David Kinney

Inductive explanation and Garber-style solutions to the problem of old evidence

Article (Published version)
(Refereed)

This version available at: http://eprints.lse.ac.uk/85907/

Available in LSE Research Online: January 2018

CrossMark

Inductive explanation and Garber–Style solutions to the problem of old evidence

David Kinney[1]

Abstract The Problem of Old Evidence is a perennial issue for Bayesian confirmation theory. Garber (Test Sci Theor 10:99–131, 1983) famously argues that the problem can be solved by conditionalizing on the proposition that a hypothesis deductively implies the existence of the old evidence. In recent work, Hartmann and Fitelson (Philos Sci 82(4):712–717, 2015) and Sprenger (Philos Sci 82(3):383–401, 2015) aim for similar, but more general, solutions to the Problem of Old Evidence. These solutions are more general because they allow the explanatory relationship between a new hypothesis and old evidence to be inductive, rather than deductive. In this paper, I argue that these solutions are either unsound or under-motivated, depending on the case of inductive explanation that we have in mind. This lends support to the broader claim that Garber–Style Bayesian confirmation cannot capture the sense in which new hypotheses that do not deductively imply old evidence nevertheless seem to be confirmed via old evidence.

Keywords Problem of old evidence · Bayesian confirmation · Philosophy of science

1 Introduction

Since Glymour (1980), the Problem of Old Evidence (POE) has posed a threat to the applicability of Bayesian confirmation theory. Put briefly, an orthodox Bayesian approach entails that if a proposition about the existence of some evidence is known to a scientist prior to the formulation of some hypothesis, then that old evidence cannot be taken to confirm the hypothesis. This is true even in cases where the hypothesis entails

✉ David Kinney
   d.b.kinney@lse.ac.uk

1   Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street Lakatos
    Building, London WC2A 2AE, UK

🠯 Springer

the old evidence in a novel way. Garber (1983) aims to solve this problem by allowing a Bayesian agent to conditionalize on the extra-systemic proposition 'H entails E'. Learning this logical fact increases the agent's degree of belief in the hypothesis H, thereby confirming the new hypothesis, despite the lack of new evidence.

I accept for the sake of argument that Garber's solution is sound. However, recent papers by Hartmann and Fitelson (2015) (hereafter 'H&F') and Sprenger (2015) have aimed for new Garber–Style solutions to the POE. These authors claim that their solutions allow us to replace the extra-systemic proposition 'H entails E' with the weaker extra-systemic proposition 'H adequately explains E', where adequate explanation need not involve a deductive relationship between a new hypothesis and old evidence. I argue that these solutions are not consistent with existing models of inductive scientific explanation. Specifically, I argue that the premises of H&F's and Sprenger's solutions are either false or under-motivated, in light of the role that frequency data plays in the formulation of inductive explanations. I do not claim that this amounts to a non-existence proof for a weaker Garber–Style solution to the POE, but it does give reason for scepticism that such a solution is forthcoming.

The plan for this essay is as follows. In Sect. 2, I provide an exegesis of the POE and Garber's solution. In Sect. 3, I lay out H&F's solution. In Sect. 4, I argue that their solution is rendered unsound or under-motivated by the role that frequency data plays in inductive explanation. In Sect. 5, I turn to Sprenger's solution. In Sect. 6, I show that Sprenger's argument runs into similar issues in cases where the truth of a hypothesis is probabilistically independent from its explanatory power with respect to some evidence. In Sect. 7, I provide some general reasons for scepticism that an account of inductive explanation will allow for a more general Gaber-style solution to the POE. In Sect. 8, I offer concluding remarks.

