A Novel Solution to the Problem of Old Evidence

Jan Sprenger*†

One of the most troubling and persistent challenges for Bayesian Confirmation Theory is the Problem of Old Evidence (POE). The problem arises for anyone who models scientific reasoning by means of Bayesian Conditionalization. This article addresses the problem as follows: First, I clarify the nature and varieties of the POE and analyze various solution proposals in the literature. Second, I present a novel solution that combines previous attempts while making weaker and more plausible assumptions. Third and last, I summarize my findings and put them into the context of the general debate about POE and Bayesian reasoning.

1. Introduction: The Problem of Old Evidence. One of the most troubling and persistent challenges for Bayesian Confirmation Theory (BCT) is the Problem of Old Evidence (POE): A phenomenon E remains unexplained by the available scientific theories. At some point, a theory T is discovered that accounts for E. Then, E is "old evidence": at the time when T is developed, the scientist is already certain or close to certain that the phenomenon E is real. Nevertheless, E apparently confirms T—at least if T was invented on independent grounds. After all, it resolves a well-known and persistent tension between theory and observation.

A famous case of old evidence in science is the Mercury perihelion anomaly (Glymour 1980; Earman 1992). For a long time, the shift of the Mercury

perihelion could not be explained by Newtonian mechanics or any other reputable physical theory. Then, Einstein realized that his General Theory of Relativity (GTR) could explain the perihelion shift. This discovery conferred a substantial degree of confirmation on GTR, much more than some pieces of novel evidence.

Also in other scientific disciplines, newly introduced theories are commonly assessed with respect to their success at accounting for observational anomalies. Think, for example, of the assessment of global climate models against a track record of historical data or of economic theories that try to explain anomalies in decision making under uncertainty (e.g., the Allais or Ellsberg paradoxes).

All this is hard to capture in the Bayesian framework, where confirmation is expressed as an increase in an agent's subjective degree of belief. On the Bayesian account, E confirms T if and only if the posterior degree of belief in T, $p'(T)$, exceeds the prior degree of belief in T, $p(T)$. These two probabilities are related by means of conditioning on the evidence E and Bayes's Theorem (e.g., Howson and Urbach 2006):

$$p'(T) := p(T \mid E) = p(T)\frac{p(E \mid T)}{p(E)}. \tag{1}$$

Here and in the sequel, reference to an accepted body of background assumptions K in the credence function $p(\cdot)$ is omitted for the sake of simplicity.

Let us now apply the Bayesian calculus to POE. When E is old evidence and already known to the scientist, the prior degree of belief in E is maximal: $p(E) = 1$. But with that assumption, it follows that the posterior probability of T cannot be greater than the prior probability: $p'(T) = p(T \mid E) = p(T) \times p(E \mid T) \leq p(T)$. Hence, E does not confirm T. The very idea of confirmation by old evidence, or equivalently, confirmation by accounting for well-known observational anomalies, seems impossible to describe in the Bayesian belief kinematics.

This article investigates the variety of the problems that POE poses for BCT, reviews existing approaches, and finally proposes a novel solution. Section 2 distinguishes the dynamic and the static version of POE and briefly comments on attempts to solve the static problem. Section 3 analyzes the solutions of the dynamic version proposed by Garber (1983), Jeffrey (1983), Niiniluoto (1983), and Earman (1992). On these accounts, confirmation occurs through conditionalizing on the proposition $T \vdash E$. Section 4 presents my own improvement on these attempts. Finally, section 5 puts my findings into the context of the general debate about POE and BCT.

Throughout the article, I work in the framework of Bayesian epistemology (Bovens and Hartmann 2003; Hájek and Hartmann 2010; Hartmann and Sprenger 2010) and Bayesian philosophy of science (Howson and Urbach 2006; Sprenger 2015). I am using Bayesian Nets for representing depen-

dencies between propositions of interest—a technique that has recently proved its merits for modeling complex confirmation phenomena (e.g., Dawid, Hartmann, and Sprenger 2015).

2. The Varieties of the POE. Part of the controversy about the POE concerns the question of what the problem really consists in. Eells (1985, 1990) has introduced a helpful conceptual distinction between different problems of old evidence:

1. *The Problem of Old New Evidence*: E is learned after T is formulated, but even after updating our degrees of belief on E, we still say that E confirms T although by now, $p'(E) = 1$. (As before, p' denotes the posterior probability distribution.)
2. *The Problem of Old Evidence*: E is known before T is formulated.
   2a) *The Problem of Old Old Evidence*: Even after formulating T and discovering that T accounts for E, E seems to confirm T.
   2b) *The Problem of New Old Evidence*: Why does E confirm T at the moment where T is discovered or when it is discovered that T accounts for E?

Items 1 and 2a describe the *static* (Eells: "ahistorical") aspect of POE: belief changes induced by the discovery of T, or the fact that T accounts for E, have already taken place. Still, we would like to say that E is evidentially relevant for T: when faced with a decision between T and a competitor T', E is a good reason for preferring T over T'. Item 2b, however, captures the *dynamic* (Eells: "historical") aspect of the problem: it refers to the moment in time when T and its relation to E are discovered. Here, the challenge for the Bayesian is to describe how the discovery of a new theory T and its explanatory successes raises our confidence in T.

The Bayesian usually approaches both problems by different strategies. The standard take on the dynamic problem consists in allowing for the learning of logical truths. In classical examples, such as explaining the Mercury perihelion shift, the newly invented theory (here: GTR) was initially not known to entail the old evidence. It took Einstein some time to find out that T entailed E (Brush 1989; Earman 1992). Learning this deductive relationship undoubtedly increased Einstein's confidence in T since such a strong consilience with the phenomena could not be expected beforehand.

However, this belief change is hard to model in BCT. A Bayesian reasoner is assumed to be logically omniscient, and the logical fact $T \vdash E$ should always have been known to her. Hence, the proposition $T \vdash E$ cannot be learned by a Bayesian: it is already part of her background beliefs.

