Ha is advanced to account for Ga.

Let us now consider the second problematic case considered by Eells, i.e., the one concerning the fact that knowledge of E saved the individual's life at some time in the past, so that had the individual's degree of belief in E been less than 1, the person would be dead now.

An example of this circumstance could be the following. Let us suppose that a primitive man sees large footprints on the ground, of a kind he has never seen (evidence E). He conjectures that the footprints were left by a huge and dangerous animal (hypothesis H). Scared for his life, he decides to change path. Later on, he confirms his hypothesis by gaining direct evidence of it: he sees from a distance the animal pass by, and once it is gone, the primitive man checks its footprints.

Schurz (2008, p. 209) considers cases like these, in which the hypothesis, advanced to explain a piece of evidence, is later on confirmed by direct evidence. And, he points out that the weak epistemic support which the initial inference conveys to the conjecture gets fully replaced by the strong epistemic support provided by the direct evidence. This means that, in our example, E confirms H, but only weakly, and, once the direct evidence is found, it is the latter that fully confirms H, and not E. Still, at the beginning, even if only weakly, E confirms H. So, in Bayesian terms we can surely say that $P(H|E) > P(H)$.

For the same considerations as above, E is old evidence. Thus, if, before direct evidence is collected, we want to explain why $P(H|E) > P(H)$ happens in a Bayesian fashion, we have to deal with an old evidence problem.

Again, if we try to solve it by using the counterfactual approach, namely by supposing that we had not known E, then we would not even have formulated H, because, as the example shows, H is advanced to account for E.

3. Eva and Hartmann's two novel solutions to the dynamic POE: the Inference to the Best Explanation's perspective

In a recent article, Eva and Hartmann (2020) propose two novel solutions to the dynamic POE. Their common denominator is the observation that what typically happens in real cases of confirmation by old evidence is that hypothesis H receives strong confirmation by old evidence E because H is the only available hypothesis that adequately explains E. Indeed, both models show that learning such a fact confirms hypothesis H.

Even if Eva and Hartmann never acknowledge this, a natural and intuitive way to express what the two models show is saying that learning that H is the best explanation of E confirms H. That is, the two models should be read in terms of inference to the best explanation. Accordingly, an explicit connection between them and IBE should be made, and the two models should be assessed in light of such a connection.

Making such an assessment is the main purpose of this section (subsection *3.3*). But, in order to do that, first, I will present the two models (subsection *3.1*), and give a brief exposition of what IBE is (subsection *3.2*).

3.1. Eva and Hartmann's two novel solutions

The first model aims to overcome the incompleteness of Hartmann and Fitelson's model (see end of subsection *2.1*), by finding plausible extra conditions, consistent with **HF1-HF4**, under which $X \wedge \neg Y$ does not disconfirm H.

There is no need to go very far, as this condition is a slight strengthening of condition **HF4**, namely **HF4***: $P(H|X, Y) = P(H|\neg X, \neg Y)$. According to Eva and Hartmann, this assumption captures the idea that, typically, hypotheses receive significant confirmation by old evidence only when they are the *only* hypotheses that adequately explain the relevant evidence. It does so – I believe – by telling us that when H's best competitor explains E as well as H does, H's probability is equal to the probability H would receive if neither H nor H' adequately explain E. And this latter probability does not confer a significant confirmation.

Then, they prove that **HF1-HF4*** are jointly sufficient to guarantee: (1) $P(H|X) > P(H)$; (2) $P(H|\neg Y) > P(H)$; (3) $P(H|X \land \neg Y) > P(H)$. Thus we have the desired result: (3).

More specifically, the aforementioned idea is justified by the observation that it is what happens in the real cases of confirmation by old evidence. Indeed, in the Mercury perihelion shift's example, GTR received such a strong confirmation because it adequately explained E, and none of its competitors did. In fact, the competing hypotheses – Le Verrier's unobserved planet 'Vulcan', and Von Seeliger's ring of a particular matter around the sun (e.g. Crelinsten, 2013) – were not serious competitors to GTR when Einstein found out it implied E. If there are competing theories which are also capable of adequately explaining the old evidence, then the degree of confirmation conferred on the hypothesis by the old evidence would intuitively be far weaker, and possibly negligible. For instance, if $H \in S$, where S is a class of mutually incompatible theories, and it is showed that all the theories in S adequately explain the old evidence, this proof will not do much to increase our confidence in H, since it does nothing to distinguish H from its competitors.