2 The problem of old evidence and Garber's solution

The POE is a problem for Bayesian confirmation theory. Typically, Bayesians hold that evidence confirms a hypothesis if and only if discovering that evidence raises a rational agent's degree of belief in that hypothesis. To state this precisely, let $P(\cdot)$ be a probability function representing an agent's degree of belief in some proposition, let H be the proposition that some hypothesis is true, let E be the proposition that some evidence exists, and let B be the conjunction of propositions comprising the agent's total background knowledge prior to learning E. We assume that $P(\cdot)$ satisfies the standard Kolmogorov probability axioms. We can write the basic definition of Bayesian confirmation as follows.

$$E \text{ confirms } H \text{ if and only if } \quad P(H|E, B) > P(H|B) \tag{1}$$

Using Bayes' Theorem, we can restate (1) as follows.

$$E \text{ confirms } H \text{ if and only if } \quad \frac{P(E|H, B)P(H|B)P(B)}{P(E, B)} > P(H|B) \tag{2}$$

The POE arises when we consider that in some instances the truth of E is learned before, rather than after, H is formulated. To use a classic example, Newtonian mechan-

ics was unable to accurately predict the precession of the perihelion of Mercury. When Einstein formulated general relativity, he found that his theory implied that the precession of the perihelion of Mercury should be exactly as it is observed. This discovery increased Einstein's degree of belief in the soundness of general relativity (Earman et al. 1993).

This story poses a problem for Bayesian confirmation theory. Let E be a proposition describing the precession of the perihelion of Mercury around the Sun, and let H be the proposition that general relativity is true. Bayes' Theorem tells us the following:

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

Since the agent knows the proposition E before she formulates H, $P(E|H, B) = 1$ and $P(E, B) = 1$. Since B is background knowledge, $P(B) = 1$. Therefore, $P(H|E, B) = P(H|B)$, meaning that the right side of (1) does not hold and that therefore the precession of the perihelion of Mercury does not confirm general relativity. This result suggests either that Bayesian confirmation theory is flawed, or that something has been missed in the application of Bayesian reasoning to the confirmatory relationship between new hypotheses and old evidence.

Garber (1983) aims to solve this problem by letting $H \rightarrow E$ denote 'general relativity entails the observed precession of the perihelion of Mercury'. The relevant question under this schema is not whether E confirms H, but whether the following is true:

$$P(H|E, B, H \rightarrow E) > P(H|E, B) \tag{4}$$

This can be restated as follows.

$$\frac{P(H \rightarrow E|H, E, B)P(H|E, B)P(E, B)}{P(H \rightarrow E|E, B)P(E, B)} > P(H|E, B) \tag{5}$$

Suppose that an agent already knows the facts regarding the precession of the perihelion of Mercury, in addition to her background knowledge. Once she supposes that general relativity is also true, her degree of belief in the entailment proposition $H \rightarrow E$ increases, and so $P(H \rightarrow E|H, E, B) > P(H \rightarrow E|E, B)$, meaning that $\frac{P(H \rightarrow E|H, E, B)}{P(H \rightarrow E|E, B)} > 1$. This is a subtle but important part of Garber's solution. Since H is a theory that is meant to apply to all large objects, learning that it is true should increase an agent's degree of belief that some known fact about a large object (e.g. the precession of the perihelion of Mercury) is entailed by H. Thus, (5) holds and Bayesian confirmation accords with our intuitions in this case. This solution requires us to relax the standard assumption that all logical truths have probability one, but Garber argues that we can do this without sacrificing coherence, provided that we interpret $H \rightarrow E$ as an "extra-systemic" proposition in a higher-level meta-language relative to H and E (1983, p. 112).

As Sprenger (2015) notes, Garber's solution is less a proof that the proposition $H \rightarrow E$ necessarily confirms H and more a possibility result. If we accept that $P(H \rightarrow E|H, E, B) > P(H \rightarrow E|E, B)$, then the solution goes through. However, Garber does not provide any restrictions on the probability function $P(\cdot)$ that would

entail this inequality; he only appeals to intuition. Following H&F, I will present Earman's (1992) proposal.[1] Earman proposes the following restrictions on $P(\cdot)$:

$$P(E, B) = 1 \tag{6}$$

$$P(H), P(H \to E), P(H \to \neg E) \in (0, 1) \tag{7}$$

$$0 < P(H \to E) < 1 \tag{8}$$

$$P(H|H \to E, E, B) > P(H|\neg(H \to E), \neg(H \to \neg E), E, B) \tag{9}$$

$$P(H, H \to \neg E) = P(H, H \to \neg E, \neg E) \tag{10}$$

Earman then shows that Garber's result follows straightforwardly from these axioms (1992, p. 130).