To solve this problem, several philosophers have relaxed the assumption of logical omniscience and enriched the algebra of statements about which we have degrees of belief. New atomic sentences of the form $T \vdash E$ are

added (Garber 1983; Jeffrey 1983; Niiniluoto 1983), such that BCT can account for our cognitive limitations in deductive reasoning. Then, it can be shown that under suitable assumptions, conditioning on $T \vdash E$ confirms T. I comment on these efforts in the next section.

The response to the static problem is different. Eells (1990, 209) proposes that "E is (*actual*) *evidence for T* . . . if, at some point in the past, the event of its confirming T took place." On that definition, the solution of the static problem in 1 and 2*a* would reduce to the solution of the dynamic problem (cf. Christensen 1999, 444): when we can show that at the time of the formulation of T, E confirmed T, then E would also be evidence for T afterward (Eells 1990, 210). However, most takes on the dynamic problem do not try to show that E confirms T at the relevant moment in time: rather, this work is done by the proposition $T \vdash E$. This strategy also fails to account for changes over time in our assessment of the strength of the evidence that E provides for T.

Therefore, Colin Howson (1984, 1985, 1991) developed a more principled take on the static problem. He gives up the Bayesian explication of confirmation as positive probabilistic relevance relative to actual degrees of belief. Rather, he suggests to subtract the old evidence E from the agent's background knowledge K and to evaluate the confirmation relation with respect to the counterfactual credence function $p_{K \setminus \{E\}}$: "The Bayesian assesses the contemporary support E gives H by how much the agent would change his odds on H *were he now* to come to know E. . . . In other words, the theory is explicitly a theory of dispositional properties of the agent's belief-structure" (Howson 1984, 246). For instance, if T is a hypothesis about the bias of a coin and E the outcome of some series of coin tosses, then eliminating E from the background knowledge allows us to describe how E raises the probability of T since we would have definite and nontrivial values for $p_{K \setminus \{E\}}(E \mid T)$ and $p_{K \setminus \{E\}}(E \mid \neg T)$ (Howson 1991, 551–52).

As the above quote indicates, Howson thinks that BCT is essentially a counterfactual or dispositional theory. I happen to agree with him on this point, but there are a couple of technical problems with Howson's specific choice of a counterfactual credence function. In particular, E may be entangled with other propositions that are part of K. As Chihara (1987) notes, just removing E from the set of background assumptions K will not work in general if K is supposed to be a deductively closed set. If we ignore this feature, we would sacrifice a main advantage of Howson's counterfactual approach: that it need not give up the elegance of the standard Bayesian rationality assumptions (e.g., epistemic closure) in order to account for the POE.

Alternatively, Howson may choose to evaluate the confirmation relation with respect to the agent's credence function at some point in the past (e.g., just before she learned E). But the agent's belief in E may have gradually grown over time, and no such time point may exist. Moreover, without

knowledge of E, T may even not have been formulated (cf. Glymour 1980, 87–93).

All in all, the credence function $p_{K \setminus \{E\}}$ that is supposed to ground the confirmation relation between T and E is rather indefinite even in cases where everybody agrees that E confirms T. As a consequence, we can hardly determine whether the degree of confirmation conferred on T by E is strong, weak, or nil.

Another solution proposal for the static problem is based on the choice of a specific confirmation measure (Fitelson 1999, 2014). Christensen (1999) contends that for the measure $s^*(T, E) = p(T \mid E) - p(T \mid \neg E)$, the POE does not arise. For instance, if T entails E, then $\neg E$ also entails $\neg T$, which implies $p(T \mid \neg E) = 0$ and $s^*(T, E) = p(T \mid E) > 0$. According to s^*, old evidence E can substantially confirm theory T, whereas the degree of confirmation is 0 for measures that compare the prior and posterior probability of T, such as $d(T, E) = p(T \mid E) - p(T)$ or $r(T, E) = \log(p(T \mid E)/p(T))$.

Christensen's move has its merits for cases where $p(E)$ is close to, but not entirely equal to, 1 (although it is questionable whether s^* is a good explicatum for the degree of confirmation; see Eells and Fitelson 2000). But in the classical POE where $p(E) = 1$, $p(T \mid \neg E)$ may not have a clear-cut definition since $p(T \mid \neg E) = p(T \wedge \neg E)/p(\neg E)$ involves a division by 0. And if $p(T \mid \neg E)$ has to be evaluated relative to a counterfactual credence function, not much may have been gained with respect to Howson's proposal.

The static POE thus proves to be a tough challenge for Bayesian confirmation theorists. Are prospects any better for the dynamic problem?

3. The GJN Approach to POE.
The first models of the dynamic POE have been developed by Daniel Garber, Richard Jeffrey, and Ilkka Niiniluoto in a group of papers that all appeared in 1983. Henceforth, we refer to the family of their solution proposals as GJN solutions. In order to properly compare my own solution proposal to the state of the art, and to assess its innovative value, I will briefly recap the achievements of the GJN models and elaborate their limitations and weaknesses.

The GJN models take into account that a scientist is typically not aware of all possible theories and their relations to the evidence, thus parting with the assumption of logical omniscience that characterizes the ideal Bayesian reasoner. Consequently, the relevant piece of evidence is not E itself but the learning of a specific relation between theory and evidence, namely, that T implies E or accounts for E. The notational convention to write this proposition as $\ulcorner T \vdash E \urcorner$ conceals that we do not necessarily deal with a strict logical deduction—also explanatory relationships may fall under the scope of this model (Garber 1983, 103; Eells 1990, 212). Such cases count, after all, as confirmatory arguments for T in ordinary scientific reasoning. How-

ever, thinking of deduction facilitates notation and can be used as a guide for intuition and the development of an adequate Bayesian model.

What the GJN models aim to show is that conditionalizing on the proposition $T \vdash E$ increases the posterior probability of T. Eells (1990, 211) distinguishes three steps in this endeavor: first, parting with the logical omniscience assumption and developing a formal framework for imperfect Bayesian reasoning; second, describing which kind of relation obtains between T and E; and third, showing that learning this relation increases the probability of T. While the GJN models neglect the second step, probably in due anticipation of the diversity of logical and explanatory relations in science, they are quite explicit on step 1 and step 3.