The point that has just been highlighted suggests that scientists are primarily concerned with the proposition A: 'H is the only available hypothesis that adequately explains E'. Accordingly, we need to show that the agent increases her confidence in H because she becomes more confident in A, without necessarily becoming certain of the truth of any individual proposition, i.e. 'H adequately explains E' and 'H is the only available hypothesis that adequately explains E'. Formally, by using Jeffrey's conditionalization, we want to show:

$$P^*(H) = P(H|A)P^*(A) + P(H|\neg A)P^*(\neg A) > P(H) \tag{3}$$

In order to do this, Eva and Hartmann use only one minimal constraint **A1**: P(H|A) > P(H|¬A). That is, H is more likely to be true assuming that it is the only available hypothesis that adequately explains the old evidence than assuming that it is not.

From **A1**, it straightforwardly follows that, when the scientist becomes more confident in A, i.e. P*(H) > P(H)[3].

When the scientist learns A for certain, then Jeffrey's conditionalization reduces to strict conditionalization, and we need to prove: P*(H) = P(H|A) > P(H), which, again, straightforwardly follows from **A1**, since the latter is equivalent to P(H|A) > P(H).

Thus, we see that both models show that learning the proposition 'H is the only available hypothesis that adequately explains E' confirms H. This proposition, in the first model, is expressed by X ∧ ¬Y, and, in the second model, by A.

3.2. What is Inference to the Best Explanation?

The core idea of Abduction or, as it is more commonly called nowadays, Inference to the Best Explanation is that explanatory considerations have a confirmational import. Such an idea can be cashed out in a variety of plausible ways. Here, following Douven (2017), I will consider three of them, which are all inference rules, starting with the formulation often encountered in textbooks of epistemology and philosophy of science:

ABD1 Given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, infer the (probable) truth of that H_i which best explains E.

The main problem with **ABD1** is that it does not appear to be normatively adequate as its reliability is based on conditions that are hard to justify. In fact, in order for **ABD1** to be reliable, we need two necessary conditions:

1. In most of the cases, the best explanation relative to the hypotheses we have considered must also be the best relative to the hypotheses we might have conceived.

[3] For the proof, see Eva and Hartmann (2020), footnote n. 4, p. 491

That is, the best *absolute* explanation of the evidence has to be among the candidate hypotheses we have come up with. Otherwise, **ABD1** would lead us to consider probably true, and so to believe, "the best of a bad lot" (van Fraassen, 1989, p. 143).

2. In most of the cases in which the best explanation of the evidence is also the best absolute explanation, the best explanation is probably true.

However, 1 can be fulfilled only when we assume a predisposition of the agent to hit the best absolute explanation among the ones she has considered. But, as van Fraassen points out (*ibid*., p. 144), it is *a priori* implausible to suppose we have such a form of privilege.

The most promising response to the 'argument of the bad lot' points out that the rule is asymmetric (e.g. Kuipers, 2000, p. 171). Namely, it has an absolute conclusion – the hypothesis is probably true – on the basis of a comparative premise – the best explanation of the data is relative to the available explanations of the data. This discrepancy can be avoided in two ways: either by making the premise absolute as well, or by making the conclusion comparative.

According to the first path, the probable truth of the best explanation is not to be inferred only when the latter is the best explanation with respect to the candidate explanations, but also when it is a satisfactory (Musgrave, 1988) or good enough (Lipton, 1993) explanation. Thus:

ABD2 Given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, infer the (probable) truth of that H_i which best explains E, provided H_i is satisfactory/good enough *qua* explanation.

The main problem with **ABD2** is that it relies on concepts, such as the satisfactoriness of an explanation or its being good enough, of which we lack a full understanding.

Conversely, as announced before, the second option derives a comparative conclusion from a comparative premise:

ABD3 Given evidence E and candidate explanations $H_1, \ldots, H_n$ of E, if H_i explains E better than any of the other hypotheses, infer that H_i is closer to the truth than any of the other hypotheses.

ABD3 requires, instead, an account of closeness to the truth. But many accounts of this kind are available today (e.g. Niiniluoto, 1998).

The bright side of the latter two definitions is that, despite the shortcomings, their reliability is not based on an implausible form of privilege as **ABD1**'s reliability.