In the general relativity example, general relativity both deductively implies and adequately explains the precession of the perihelion of Mercury, at least according to Hempel's deductive-nomological account of explanation (1965). The deductive-nomological account, of course, is not the sole account of explanation on offer, and many philosophers of science would hold that adequate explanation can occur in the absence of a deductive relationship from explandum to explanans. However, (10) imposes the requirement that the relationship between H and the old evidence E *must* be deductive. To see why, note that we can re-write (10) as follows:

$$P(H, H \to \neg E) = P(H, H \to \neg E)P(\neg E|H, H \to \neg E). \tag{11}$$

So (10) is true if and only if $P(\neg E|H, H \to \neg E) = 1$, which in turn is true if and only if we interpret the arrow in $H \to \neg E$ as denoting a deductive implication relation. Thus, the scope of Earman's solution is limited to cases in which an explanans both adequately explains and deductively implies the explanandum.

As stated in the introduction, both Hartmann and Fitelson (2015) and Sprenger (2015) seek solutions to the POE that do not require that a new hypothesis deductively entail old evidence. Instead, they require only that a new hypothesis adequately explain some old evidence, where adequate explanation need not involve deductive entailment. In what follows, I will argue that neither of these solutions succeed.

3 H&F's solution

Like Garber, H&F aim to solve the POE by conditionalizing on an extra-systemic relation between hypothesis and evidence. However, they also aim to relax the requirement that a hypothesis must deductively entail old evidence in order for that evidence to confirm the hypothesis in question (2015, p. 714). Their formal argument begins by letting X denote 'H adequately explains E' and letting Y denote 'H' adequately explains E', where H' is the leading competition to H, e.g. Newtonian mechanics as opposed to general relativity. To clarify, I take it that by 'adequately explains' here H&F mean

---

[1] See Jeffrey (1983) and Niiniluoto (1983) for earlier attempts to restrict $P(\cdot)$ so as to entail Garber's result. See Sprenger (2015) for an argument that Earman's restrictions are circular.

precisely 'adequately potentially explains'; the hypothesis H would explain E if H were true. So a slightly more precise interpretation would be one such that X denotes 'If H were true, then H adequately explains E', and Y denotes 'If H' were true, then H' adequately explains E'. I take it that we can make this substitution without altering the intended upshot of H&F's argument.

H&F's argument consists of the following four premises:

$$P(H|E, B, X, \neg Y) > P(H|E, B, \neg X, \neg Y) \tag{12}$$
$$P(H|E, B, X, \neg Y) > P(H|E, B, \neg X, Y) \tag{13}$$
$$P(H|E, B, X, Y) > P(H|E, B, \neg X, Y) \tag{14}$$
$$P(H|E, B, X, Y) \geq P(H|E, B, \neg X, \neg Y) \tag{15}$$

The law of total probability implies that the following two equations hold:

$$P(H|E, B, X) = P(H|E, B, X, \neg Y)P(E, B, \neg Y|X)$$
$$+ \ P(H|E, B, X, Y)(1 - P(E, B, \neg Y|X)) \tag{16}$$
$$P(H|E, B, \neg X) = P(H|E, B, \neg X, \neg Y)P(E, B, \neg Y|\neg X)$$
$$+ P(H|E, B, \neg X, Y)(1 - P(E, B, \neg Y)|\neg X)) \tag{17}$$

It follows from (12)–(17) that $P(H|E, B, X) > P(H|E, B, \neg X)$, and therefore that $P(H|E, B, X) > P(H|E, B)$. This result entails that the explanatory proposition X confirms H, if we accept that (12)–(15) are true.

4 Problems for H&F's solution

While H&F want to remain neutral on questions concerning the nature of scientific explanation, it remains an open question whether there are any non-deductive explanatory relations that can be denoted by X in a sound interpretation of (12)–(15). If such explanations do exist, then H&F's solution expands the class of solutions to the POE as compared to Garber's solution, which requires a deductive relationship between H and E. In what follows, I show that the role that frequency data plays in the formulation of inductive explanations of old evidence renders H&F's solution either unsound or under-motivated in a broad class of conceivable use cases.