Garber's (1983) model focuses on step 1 and on learning logical truths in a Bayesian framework. After all, no reasoner is ever logically omniscient. Learning logical/mathematical truths can be quite insightful and lead to great progress in science. The famous, incredibly complex proof of Fermat's Last Theorem may be a good example (see Wiles [1995] for a short version). Garber therefore enriches the underlying language L in a way that $T \vdash E$ is one of the atomic propositions of the extended language L'.

Garber also demands that the agent recognize some elementary relations in which the proposition $T \vdash E$ stands to other elements of L':

$$p(E \mid T,\ T \vdash E) = 1 \qquad p(T,\ E,\ T \vdash E) = p(T,\ T \vdash E). \qquad (2)$$

These constraints express the closure of degree of belief under modus ponens. Or alternatively, if an agent takes T and $T \vdash E$ for granted, then she maximally believes E. Even in a boundedly rational picture of Bayesian reasoning, such as Garber's, knowledge of such elementary inference schemes sounds eminently sensible. Garber then proves the following theorem: there is at least one probability function on L' such that every nontrivial atomic sentence of the form $T \vdash E$ gets a value strictly between 0 and 1. Thus, one can coherently have a genuinely uncertain attitude about all propositions in the logical universe, including tautologies. Finally Garber shows that for any atomic L'-sentence of the form $\ulcorner T \vdash E \urcorner$, there are infinitely many probability functions such that $p(E) = 1$ and $p(T \mid T \vdash E) > p(T)$. A similar point is made by Niiniluoto (1983), although with less formal rigor and elaboration.

While Garber's efforts are admirable, they only address the first step of solving the dynamic POE: he provides an *existence proof* for a solution to the POE but no set of conditions that guide our judgments on when learning $T \vdash E$ confirms T. This lacuna is closed by Richard Jeffrey (1983), who published his solution in the same volume in which Garber's paper appeared.

Jeffrey considers the proposition $T \vdash E$ as an object of subjective uncertainty. We follow Earman's (1992) presentation of Jeffrey's (1983, 150–51) solution, which contains the following assumptions:

α. $p(E) = 1$.
β. $p(T)$, $p(T \vdash E)$, $p(T \vdash \neg E) \in (0, \ 1)$.
γ. $p([T \vdash E] \wedge [T \vdash \neg E]) = 0$.
δ. $p(T \mid [T \vdash E] \vee [T \vdash \neg E]) \geq p(T)$.
η. $p(T, \neg E, T \vdash \neg E) = p(T, \ T \vdash \neg E)$.

From these assumptions, Jeffrey derives $p(T \mid T \vdash E) > p(T)$, which amounts, given the constraint $p(E) = 1$, to a solution of the dynamic POE.

The strength of Jeffrey's solution crucially depends on how well we can motivate condition δ. The other conditions are plausible: α is just the standard presumption that at the time where confirmation takes place, E is already known to the agent; β demands that we not be certain about the truth of T or $T \vdash \pm E$ beforehand, in line with the typical description of POE; and γ requires that T do not entail E and $\neg E$ at the same time. In particular, T has to be consistent. Finally, η is a modus ponens condition similar to equation (2): the joint degree of belief in T, $\neg E$, and $T \vdash \neg E$ is equal to the joint degree of belief in T and $T \vdash \neg E$, demanding that the agent recognize the implications that the latter two propositions have on $\neg E$.

Hence, δ really carries the burden of Jeffrey's argument. This condition has some odd technical consequences, as pointed out by Earman (1992, 127). Assume, for instance, that $p(T \vdash E) = p(T \vdash \neg E)$, which may be a plausible representation of our momentary ignorance regarding the implications of T for $\pm E$. Then it follows that $p(T \mid T \vdash E) \geq 2p(T)$, which implies that the prior degree of belief $p(T)$ must have been be smaller than .5. In other words, δ cannot be satisfied for theories that are rather probable a priori. This restriction is ad hoc and quite troubling since it severely limits the scope of Jeffrey's solution: why should probable theories not be confirmed by old evidence? To this I would like to add that in the absence of specific motivation, it is very surprising that the posterior probability of T should be at least twice as large as the prior probability.

The real problem with δ is, however, not technical but philosophical. Jeffrey (1983, 148–49) supports δ by mentioning that Newton was, on formulating his theory of gravitation G, convinced that it would bear on the phenomena he was interested in, namely, the mechanism governing the tides. Although Newton did not know whether G would entail the phenomena associated to the tides or be inconsistent with them, he used his knowledge that G would bear on the tides as an argument for endorsing it and for temporarily accepting it as a working hypothesis.

To my mind, this reconstruction conflates an evidential virtue of a theory with a methodological one. We are well advised to cherish theories of which we know that they make precise predictions on an interesting subject matter, even if we do not yet know what these predictions look like in detail. This is basically a Popperian rationale for scientific inquiry: go for theories that have high empirical content and that make precise predictions and develop

them further. They are the ones that will finally help us to solve deep scientific problems. It is plausible that Newton, on deciding to further pursue his theory of gravitation, followed this methodological rule when discovering that it would have some implications for the tides phenomena. But following this rule is very different from arguing that the informativity and empirical content of a theory increases its plausibility. Actually, Popper (1959/ 2002, 268–69) thought the other way round: theories with high empirical content rule out more states of the world and will have low (logical) probability. This is just because they take, in the virtue of making many predictions, a higher risk of being falsified. Indeed, it is hard to understand why increasing the (unconfirmed and unrefuted) empirical content of T provides an argument that T is more likely to be true. The reply that p describes a subjective rather than a logical probability function will not help much: even if p is a purely subjective function, it remains opaque why increasing the class of potential falsifiers of T will increase its plausibility. Jeffrey's condition δ is therefore ill-grounded and at the very least too controversial to act as a premise in a solution of the POE.