3.3. Weaknesses and strengths of the two models from IBE's point of view

Armed with what we have just said in the previous two sections, it is now time to make explicit the connection between IBE and Eva and Hartmann's two models, and to assess the latter in light of such a connection.

In order to do that, the first thing we need to understand is which of the three formulations of IBE considered in subsection *3.2*, if any, originates from Eva and Hartmann's discussion.

As pointed out in subsection *3.1*, Eva and Hartmann's two models rely on a key idea, namely:

E&H Idea In the real cases of confirmation by old evidence, hypothesis H receives strong confirmation by old evidence E, because H is the only available hypothesis that adequately explains E.

At a closer look, it can be seen that **E&H Idea** is endorsed by the two authors both in terms of *confirmation as firmness*, and in terms of *confirmation as increase in firmness*[4].

Let us start by seeing in which sense **E&H Idea** is expressed in terms of confirmation as firmness.

[4] To expand on the distinction between these two senses of confirmation, see Hartmann and Sprenger 2019, variation 1.

In the literature (e.g., Hartmann and Sprenger, 2019, variation 1), confirmation as firmness is defined as follows:

Confirmation as Firmness Evidence E confirms hypothesis H if and only if $P(H|E) \geq t$ for some – possibly context-dependent – $t \in [0,1]$.

Thus, **E&H Idea** in terms of confirmation as firmness would say:

learning that H is the only available hypothesis that adequately explains E (e.g., $X \wedge \neg Y$) strongly confirms H because $P(H|X \wedge \neg Y) = t$, where t is a high value[5].

This is how **E&H Idea** is captured by condition **HF4***: $P(H|X,Y) = P(H|\neg X,\neg Y)$. Here is why. As we have seen in subsection *3.1*, condition **HF4*** captures **E&H Idea** by, presumably, telling us the following: when H's best competitor explains E as well as H does, H's probability is equal to the probability H would receive if neither H nor H' adequately explain E; and the latter probability does not confer a significant degree of confirmation.
Thus, ultimately, **HF4*** captures **E&H Idea** by saying that learning $X \wedge Y$ does not strongly confirm H, i.e., $P(H|X,Y) = t$, where t is not a significantly high number. In other words, $X \wedge Y$ does not strongly confirm H according to the notion of confirmation as firmness. So, when **E&H Idea** is captured by **HF4***, the sense of confirmation in **E&H Idea** is expressed in terms of confirmation as firmness. That is, $P(H| X \wedge Y) = t$, where t is high number.
As for **E&H Idea** in terms of confirmation as increase in firmness, it is just what the two models show. More precisely, in the literature (e.g., Hartmann and Sprenger, 2019, variation 1), the concept of confirmation as increase in firmness is defined by saying:

Confirmation as increase in firmness Evidence E confirms hypothesis H if and only if $P(H|E) > P(H)$.

[5] I say that t is a high value because in **E&H Idea**, it is said that H receives *strong confirmation*.

Accordingly, **E&H Idea** in terms of confirmation as increase in firmness would say:

learning that H is the only available hypothesis that adequately explains E ($X \wedge \neg Y \equiv A$) strongly confirms H because $P(H|X \wedge \neg Y \equiv A) > P(H)$.

As we have seen in subsection *3.1*, $P(H|X \wedge \neg Y) > P(H)$ and $P(H|A) > P(H)$ is what Eva and Hartmann's two models respectively show.

Now, the three formulations of IBE assign to explanatory considerations firm confirmation judgements (see subsection *3.2*): the best explanation (and good enough explanation) *is probably true* (**ABD1**, **ABD2**); the best explanation *is closer to the truth* than any of the other hypotheses (**ABD3**). Thus, the best way to see if one of these formulations is implicitly endorsed by Eva and Hartmann is to stick to **E&H Idea** in terms of confirmation as firmness.

E&H Idea in terms of confirmation as firmness can be read as saying that if H is the best explanation of E (where best explanation of E means the only available hypothesis that adequately explains E), then its probability is equal to a high value. From now on, I will refer to this inference rule as 'Eva and Hartmann's IBE'.

At first glance, Eva and Hartmann's IBE is very similar to **ABD1**: its premise is comparative (H is the best explanation of E among the explanations we have so far considered, since H is the best explanation of E in that it is the only *available* hypothesis that adequately explains E); and its conclusion is absolute.

However, such a similarity holds only if 'best explanation of E' can be understood as 'the only available hypothesis that adequately explains E'.