One variety of non-deductive explanation that H&F believe is compatible with their solution is Hempel's (1965) model of "inductive-statistical" explanation (H&F 2015, p. 716). Hempel's basic idea is that an explanans (which includes some hypothesis) can explain an explanandum in an inductive-statistical sense if the truth of the explanans renders the truth of the explanandum sufficiently probable. Thus, we can define the extra-systemic propositions X and Y as follows.

$$X := \mathfrak{p}(E|H) \geq p \tag{18}$$
$$Y := \mathfrak{p}(E|H') \geq p \tag{19}$$

Where p is some threshold probability and $\mathfrak{p}(\cdot)$ is a counterfactual conditional probability function. This function "represent[s] 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 empirical knowledge" (Sprenger 2015, pp. 391–392, cf. Earman 1992, p. 134). To get a feel for how this function works, suppose that a scientist knows that some theory H is true, and knows that it assigns some probability to some event E, but, counterfactually, does not know whether E occurred. The function $\mathfrak{p}(E|H)$ represents the degree of belief that the scientist would have in E in this counterfactual scenario.

We can construct cases in which the explanatory relationship between hypothesis and evidence is inductive-statistical that could putatively give rise to the POE. For example, suppose that doctors have known for some time that a person x suffers from lung cancer, but have only recently discovered that smoking explains lung cancer in the inductive-statistical sense described above. Upon learning this explanatory proposition, should x's doctors raise their degree of belief in the hypothesis that x is a smoker? I take it that the answer depends on how the doctors arrived at the realization that smoking explains lung cancer. As Roche and Sober (2013) point out, if the explanation was formulated through the analysis of previously known background data, then the explanatory proposition should not confer any additional credence on the hypothesis that x is a smoker. The hypothesis that x is a smoker entails a population frequency from which x's cancer can be induced. This entailment, rather than the proposition that smoking explains lung cancer, does the confirmatory work in a Bayesian sense.

To see why, let H be the proposition that patient x is a smoker, let E be the proposition that x has lung cancer, and let B denote the doctors' background knowledge. The extra-systemic propositions X and Y are defined according to H&F's schema. It is clear that the following is true:

$$P(H|E, B, X) = P(H|E, B) \qquad (20)$$

If a Bayesian agent had to bet on whether x is a smoker, then the amount that she would be willing to wager would not change depending on whether or not she supposed that X were true. Rather, her credence in x's being a smoker would be determined solely by the relevant frequency data, which is part of her background knowledge B. In other words, the statistical evidence "screens off" the hypothesis from the extra-systemic proposition (Roche and Sober 2013, p. 660). Thus, cases where frequency data is known prior to the formulation of the explanatory hypothesis are cases in which the conclusion of H&F's argument is false, and therefore are not interesting use cases for their solution.[2] To put the point slightly differently, in cases where new hypotheses that explain old evidence are learned through statistical analysis, the explanatory relationship between the new hypothesis and the old evidence is transparent to the scientist;

---

[2] The truth of this "Screening Off Thesis" has been challenged in the literature. See McCain and Poston (2014) and Climenhaga (2017) for objections to the thesis, and Roche and Sober (2014) and Roche and Sober (forthcoming) for what I take to be convincing responses to these objections.

there is no moment of surprise when the scientist realizes that the new theory explains the old evidence.

So it must be that H&F have cases in mind in which the relevant frequency data is not included in the agent's background knowledge. For instance, doctors might observe a patient with lung cancer but not know that smokers are more likely than non-smokers to develop lung cancer. When they learn the frequency data that implies this statistical fact, they might increase their degree of belief that H is a smoker; perhaps this is due to the additional supposition that smoking explains lung cancer in some causal sense. In these cases, there are two possibilities with regard to confirmation. On the one hand, it could be that learning new frequency data does all of the confirmatory work with respect to H. In this case, there is no problem of old evidence, since the new evidence, i.e. the frequency data, is doing all of the confirmatory work. On the other hand, it might be that the explanatory relationship between the hypothesis and the old evidence increases the probability that H is true in a way that is over and above any confirmatory power of the frequency data.