Earman (1992, 128–29) considers two alternative derivations of $p(T \mid T \vdash E) > p(T)$ where different assumptions carry the burden of the argument. One of them is the inequality

$$\phi. \quad p(T \mid T \vdash E) > p(T \mid [T \nvdash E] \wedge [T \nvdash \neg E]),$$

but it is questionable whether this suffices to circumvent the above objections. What Earman demands here is very close to what is supposed to be shown: that learning $T \vdash E$ is more favorable to T than learning that T gives no definite prediction for the occurrence of E or $\neg E$. In the light of the above arguments against δ and in the absence of independent arguments in favor of ϕ, this condition just seems to beg the question.

The second alternative derivation of $p(T \mid T \vdash E) > p(T)$ is proposed by Jeffrey (1983, 149) himself and relies on the equality

$$\psi. \quad p([T \vdash E] \vee [T \vdash \neg E]) = 1.$$

However, as Earman admits himself, this condition is too strong: it amounts to demanding that on formulating T, the scientist was certain that it either implied E or $\neg E$. In practice, such relationships are rather discovered gradually. As Earman continues, discussing the case of GTR: "the historical evidence goes against this supposition: . . . Einstein's published paper on the perihelion anomaly contained an incomplete explanation, since, as he himself noted, he had no proof that the solution of the field equations . . . was the unique solution for the relevant set of boundary conditions" (1992, 129). Taking stock, we conclude that Garber, Jeffrey, Niiniluoto, and Earman make

interesting proposals for solving the dynamic POE but that their solutions are either incomplete (Garber, Niiniluoto) or based on highly problematic assumptions (Earman, Jeffrey). I now present a novel solution proposal that also aims at the dynamic problem but that makes use of a slightly different conceptualization.

4. A Novel Solution of POE. The troubles with both Howson's approach to the static problem and the GJN approach to the dynamic problem seem to be technical at first sight. Howson has trouble with the counterfactual credence function $p_{K\setminus\{E\}}$, and the GJN-style solutions by Jeffrey and Earman are based on implausible assumptions. How are we supposed to take it from there?

The aim of this section consists in solving the dynamic problem and in showing that if E is already known, learning $R := T \vdash E$ raises the subjective probability of T. To achieve this goal, we combine ideas from Howson with the machinery of the GJN models. Specifically, we assume a couple of conditions on when R confirms T relative to E. The first condition has the following qualitative form:

(+) If $\neg T$ is already known, then learning $R = T \vdash E$ leaves the probability of E unchanged.

To motivate this constraint, consider the following fictitious scenario. Consider the hypothesis that Steven is going on a ski trip in the Alps. We already know that Steven had to cancel the trip because his employer did not give him leave. Now, Steven tells us: "Oh, by the way, if I were to go on this ski trip (T), I would soon buy a new winter jacket (E)." Does this utterance raise our credence that Steven will buy a new winter jacket soon? Plausibly not. Remember that we already know that the ski trip is canceled when learning $T \vdash E$. Does it lower it? Plausibly not. Steven's statement neither undermines nor supports those reasons for buying a winter jacket that are independent of his (not) going on a ski trip (e.g., the winter temperatures in his home place). More generally, learning that certain events are predicted by a refuted hypothesis is just irrelevant to our assessment of the plausibility of these events. Nostradamus's astrological theory is in all probability wrong; therefore, on learning the content of his predictions, we should neither raise nor lower our credence in the events that his theory predicted to happen.

Formally, (+) can be written as $p(E \mid \neg T, R) = p(E \mid \neg T)$ or $p(E \mid \neg T, R) = p(E \mid \neg T, \neg R)$. However, one may worry about the credence function involved. If E is old evidence, then $p(E \mid \cdot) = 1$, regardless of which proposition stands to the right of the vertical dash. In this reading, (+) would be utterly trivial. Therefore, we have, like Howson, to define a counterfactual credence function $\mathfrak{p}$ and to give it an interpretation. Instead of just eliminating E from the background knowledge, $\mathfrak{p}$ should represent the degrees of

belief of a scientist who has a sound understanding of theoretical principles and their impact on observational data but who is an imperfect logical reasoner and lacks full knowledge of the observational history (cf. Earman 1992, 134). Such probabilistic assessments are typically made by scientists who assess evidence and review journal articles: How probable would the actual evidence E be if T were true? How probable would E be if T were false? When T and $\neg T$ are two definite statistical hypotheses, such judgments are immediately given by the corresponding sampling distribution. Also in more general contexts, such judgments may be straightforward, or at least a matter of consensus in the scientific community.

This proposal resembles Howson's proposal for introducing counterfactual reasoning into the POE, but it is not bound to a particular credence function. I say more about this in section 5. Let us now proceed by formulating constraints on $\mathfrak{p}$ that allow us to solve the dynamic POE. The first one characterizes the elementary inferential relations between E, T, and R:

[1] $\mathfrak{p}(E \mid T,\, R) = 1.$

If T is true and T entails E, then E can be regarded as a certainty. In this scenario, R codifies a strict deductive relation between T and E; later, we relax this condition and also consider weaker (e.g., explanatory) dependencies.

The second condition just spells out $(+)$ in terms of the $\mathfrak{p}$-function:

[2] $\mathfrak{p}(E \mid \neg T,\, R) = \mathfrak{p}(E \mid \neg T,\, \neg R) > 0.$

In other words, if $\neg T$ is already known, then learning R or $\neg R$ does not change the probability of E, as argued above. Moreover, we should not be certain that E can only occur if T is true.