It is often argued that the best explanation is a hypothesis that does best, on balance, on epistemic virtues like simplicity, generality, coherence with well-established theories, and so on (e.g. Thagard, 1978; McMullin, 1996). But, since Eva and Hartmann do not give an account of 'adequate explanation', it is unclear how to conciliate their understanding of best explanation with the common one.

That said, the concept of best explanation does not have a straightforward interpretation (Douven, 2017, section 2). Thus, I believe that there is no harm in interpreting it as intuitively meaning the only available hypothesis that adequately explains E, as Eva and Hartmann implicitly do, and therefore to state a similarity between Eva and Hartmann's IBE and **ABD1**. Thus, summing up, the explicit connection between Eva and Hartmann's novel contributions and IBE consists of two points:

1. Eva and Hartmann's IBE: learning that H is the best explanation of E confers to H a high probability.
2. In real cases of confirmation by old evidence, learning that H is the best explanation of E increases H's probability.

Let us now see the assessment of the two models in light of such an explicit connection. As highlighted just now, they seem to rely on an inference rule which is very similar to **ABD1**. However, so far, it is still an open question which of the three formulations of IBE is descriptively used, or if some further rule is used or whether some version is used in some contexts and another version in others (Douven, 2017, section 2). So, a remark that can be made to Eva and Hartmann's two models, as well as to any philosophical work which assumes a particular version of IBE, is that there is an empirical descriptive question that has yet to be answered: on which version of IBE do scientists rely?

As for the first model alone, it seems to me that condition **HF4*** is not expression of IBE's core idea that explanatory considerations have a confirmational import. Let us see why.

Before, we have seen that Eva and Hartmann's IBE is very similar to **ABD1**, in that it can be read as saying: if H is the best explanation of E in the sense that it is the only hypothesis that adequately explains E, then H receives a high probability. So, such a high probability is given to H because: (i) it is an adequate explanation itself; (ii) it is the only adequate explanation of E.

Now, the first supposition of condition **HF4***, i.e. $X \wedge Y$, tells us that H is an adequate explanation of E itself, but that it is not the only adequate explanation of E, as H' is an explanation as adequate as H. The second supposition of **HF4***, i.e. $\neg X \wedge \neg Y$, on the other hand, tells us that H is not an adequate explanation at all, and so it cannot be the only adequate explanation of E. Thus, in the first supposition, we have one of the two reasons why H receives a high probability, whereas, in the second one, none of them.

Consequently, the probability of H given the first supposition should be higher than its probability given the second one, namely: $P(H|X, Y) > P(H|\neg X, \neg Y)$.

This is not to say that when H's best competitor explains E as well as H does, H receives significant confirmation, but that the **E&H Idea** is not captured by condition **HF4***.

Regarding the second model, the situation is different. In fact, condition **A1**: $P(H|A) > P(H|\neg A)$ is expression of IBE's core idea. Indeed, Eva and Hartmann's IBE implies that when H is the best explanation of E, then H receives a high probability. Conversely, when H is not, it does not receive such a high probability.

Therefore, we can conclude that, even if Eva and Hartmann's IBE turns out to be the one used by scientists in the contexts of confirmation by old evidence, only their second model is appropriate to model them, as the crucial condition of their first one does not capture Eva and Hartmann's IBE.

4. Bayesian IBE and the Inference to the Best Explanation's perspective on the static dimension of POE

By focusing on the origin of confirmation by old evidence, i.e. on how H is confirmed at the moment in which H and its relationship to E are discovered, Eva and Hartmann tackle only the dynamic dimension of POE. However, the reading in IBE's terms of Eva and Hartmann's two novel solutions to the dynamic POE implies that the static POE should also be read in IBE's terms. This is because the IBE's character of the dynamic dimension is inherited by the static dimension: if, at the moment in which H and its relationship to E are discovered, we want to explain why learning that H is the best explanation enhances H's probability, then, after that discovery, we want to explain why E confirms the best explanation more than the other hypotheses that are not the best explanations. Accordingly, the counter-factual approach should be assessed from an IBE's perspective.

But, before expanding on this point, let us consider a particular version of IBE, i.e. the Bayesian IBE, which will come in handy to solve the static POE from the IBE's perspective. More specifically, the latter is the product of a response to an incompatibility between IBE and Bayesianism claimed by some philosophers.