In response to this second case of putative confirmation via deductive explanation, I maintain Roche and Sober's (2013) position that any learned frequency data will screen off the hypothesis from the extra-systemic, explanatory proposition. If we let $F_{n,k}$ be the proposition that out of n smokers, k get lung cancer, and let Z be the proposition that the frequency $F_{n,k}$ is explained by a causal or otherwise explanatory relationship between H and E, the following equation is still true:

$$P(H|E, B, F_{n,k}, Z) = P(H|E, B, F_{n,k}) \qquad (21)$$

Thus, if we separate the learned frequency data from the newly formulated explanatory claim, it is not the case that the explanatory claim will have confirmatory power on its own; any increase in the probability of H will be due to the frequency data.[3]

Thus, the best test case for H&F's solution is one in which there is no frequency data linking the hypothesis to the evidence, but the relationship between the hypothesis and the old evidence is still probabilistic in nature. For example, imagine that the first and only time that two materials a and b are stored together, the warehouse containing them unexpectedly catches fire. Later, scientists develop a new hypothesis, which entails that if materials a and b are in close proximity, then fire will break out with probability $p < 1$. In this example, the warehouse fire, which is old evidence, seems to confirm the new hypothesis solely in virtue of the explanatory relationship between the warehouse fire and the chemical hypothesis. This explanatory relationship is inductive; the hypothesis implies only that the fire was likely, not that it had to occur as a matter of necessity. If we suppose in addition that the new chemical hypothesis' nearest competitor did not imply that the fire had as high a probability of occurring, then we might think that H is confirmed by the fact that H implies that the fire had a high probability of occurring, independently of any statistical data.

---

[3] This observation also implies that statistical-relevance explanations, in the style of Salmon (1971, 1989), cannot provide a sound interpretation of H&F's argument.

In these cases, I argue that H&F's axioms lack any independent justification. Consider the axiom (12). It is true if and only if the following is true:

$$P(X|H, E, B, \neg Y) > P(X|\neg H, E, B, \neg Y) \qquad (22)$$

We might attempt to justify (22) on the grounds that when H and E are both true, it is more likely that H explains E, since H is a true theory of events like E. The problem with this justification is that, as per the terms of the thought experiment, *there are no events like E*. If there were, then there would be frequency data to do the confirmatory work, and screen off the hypothesis from the explanatory proposition. Thus, while the example appears to be a promising one for H&F's solution, it ultimately leaves their axioms unjustified.

In response, one could argue that even if materials a and b are only stored in close proximity one time, it is not the case that there are no events like E. After all, H is introduced as a more general hypothesis from chemistry, and the materials a and b might have more fundamental chemical properties with behaviors that have been observed previously. If H is a theory of the behaviors of these more fundamental properties, then it is the case that H is a theory of events like E, and that therefore (22) is justified. This move is similar to one that Lewis (1994) makes in his discussion of the half-lives of elements. He considers a heavy element, unobtanium, that exists only twice and decays once after $4.8\,\mu s$ and once after $6.1\,\mu s$. He argues that these decay times in themselves do not tell us much about the half-life of unobtanium. What we need to do is to consider "general laws of radioactive decay", laws that are "function[s] of the nuclear structure of the atom in question" (Lewis 1994, p. 478). So even if a chancy event occurs very infrequently, we can subsume the event under more general laws, such that a justification for (22) might go through.

The problem with this kind of response is that when Lewis appeals to more general properties of nuclear structure in order to assign half-lives to atoms of very rare elements, he is doing so in order to arrive at a sufficiently large set of frequency data. This data set will then be informative as to the correlations between types of nuclear structure and an element's half-life. Similarly, if we can use evidence about more fundamental properties of a and b to assign a high counterfactual probability to the event that materials a and b would explode when stored together, we can do so on the basis of frequency data linking these fundamental properties to events that might be associated with combustion, such as increased temperature, friction, etc. As demonstrated above, this frequency data will screen off the hypothesis H from the explanatory proposition X, rendering (22), and therefore (12), false.