Finally, we have the following inequality:

[3]
$$\mathfrak{p}(E \mid T,\, \neg R) < \frac{1 - \mathfrak{p}(R \mid \neg T)}{\mathfrak{p}(R \mid \neg T)} \times \frac{\mathfrak{p}(R \mid T)}{1 - \mathfrak{p}(R \mid T)}.$$

This condition demands that the value of $\mathfrak{p}(E \mid \neg R,\, T)$ be smaller than the threshold on the right-hand side.[1] When the (dubious) Jeffrey condition δ is satisfied and R and T are positively relevant or neutral to each other, [3] is

1. In this formulation of condition [3], it is implicit that $1 > \mathfrak{p}(R \mid T),\, \mathfrak{p}(R \mid \neg T) > 0$ since, otherwise, either the expression would be undefined (divison by 0) or $\mathfrak{p}(E \mid \neg R,\, T)$ would have to be smaller than 0, which is impossible. We will use this fact in the proofs.

trivially satisfied since in that case, $\mathfrak{p}(R \mid T) \geq \mathfrak{p}(R \mid \neg T)$, implying that the right-hand side of [3] is greater or equal to 1. But also if R and T are negatively relevant to each other, [3] is plausibly satisfied. To see this, note that [3] can be written in the form $\mathfrak{p}(E \mid \neg R, T) < O(R \mid \neg T)/O(R \mid T)$, where O represents the betting odds for a proposition that correspond to a particular degree of belief. When the mutual negative impact of R and T is not too strong and the two betting odds are quite close to each other, [3] will be satisfied as long as $\mathfrak{p}(E \mid \neg R, T)$ is not too close to 1. And given that T is assumed to be true, but that by $\neg R$ it does not fully account for E, E should not be a matter of course for a rational Bayesian agent.

These three conditions are jointly sufficient to prove the following result (proof in the appendix):

> **Theorem 1**. Let T, R, and E be three elements of an algebra $\mathcal{A}$ with associated probability measure $\mathfrak{p}$, and let the following three conditions be satisfied:
>
> [1] $\mathfrak{p}(E \mid T, R) = 1$.
>
> [2] $\mathfrak{p}(E \mid \neg T, R) = \mathfrak{p}(E \mid \neg T, \neg R) > 0$.
>
> [3]
>
> $$\mathfrak{p}(E \mid T, \neg R) < \frac{1 - \mathfrak{p}(R \mid \neg T)}{\mathfrak{p}(R \mid \neg T)} \times \frac{\mathfrak{p}(R \mid T)}{1 - \mathfrak{p}(R \mid T)}.$$
>
> Then, R confirms T relative to old evidence E; that is, $\mathfrak{p}(T \mid E, R) > \mathfrak{p}(T \mid E)$. In other words, learning that T entails E also confirms T.

This result solves the POE on the basis of a conceptualization that combines elements from the GJN and Howson's counterfactual strategy. We use the main idea from the GJN models—the confirming event is the discovery that T accounts for/explains E—but we spell out the confirmation relation relative to counterfactual credences, as Howson envisions.

However, in many cases of scientific reasoning, condition [1]; that is, $\mathfrak{p}(E \mid T, R) = 1$, may be too strong. It may apply well to the Mercury perihelion shift, which is deductively implied by GTR, but it may fail to cover cases where T accounts for E in a less rigorous manner (see also Earman 1992, 121; Fitelson 2014). If we allow for a weaker interpretation of R (e.g., as providing some explanatory mechanism), then we are faced with the possibility that even if we are certain that T is true, and that T explains E, the conditional degree of belief in E may not be a certainty. And $\mathfrak{p}(E \mid T) < 1$ could even make sense if the relationships between T and E are deductive:

the proof of $T \vdash E$ could so complex that the involved scientists have some doubts about its soundness and refrain from assigning it maximal degree of belief. Again, Fermat's Last Theorem may be a plausible intuition pump.

For covering this case, I prove a second theorem that covers the case of $\mathfrak{p}(E \mid T, R) = 1 - \varepsilon$ for some small $\varepsilon > 0$ (proof in the appendix).

> **Theorem 2**. Let T, R, and E be three elements of an algebra $\mathcal{A}$ with associated probability measure $\mathfrak{p}$, and let the following three conditions be satisfied:
>
> [1′] $\mathfrak{p}(E \mid T, R) = 1 - \varepsilon$ for some $0 < \varepsilon < 1$.
>
> [2′] $\mathfrak{p}(E \mid \neg T, R) = \mathfrak{p}(E \mid \neg T, \neg R) > 0$.
>
> [3′]
>
> $$\mathfrak{p}(E \mid T, \neg R) < (1 - \varepsilon) \frac{1 - \mathfrak{p}(R \mid \neg T)}{\mathfrak{p}(R \mid \neg T)} \times \frac{\mathfrak{p}(R \mid T)}{1 - \mathfrak{p}(R \mid T)}.$$
>
> Then, R confirms T relative to old evidence E; that is, $\mathfrak{p}(T \mid E, R) > \mathfrak{p}(T \mid E)$. In other words, learning that T accounts for E also confirms T.

The motivations and justifications for the above assumptions are the same as in theorem 1. Condition [1′] just accounts for lack of full certainty about the old evidence, and [2′] is identical to [2]. Moreover, condition [3] of theorem 1 can, with the same line of reasoning, be extended to condition [3′] in theorem 2. Condition [3′] sharpens [3] by a factor of $1 - \varepsilon$ but leaves the qualitative argument for [3] intact. As long as $\mathfrak{p}(E \mid T, \neg R)$ and $\mathfrak{p}(E \mid T, R)$ decrease by roughly the same margin, the result of theorem 1 transfers to theorem 2.

Thus, we can extend the novel solution of POE to the case of residual uncertainty about the old evidence E—a case that is highly relevant for case studies in the history of science. If we compare this solution of the POE to Jeffrey's and Earman's proposals, we note that our assumptions [1], [2], and [3] are silent on whether Jeffrey's δ—or Earman's ϕ and ψ, for that matter—is true or false.[2] Hence, we can discard Jeffrey's dubious assumption δ that increasing empirical content makes a theory more plausible, without jeopardizing our own results.

---

2. A Mathematica notebook file that checks the relevant probability models is appended in a zip file online as proof of this claim. In the model checking, I used the PrSAT decision procedure for probabilistic models invented by Branden Fitelson (2008).

I have thus provided a solution of the dynamic POE that makes less demanding assumptions than Jeffrey and Earman and that achieves stronger results than Garber and Niiniluoto. The final section discusses the repercussions of these results on the general debate about POE and BCT.