4.1. Incompatibility between Bayesianism and IBE, and Bayesian IBE

The confirmational role that IBE assigns to explanatory considerations (see subsection *3.2*), directly suggests a comparison with Bayesian confirmation theory, the dominant view on confirmation.

In this regard, some philosophers have stated an incompatibility between IBE and Bayesianism. For example, Salmon (2001) stresses that Bayesian confirmation theory is not guided by the concept of explanation. In fact, he concedes that the prior probability in Bayes' theorem can be identified with the goodness, or "loveliness" (Lipton, 2001, p. 105), of an explanation, that is with the degree of understanding the explanation at hand provides. But, such a loveliness is a consequence of the prior probability of the hypothesis and not the other way around.

The prior probability of the hypothesis is determined by considering how the latter relates to our background knowledge. This can contain all sorts of information: theories known at the time, frequencies of the data, the evidential record available. On this basis, one can determine the prior probability of the hypothesis considering, for example, its external consistency, or its *ad hoc* or *non-ad hoc* character.

It may well be that such epistemic virtues – i.e. features of the theories that enhance the theories' probability of being true or accepted – coincide with explanatory virtues – i.e. features of theories that enhance the degree of understanding the theories provide. But the prior probability of the hypotheses is evaluated relying on the epistemic character of the virtues, and not on their explanatory character. According to Salmon, we do not say that a given hypothesis deserves a low or high prior probability because it is a bad or good explanation in that it scores badly or well on the aforementioned virtues. Rather, we say that a given hypothesis deserves a low or high prior probability because it scores badly or well on the aforementioned virtues. And that is it.

Moreover, Salmon continues, the prior probability is not enough to make a choice among hypotheses. We need the posterior probability of the hypotheses, of which prior probability is just a component, as Bayes' theorem tells us. At least, he says, this is what scientists implicitly do when they choose their hypotheses.

On the normative side, there is van Fraassen's criticism (1989), according to which a Bayesian agent that uses IBE as a rule, is liable to diachronic Dutch Book (Teller, 1973). In fact, such an agent adopts an explicit strategy that consists in adding bonus points to the hypotheses that explain the evidence particularly well *after* conditionalization. On this basis then, the bookie can construct a series of bets that leave the Bayesian agent with a certain loss.

Reactions to these criticisms have come respectively from Lipton (2001; 2004) and Okasha (2000). They both endorse a 'Bayesian IBE', namely a probabilistic version of IBE, in which explanatory considerations may act as heuristics, which help to determine, even if roughly, the probabilities in Bayes' theorem.

More precisely, *contra* Salmon, Lipton (2001) makes different claims, all of which are directed to support his heuristic endeavour which he expresses by saying that "the Bayesian and the Explanationist should be friends" (p. 94).

One step in this direction is to show that the epistemic character of the virtues used to estimate the prior probability of the theories is a symptom of their explanatory character. This is what Lipton calls "the guiding challenge" (pp. 107-109), which he resolves by saying that the best explanation of the match between epistemic and explanatory virtues is that scientists select hypotheses on the basis of their explanatory virtues.

The second step consists of arguing that the loveliness is not related solely to the prior probability, but to all the components of Bayes' theorem. For example, explanatory considerations might help to evaluate the likelihoods because lovelier explanations tend to make what they explain likelier. Moreover, they could help to determine the priors in different ways. Firstly, priors are generally determined by earlier conditionalization, where the assessment of the likelihood is essential. But, as pointed out just now, such an assessment might be aided by explanatory considerations. Secondly, explanatory virtues such as scope, mechanism, precision, unification, simplicity[6], could be used by the Bayesian to estimate the

[6] Lipton (2001, p. 106) explains why each of these virtues is an explanatory virtue by saying, respectively: "better explanations explain more types of phenomena, explain them with greater precision, provide more information about underlying mechanisms, unify apparently disparate phenomena, or simplify our overall picture of the world".

prior probability, taking for granted the success of the guiding challenge. Finally, explanatory considerations might help to determine why certain bits of evidence, and not others, are involved in the Bayesian process of conditionalization. In fact, we can come to see that a datum is relevant for the hypothesis precisely because the hypothesis would explain that datum.

Contra van Fraassen, Okasha (2000), instead, highlights that van Fraassen's way to represent IBE in probabilistic terms is idiosyncratic. For, it does not capture the phenomenology of IBE where there is no hint of a two-stage process. We do not first respond to the evidence, and take explanatory considerations into account at a later stage. Rather, we use explanatory considerations to decide how to respond to the evidence – just one process.