Still, it could be argued that one case that is particularly amenable to H&F's solution is a case in which a scientist possesses some old evidence, and then learns through the testimony of a second agent that a hypothesis H explains E better than H's nearest rival H'. In such a case, there would be no frequency data to screen off H from X, and so one would be tempted to conclude that H&F's solution goes through. However, in order for the first agent to accept the testimony of the second agent, the two agents must be epistemic peers to some extent. In this context, I take this to mean that they must both be scientists. Thus, there must ultimately be a scientist at the beginning of a

chain of testimony who learned H and discovered that it explained, inductively, some old evidence E, or else there is no problem of old evidence.

In this context, the problem of old evidence is as follows. Why should the scientist at the beginning of the chain of testimony, i.e. the scientist who first realizes that H explains E, increase her degree of belief in H? Given that such a scientist would not be able to rely on testimony to learn this explanatory fact, it stands to reason that in the case where H inductively explains E, she would learn H through either frequency data or theoretical analysis, and then the worries discussed above would once again apply. In the context of the Bayesian approach to confirmation via scientific theory testing, we are concerned with modelling the the beliefs of this first scientist, and thus instances of updating one's beliefs via testimony are not applicable use cases for H&F's solution.

Finally, Branden Fitelson has suggested in correspondence that one could take on board the criticisms offered above and still maintain the novelty of H&F's solution. The key move would be to change the interpretation of the extra-systemic propositions X and Y in the premises (12)–(15). Under the proposed change, X becomes either 'H is empirically adequate with respect to E' or 'H is predictively accurate with respect to E'. Y then becomes 'H' is empirically adequate with respect to E', or 'H' is predictively accurate with respect to E', where E is old evidence. With this reinterpretation in mind, suppose that our old evidence is a data set plotting the relationship between two variables w and z. A linear regression model with coefficient β has the greatest degree of fit relative to the data, but no substantive theory has been put forward as to why this is the case. Independently of the data set in question, scientists develop a theory implying that for any w, $w = \beta z$. If we let E be the proposition that the data set is observed, let H be the proposition that theory described above is true, let X be the proposition that the hypothesis H implies that a model with coefficient β best fits the data, and let Y be the proposition that a competitor hypothesis H' implies that a model with coefficient β best fits the data, then one might think that H&F's argument would be sound.

I take it that the argument to this effect would proceed roughly as follows. At first, we just have some data E. After the analysis, we learn that the best empirical model of the data is implied by the hypothesis H. This discovery, which does not involve the acquisition of any new data, should increase our degree of belief in the hypothesis H, regardless of whether or not H' also entails the best model of the data. This seems to hold whether we interpret a model's goodness of fit as a matter of empirical adequacy (i.e. the model accurately represents the existing data) or as a matter of predictive accuracy (i.e. given the value of one variable, the model does a good job of telling us the value that the other variable should take).

In response, I argue that the explanatory relationship appealed to here is deductive, rather than inductive. Recall that H entails that the relationship between w and z is best modelled by the equation $w = \beta z$. The evidential proposition E describes one body of evidence that is best modelled by the line $w = \beta z$. However, indefinitely many other bodies of evidence could also be optimally modelled by the equation $w = \beta z$. So the evidence E confirms H in virtue of the fact that E is a member of this wider set of evidential propositions. The relationship between H and the existence of *some* data set that is best modelled by the line $w = \beta z$ is deductive, rather than inductive.

So this example does not show that H&F's solution improves on Garber's solution with respect to the set of cases that it can successfully accomodate.

It might be said in response to this that H may not deductively entail that $w = \beta z$ ought to be the best-fitting model of the data set in question. Instead, H may only assign a high probability to the proposition that $w = \beta z$ best fits the data. Here, I need only note that if this probability is generated via frequency data, then this frequency data will screen off the hypothesis from the explanatory proposition, as argued above. If the probability is generated via purely *a priori* analysis, then we are back to the problem that was noted in the warehouse example. That is, if there are truly no other events like E to be observed, then it is unclear why we ought to think that the truth of both H and E renders the proposition 'H adequately explains E' more likely.