5. Discussion. This article has analyzed some Bayesian attempts to solve the Problem of Old Evidence (POE), and it has proposed a new solution. I started with a distinction between the static and the dynamic aspect of the problem. Then I presented my criticism of the Garber-Jeffrey-Niiniluoto (GJN) approach and my own solution proposal to the dynamic POE. Like Howson, I rely on counterfactual credences, but unlike him, I use them for developing a more plausible and robust solution to the dynamic POE. Since I have already defended the specific assumptions of theorems 1 and 2, this last section is devoted to placing my proposal in the literature and to a general synopsis of the relation between POE and Bayesian Confirmation Theory (BCT).

Let us first reconsider the relation between the static and the dynamic POE. The static problem concerns, in a nutshell, the question whether E provides evidence for T, and the dynamic problem concerns the question whether discovering $T \vdash E$ confirms T. The words "evidence" and "confirmation" are often used synonymously, but they have a different emphasis, namely, explanatory power for the data versus an increase in degree of belief. This is mirrored in the literature in the foundations of statistics (e.g., Berger and Wolpert 1984; Royall 1997; Lele 2004) and also in Bayesian measures of evidential support. Some of them, like the log-likelihood ratio $l(T, E) = p(E \mid T)/p(E \mid \neg T)$, express the difference in explanatory power of T and $\neg T$ for the evidence, whereas others, such as the difference measure $d(T, E) = p(T \mid E) - p(T)$, focus on the increase in degree of belief in T. If we are interested in the dynamic POE—modeling our increase in degree of belief in T—then by assumption, old evidence E cannot confirm T. This is why many philosophers have shifted their attention to studying how learning the proposition $R := T \vdash E$ affects our degree of belief in T. In the static POE, however, we compare how expected E is under the competing hypotheses. In scientific practice, such assessments are usually counterfactual, as Howson rightfully remarks. That is, we standardly interpret $p(E \mid T)$ and $p(E \mid \neg T)$ as principled statements about the predictive import of $\pm T$ on E, without referring to our complete observational record. Such judgments are part and parcel of scientific reasoning, for example, in statistical inference, where theories T, T', and so on, impose definite probability distributions on the observations, and our credences $p(E \mid T)$, $p(E \mid T')$, and so on, follow suit (see also Ramsey 1926; Sprenger 2010).

So I believe that a resolution of the static problem should be not technical but conceptual—by spelling out why central aspects of scientific reason-

ing are suppositional and counterfactual. If we managed to substantiate this claim, the static POE would vanish because the relevant concept of evidence would differ from an increase in degrees of belief. This would not imply a farewell to Bayesian reasoning since subjective degrees of belief are required for weighing hypotheses within complex models, integrating out nuisance parameters, and so on. However, making such a counterfactual theory of scientific reasoning explicit and articulating the role of BCT in this context goes beyond the scope of this article. I hope to address the challenge in a future paper by means of a detailed analysis of scientific reasoning patterns.

It is also worth mentioning that our treatment of the POE allows for a distinction between theories that have been constructed to explain the old evidence E and those that explain E surprisingly (like Einstein's GTR). In the first case, we would not speak about proper confirmation. Indeed, if we accommodate the parameter values of a general theory T such that it explains the old evidence E, then R is actually a certainty, conditional on E: $p(R \mid E) = 1$. This is because T has been designed to explain E. As a consequence, $p(T \mid E, R) = p(T \mid E)$, and R fails to confirm T. Whereas in the case of a surprising discovery of an explanatory relation between T and E, $p(R \mid E) < 1$. The degree of confirmation that R confers on T then gets bigger, the more surprising R is—a property that aligns well with our intuitive judgment that surprising explanations have special cognitive value.

Finally, a general but popular critique of Bayesian approaches to the POE is inspired by the view that the POE poses a principled and insoluble problem for BCT. For instance, Glymour writes at the end of his discussion of the POE: "our judgment of the relevance of evidence to theory depends on the perception of a structural connection between the two, and . . . degree of belief is, at best, epiphenomenal. In the determination of the bearing of evidence on theory there seem to be mechanisms and stratagems that have no apparent connections with degrees of belief" (1980, 92–93).

What Glymour argues here is not so much that a specific formal aspect of the Bayesian apparatus (e.g., logical omniscience) prevents it from solving the POE but that these shortcomings are a symptom of a more general inadequacy of BCT: the inability to capture structural relations between evidence and theory. This criticism should not be misunderstood as claiming that confirmation has to be conceived of as an objective relation that is independent of contextual knowledge or contingent background assumptions. Rather, it suggests that solutions to the dynamic POE mistake an increase in degree of belief for a structural relation between T and E. But what makes E relevant for T is not the increase in degree of belief $p(T \mid E) > p(T)$ but the entailment relation between T and E—hence, Glymour's verdict that BCT gives "epiphenomenal" results.

To my mind, this criticism is too fundamental to be a source of concern: it does not only affect solutions to the POE, but it directly attacks the en-

tire Bayesian explication of confirmation as increase in degree of belief. However, BCT can point to a lot of success stories: explaining the confirmatory value of evidential diversity, mitigating the tacking by conjunction paradoxes, resolving the raven paradox, and so on (see Crupi 2013 for a detailed review). What I have shown in this article is that the confirmatory power of old evidence might be added to this list.