This suggests that the best way to represent IBE in probabilistic terms is to use explanatory considerations in the process that realizes conditionalization. That is, the better the explanation, the higher its prior and/or likelihood, and, in any case, given the same body of evidence E, the better explanation of E will end up having the higher product of prior and likelihood.

Indeed, like Lipton, Okasha underlines that explanatory considerations help to determine the likelihoods because better explanatory hypotheses tend to give a higher probability to the evidence. Contrastingly to Lipton, and agreeing with Salmon, Okasha argues that explanatory considerations help to determine the prior probability in the sense that the goodness of an explanation is a consequence of its plausibility, i.e. of its prior probability.

4.2. The IBE's perspective on the static dimension of POE

As seen in subsection *2.2*, in the static dimension of POE, we find ourselves in a moment in time in which belief changes caused by the discovery of H and its relationship to E already happened. Still, we want to say why E is and will be a reason to prefer H over its competitors. In IBE terms, we want to explain why E confirms the best explanation H more than the other theories that are not best explanations of E. Indeed, this is what happens: (i) if Eva and Hartmann's IBE is descriptively accurate; (ii) given IBE's "self-evidencing character"

(Lipton, 2001, p. 96). In fact, according to (i) when H is the best explanation of E in the sense that it is the only hypothesis that adequately explains E, then it receives a high probability. Conversely, the other explanations do not receive such a high probability. And according to (ii), the datum explained by the hypothesis, in turn, confirms the hypothesis precisely because it is explained by the hypothesis.

In subsection *2.2*, we also saw that the standard approach to the static POE urges us to give up the actual degrees of belief in E so to allow a meaningful comparison between the likelihoods to establish confirmation judgements. That is, $P(H|E) > P(\neg H|E)$ if and only if $P(H)P(E|H) > P(\neg H)P(E|\neg H)$.

As we have seen in subsection *4.1*, the heuristic conciliatory approaches prove that better explanations are the ones with higher priors and/or likelihoods, and, in any case, the ones with the higher product of these two quantities. Thus, we have $P(H|E) > P(\neg H|E)$, i.e. E confirms the best explanation H more than the competing theories that are not the best explanations of H.

That said, the problems connected with the counterfactual interpretation of probability are still present when the aforementioned approach is used to solve the static POE in IBE terms. For instance, in the Mercury perihelion shift's example, it is unlikely that we could say which our degree of belief in ¬H given E would have been, if we had not known E.

Indeed, Le Verrier's unobserved planet 'Vulcan', and Von Seeliger's ring of a particular matter around the sun were generated to account for the anomalous shift (see Crelinsten, 2013, pp. 51-54). That is, these hypotheses would not have been formulated had E not been known. The same cannot be said for Einstein's GTR. The latter, in fact, was not formulated to explain the anomalous shift, and derivation of the shift from GTR was surprising in that it was not expected beforehand (see subsection *2.1*). Still, we would lack what we want to explain, i.e. $P(H|E) > P(\neg H|E)$, as we would miss the term of comparison $P(\neg H|E)$.

5. Conclusion

In the foregoing, I have made an explicit connection between Eva and Hartmann's two novel solutions to the dynamic POE and the inference to the best explanation.

This has allowed me to evince that Eva and Hartmann have in mind a specific understanding of IBE's core idea, which is very similar, although not equal, to **ABD1**, which I labelled Eva and Hartmann's IBE. On this basis, I have assessed the two models. Specifically, I have shown that – taking for granted the open question regarding if Eva and Hartmann's IBE is descriptively accurate – their first model is not adequate to solve the dynamic POE in IBE's terms, while the second one is. The reason is that the crucial condition of the former, **HF4***, is not expression of Eva and Hartmann's IBE. Conversely, the only condition of the latter, **A1**, is.

Finally, I have pointed out that the explicit realization that real cases of confirmation by old evidence are instances of IBE implies that the static dimension of POE has to be, now, expressed in IBE terms. I have attempted to solve the static dimension of POE so expressed by remaining inside the frame of the standard counterfactual approach. This has been possible by using the results of the heuristic conciliatory approaches which show that better explanations have higher prior and/or likelihood, and, in any case, that they are the ones with the higher product of these two quantities. However, I have stressed that the problems related to the counterfactual interpretation of probability are still present when the counterfactual approach is used to solve the static POE in IBE terms. For instance, I have pointed out that in the GTR's example, if the old evidence E had not been known, we would have lost what we wanted to explain.