Appendix

Proofs

Proof of Theorem 1. First, we define

$$e_1 = \mathfrak{p}(E \mid T, R) \qquad e_2 = \mathfrak{p}(E \mid \neg T, R)$$
$$e_3 = \mathfrak{p}(E \mid T, \neg R) \qquad e_4 = \mathfrak{p}(E \mid \neg T, \neg R)$$
$$t = \mathfrak{p}(T) \qquad r = \mathfrak{p}(R \mid T)$$
$$t' = \mathfrak{p}(T \mid R) \qquad \bar{r} = \mathfrak{p}(R \mid \neg T).$$

By making use of [1] ($e_1 = 1$), [2] ($e_2 = e_4 > 0$), and the extension theorem $\mathfrak{p}(X \mid Z) = \mathfrak{p}(X \mid Y, Z)\mathfrak{p}(Y \mid Z) + \mathfrak{p}(X \mid \neg Y, Z)\mathfrak{p}(\neg Y \mid Z)$, we can quickly verify the identities

$$\mathfrak{p}(E \mid T) = \mathfrak{p}(E \mid T, R)\mathfrak{p}(R \mid T) + \mathfrak{p}(E \mid T, \neg R)\mathfrak{p}(\neg R \mid T)$$
$$= r + e_3(1 - r)$$
$$\mathfrak{p}(E \mid \neg T) = \mathfrak{p}(E \mid \neg T, R)\mathfrak{p}(R \mid \neg T) + \mathfrak{p}(E \mid \neg T, \neg R)\mathfrak{p}(\neg R \mid \neg T)$$
$$= e_2\bar{r} + e_4(1 - \bar{r})$$
$$= e_2$$

that will be useful later. Second, we note that by Bayes's Theorem and assumption [1],

$$\mathfrak{p}(T \mid E, R) = \mathfrak{p}(T \mid R)\frac{\mathfrak{p}(E \mid T, R)}{\mathfrak{p}(E \mid R)}$$
$$= \left(1 + \frac{\mathfrak{p}(\neg T \mid R)}{\mathfrak{p}(T \mid R)} \frac{\mathfrak{p}(E \mid \neg T, R)}{\mathfrak{p}(E \mid T, R)}\right)^{-1} \qquad \text{(A1)}$$
$$= \left(1 + \frac{1 - t'}{t'} \times e_2\right)^{-1}.$$

Third, we observe that by [1], [2], and the above identities for $\mathfrak{p}(E \mid T)$ and $\mathfrak{p}(E \mid \neg T)$,

$$
\begin{aligned}
\mathfrak{p}(T \mid E) &= \mathfrak{p}(T) \frac{\mathfrak{p}(E \mid T)}{\mathfrak{p}(E)} \\
&= \left(1 + \frac{\mathfrak{p}(\neg T)}{\mathfrak{p}(T)} \frac{\mathfrak{p}(E \mid \neg T)}{\mathfrak{p}(E \mid T)}\right)^{-1} \quad\quad (A2) \\
&= \left(1 + \frac{1-t}{t} \frac{e_2}{r + e_3(1-r)}\right)^{-1}.
\end{aligned}
$$

We also note by [3] that $e_3 = [(1 - \bar{r})/\bar{r}][r/(1 - r)]$. This allows us to derive

$$
\begin{aligned}
r + e_3(1 - r) &< r + \frac{1 - \bar{r}}{\bar{r}} \frac{r}{1 - r}(1 - r) \\
&= r \times \left(1 + \frac{1 - \bar{r}}{\bar{r}}\right) \qu\quad\quad (A3) \\
&= \frac{r}{\bar{r}},
\end{aligned}
$$

and in addition,

$$
\frac{1 - t'}{t'} = \frac{\mathfrak{p}(\neg T \mid R)}{\mathfrak{p}(T \mid R)} = \frac{\mathfrak{p}(R \mid \neg T)}{\mathfrak{p}(R \mid T)} \times \frac{\mathfrak{p}(\neg T)}{\mathfrak{p}(T)} = \frac{1 - t}{t} \times \frac{\bar{r}}{r}. \quad (A4)
$$

All this implies that

$$
\begin{aligned}
\frac{\mathfrak{p}(T \mid E, R)}{\mathfrak{p}(T \mid E)} &\overset{(A3),(A4)}{=} \left(1 + \frac{1 - t'}{t'} \times e_2\right)^{-1} \times \left(1 + \frac{1 - t}{t} \frac{e_2}{r + e_3(1 - r)}\right) \\
&\overset{(A5)}{>} \left(1 + \frac{1 - t'}{t'} \times e_2\right)^{-1} \times \left(1 + \frac{1 - t}{t} \frac{e_2 \bar{r}}{r}\right) \\
&\overset{(A6)}{=} \left(1 + \frac{1 - t'}{t'} \times e_2\right)^{-1} \times \left(1 + \frac{1 - t'}{t'} e_2\right) \\
&= \quad 1,
\end{aligned}
$$

completing the proof. The second line has also made use of $e_2 > 0$, as ensured by [2].[3] QED

3. It might be objected that we have tacitly assumed that $t = \mathfrak{p}(T) < 1$. This is right insofar as if $\mathfrak{p}(T) = 1$, then no confirmation can occur. However, $\mathfrak{p}(T) < 1$ is entailed by [1]–[3]. If $\mathfrak{p}(T) = 1$, then also $\mathfrak{p}(E \mid R) = 1$, because R states that T implies E, and T is (almost) certain. However, $\mathfrak{p}(E \mid R) = \mathfrak{p}(E \mid \neg T, R)\mathfrak{p}(\neg T \mid R) + \mathfrak{p}(E \mid T, R)\mathfrak{p}(T \mid R)$,

Proof of Theorem 2. By means of performing the same steps as in the proof of theorem 1, we can easily verify the equalities

$$\mathfrak{p}(T \mid E, R) = \left(1 + \frac{\mathfrak{p}(\neg T \mid R)}{\mathfrak{p}(T \mid R)} \frac{\mathfrak{p}(E \mid \neg T, R)}{\mathfrak{p}(E \mid T, R)}\right)^{-1}$$

$$= \left(1 + \frac{1 - t'}{t'} \times \frac{e_2}{1 - \varepsilon}\right)^{-1}$$

$$\mathfrak{p}(T \mid E) = \left(1 + \frac{\mathfrak{p}(\neg T)}{\mathfrak{p}(T)} \frac{\mathfrak{p}(E \mid \neg T)}{\mathfrak{p}(E \mid T)}\right)^{-1}$$

$$= \left(1 + \frac{1 - t}{t} \frac{e_2 \bar{r} + e_4(1 - \bar{r})}{(1 - \varepsilon)r + e_3(1 - r)}\right)^{-1}$$

$$= \left(1 + \frac{1 - t}{t} \frac{e_2}{(1 - \varepsilon)r + e_3(1 - r)}\right)^{-1},$$