References

Brush, S. G. (1989). Prediction and theory evaluation: The case of light bending. *Science*, *246*(4934), 1124-1129. https://doi.org/10.1126/science.246.4934.1124

Crelinsten, J. (2013). *Einstein's jury*. Princeton University Press.

Douven, I. (2017). Abduction. In E. N. Zalta (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). Stanford University. <https://plato.stanford.edu/archives/sum2017/entries/abduction/>

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press.

Eells, E. (1985). Problems of old evidence. *Pacific Philosophical Quarterly*, *66*(3-4), 283-302. https://doi.org/10.1111/j.1468-0114.1985.tb00254.x

Eells, E. (1990). Bayesian problems of old evidence. In C. W. Savage (Eds.), *Scientific Theories* (pp. 205–223). University of Minnesota Press.

Eva, B., & Hartmann, S. (2020). On the origins of old evidence. *Australasian Journal of Philosophy*, *98*(3), 481-494. https://doi.org/10.1080/00048402.2019.1658210

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In J. Earman (Eds.), *Testing Scientific Theories* (pp. 99–132). University of Minnesota Press.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press.

Hartmann, S., & Fitelson, B. (2015). A new Garber-style solution to the problem of old evidence. *Philosophy of Science*, *82*(4), 712-717. https://doi.org/10.1086/682916

Hartmann, S., & Sprenger, J. (2019). *Bayesian philosophy of science*. Oxford University Press.

Howson, C. (1984). Bayesianism and support by novel facts. *The British journal for the Philosophy of Science*, *35*(3), 245-251. https://doi.org/10.1093/bjps/35.3.245

Howson, C. (1985). Some recent objections to the Bayesian theory of support. *The British journal for the philosophy of science*, *36*(3), 305-309. https://doi.org/10.1093/bjps/36.3.305

Howson, C. (1991). The 'old evidence' problem. *The British Journal for the Philosophy of Science*, *42*(4), 547-555. https://doi.org/10.1093/bjps/42.4.547

Jeffrey, R. C. (1983). Bayesianism with a human face. In J. Earman (Eds.), *Testing Scientific Theories* (pp. 133–156). University of Minnesota Press.

Kuipers T. A. F. (2000). *From Instrumentalism to Constructive Realism*. 2000. Kluwer.

Lipton, P. (1993). Is the best good enough?. In *Proceedings of the Aristotelian Society* (Vol. 93, pp. 89-104). Aristotelian Society, Wiley.

Lipton, P. (2001). Is explanation a guide to inference? A reply to Wesley C. Salmon. In G. Hon & S. S. Rakover (Eds.), *Explanation* (pp. 93-120). Kluwer Academic.

Lipton, P. (2004). *Inference to the best explanation*. Routledge.

McMullin, E. (1996). Epistemic Virtue and Theory Appraisal. In I. Douven and L. Horsten (Eds.), *Realism in the Sciences* (pp. 13–34). Leuven University Press.

Musgrave, A. (1988). The ultimate argument for scientific realism. In R. Nola (Eds.), *Relativism and realism in science* (pp. 229–252) Kluwer.

Niiniluoto, I. (1983). Novel facts and Bayesianism. *The British Journal for the Philosophy of Science*, *34*(4), 375-379. https://doi.org/10.1093/bjps/34.4.375

Niiniluoto, I. (1998). Verisimilitude: The third period. *The British Journal for the Philosophy of Science*, *49*(1), 1-29. https://doi.org/10.1093/bjps/49.1.1

Okasha, S. (2000). Van Fraassen's critique of inference to the best explanation. *Studies in History and Philosophy of Science*, *31*(4), 691-710. https://doi.org/10.1016/S0039-3681(00)00016-9

Salmon, W. C. (2001). Explanation and confirmation: A Bayesian critique of inference to the best explanation. In G. Hon & S. S. Rakover (Eds.), *Explanation* (pp. 61-91). Kluwer Academic.

Schurz, G. (2008). Patterns of Abduction", *Synthese,* 164, 201-234. https://doi.org/10.1007/s11229-007-9223-4

Teller, P. (1973). Conditionalization and observation. *Synthese*, *26*(2), 218-258. https://doi.org/10.1007/BF00873264

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

Van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford University Press.