where we have made use of $[1']$ and $[2'] = [2]$. We also note that $[3']$ implies

$$(1 - \varepsilon)r + e_3(1 - r) < (1 - \varepsilon)r + (1 - \varepsilon)\frac{1 - \bar{r}}{\bar{r}} \frac{r}{1 - r}(1 - r)$$

$$= (1 - \varepsilon) \times r \times \left(1 + \frac{1 - \bar{r}}{\bar{r}}\right)$$

$$= (1 - \varepsilon)\frac{r}{\bar{r}}.$$

This brings us to the final calculation:

$$\frac{\mathfrak{p}(T \mid E, R)}{\mathfrak{p}(T \mid E)} = \frac{1 + [(1 - t)/t]\{e_2/[(1 - \varepsilon)r + e_3(1 - r)]\}}{1 + [(1 - t')/t'] \times e_2/(1 - \varepsilon)}$$

$$> \frac{1 + [(1 - t)/t][1/(1 - \varepsilon)](\bar{r}/r)e_2}{1 + [(1 - t)/t] \times (\bar{r}/r) \times e_2/(1 - \varepsilon)}$$

$$= 1,$$

where we have, in the penultimate step, also applied equation (A4). This completes the proof. QED

---

and from [2] and the implicit condition in [3] ($\mathfrak{p}(T \mid R) < 1$) it follows that this expression is smaller than 1. Hence, we have a contradiction, and $\mathfrak{p}(T)$ must be smaller than 1, too.

REFERENCES

Berger, J. O., and R. L. Wolpert. 1984. *The Likelihood Principle*. Hayward, CA: Institute of Mathematical Statistics.

Bovens, L., and S. Hartmann. 2003. *Bayesian Epistemology*. Oxford: Oxford University Press.

Brush, S. G. 1989. "Prediction and Theory Evaluation: The Case of Light Bending." *Science* 246:1124–29.

Chihara, C. 1987. "Some Unresolved Problems in Bayesian Confirmation Theory." *British Journal for the Philosophy of Science* 38:551–60.

Christensen, D. 1999. "Measuring Confirmation." *Journal of Philosophy* 96:437–61.

Crupi, V. 2013. "Confirmation." In *Stanford Encyclopedia of Philosophy*, ed. E. Zalta. Stanford, CA: Stanford University. http://plato.stanford.edu/entries/confirmation/#BayConThe.

Dawid, R., S. Hartmann, and J. Sprenger. 2015. "The No Alternative Argument." *British Journal for the Philosophy of Science* 66 (1): 213–34.

Earman, J. 1992. *Bayes or Bust?* Cambridge, MA: MIT Press.

Eells, E. 1985. "Problems of Old Evidence." *Pacific Philosophical Quarterly* 66:283–302.

———. 1990. "Bayesian Problems of Old Evidence." In *Scientific Theories*, ed. C. W. Savage, 205–23. Minnesota Studies in the Philosophy of Science 14. Minneapolis: University of Minnesota Press.

Eells, E., and B. Fitelson. 2000. "Measuring Confirmation and Evidence." *Journal of Philosophy* 97:663–72.

Fitelson, B. 1999. "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity." *Philosophy of Science* 66:S362–S378.

———. 2008. "A Decision Procedure for Probability Calculus with Applications." *Review of Symbolic Logic* 1:111–25.

———. 2014. "Earman on Old Evidence and Measures of Confirmation." Unpublished manuscript, University of California, Berkeley. http://fitelson.org/research.htm.

Garber, D. 1983. "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory." In *Testing Scientific Theories*, ed. J. Earman, 99–132. Minnesota Studies in the Philosophy of Science 10. Minneapolis: University of Minnesota Press.

Glymour, C. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Hájek, A., and S. Hartmann. 2010. "Bayesian Epistemology." In *A Companion to Epistemology*, ed. J. Dancy et al., 93–106. Oxford: Blackwell.

Hartmann, S., and J. Sprenger. 2010. "Bayesian Epistemology." In *Routledge Companion to Epistemology*, ed. S. Bernecker and D. Pritchard, 609–20. London: Routledge.

Howson, C. 1984. "Bayesianism and Support by Novel Facts." *British Journal for the Philosophy of Science* 35:245–51.

———. 1985. "Some Recent Objections to the Bayesian Theory of Support." *British Journal for the Philosophy of Science* 36:305–9.

———. 1991. "The 'Old Evidence' Problem." *British Journal for the Philosophy of Science* 42: 547–55.

Howson, C., and P. Urbach. 2006. *Scientific Reasoning: The Bayesian Approach*. 3rd ed. La Salle, IL: Open Court.

Jeffrey, R. 1983. "Bayesianism with a Human Face." In *Testing Scientific Theories*, ed. J. Earman, 133–56. Minnesota Studies in the Philosophy of Science 10. Minneapolis: University of Minnesota Press.

Lele, S. 2004. "Evidence Functions and the Optimality of the Law of Likelihood." In *The Nature of Scientific Evidence*, ed. M. Taper and S. Lele, 191–216. Chicago: University of Chicago Press.

Niiniluoto, I. 1983. "Novel Facts and Bayesianism." *British Journal for the Philosophy of Science* 34:375–79.

Popper, K. R. 1959/2002. *Logic of Scientific Discovery*. London: Routledge.

Ramsey, F. P. 1926. "Truth and Probability." In *Philosophical Papers*, ed. D. H. Mellor, 52–94. Cambridge: Cambridge University Press.

Royall, R. 1997. *Scientific Evidence: A Likelihood Paradigm*. London: Chapman & Hall.

Sprenger, J. 2010. "Statistical Inference without Frequentist Justifications." In *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association*, vol. 1, ed. M. Suárez et al., 289–97. Berlin: Springer.

———. 2015. "Confirmation and Induction." In *Oxford Handbook of the Philosophy of Science*, ed. P. Humphreys. Oxford: Oxford University Press, forthcoming.

Wiles, A. 1995. "Modular Elliptic Curves and Fermat's Last Theorem." *Annals of Mathematics* 142:443–551.