5 Sprenger's solution

Sprenger (2015) offers an alternative argument that an inductive, explanatory relationship between a new hypothesis and old evidence can be used to confirm the new hypothesis. The argument is Garber–Style in that it makes essential use of an extra-systemic proposition of the form 'H explains E'. His first move is to follow Howson (1991), and offer an account of confirmation that makes explicit use of the counterfactual probability function $\mathfrak{p}$, which retains the definition given above. Having introduced this counterfactual probability function, Sprenger adopts the following three premises (the interpretation of notation here matches the interpretation of H&F's solution):

$$\mathfrak{p}(E|H, X) = 1 - \epsilon, \text{ where } 0 \leq \epsilon < 1 \tag{23}$$

$$\mathfrak{p}(E|\neg H, X) = \mathfrak{p}(E|\neg H, \neg X) > 0 \tag{24}$$

$$\mathfrak{p}(E|H, \neg X) < (1 - \epsilon)\frac{1 - \mathfrak{p}(X|\neg H)}{\mathfrak{p}(X|\neg H)} \frac{\mathfrak{p}(X|H)}{1 - \mathfrak{p}(X|H)} \tag{25}$$

Sprenger then proves that (23)–(25) collectively imply that $\mathfrak{p}(H|E, X) > \mathfrak{p}(H|E)$, and therefore that the explanatory proposition X confirms the hypothesis H. Although for brevity's sake I do not reproduce the proof here (see Sprenger 2015, pp. 397–399), it is similar to H&F's proof in that it makes essential and novel use of ordinal constraints on credences, along with the law of total probability.

6 Problems for Sprenger's solution

The soundness of Sprenger's three premises depends crucially on how one interprets the proposition X. If one were to interpret it as 'H deductively entails E', then all three premises are plausibly true. If an agent were to learn that H was a true theory of events like those described by E, and that H entailed that E was true, then $\mathfrak{p}(E|H, X) = 1$, so that $\epsilon = 0$. Thus, (23) is true. Similarly, the events described by E should have the same positive probability of occurring when H is false, regardless of whether or not H deductively implies E, and so (24) is true as well. As for (25), if an agent were to learn that H was a true theory of events like those described by E, and also that H did

not deductively entail E, then her credence that E is true would be low. By contrast, learning that H is true or false is uninformative as to whether or not H entails E, so that $\mathfrak{p}(X|\neg H) = \mathfrak{p}(X|H)$. It follows from this that $(1 - \epsilon)\frac{1-\mathfrak{p}(X|\neg H)}{\mathfrak{p}(X|\neg H)} \frac{\mathfrak{p}(X|H)}{1-\mathfrak{p}(X|H)} = 1$ when $\epsilon = 0$. Thus (25) is true as well, and Sprenger's solution is sound.

Once we allow that X is not a relationship of deductive implication between H and E, it is much less clear that (25) is consistent with most accounts of scientific explanation. Just as the truth of H was probabilistically independent of its implying some evidence E, the truth of H does not on its own make it any more likely to explain anything. As an illustration of this point, consider the fact that economists and other social scientists often speak of models and theories that are both false and explanatory. One way of accommodating the ability of false hypotheses to explain is to suppose that the truth value of H makes no difference to the degree of belief that an agent would assign to the proposition that H explains some event E, i.e. that $\mathfrak{p}(X|H) = \mathfrak{p}(X|\neg H)$.[4] Under this condition, $\frac{1-\mathfrak{p}(X|\neg H)}{\mathfrak{p}(X|\neg H)} \frac{\mathfrak{p}(X|H)}{1-\mathfrak{p}(X|H)} = 1$, so that (25) can be re-written as follows.

$$\mathfrak{p}(E|H, \neg X) < \mathfrak{p}(E|H, X) \tag{26}$$

Thus, if we suppose that the explanatory power of a hypothesis can be independent of its truth value, then Sprenger's argument depends on roughly the same assertion as H&F's, i.e. that the evidence available to an agent does not screen off the extra-systemic explanatory proposition from the hypothesis.

As demonstrated above, we have good reason to hold that this kind of screening off *does* occur, and that therefore (26) is false. That is, the explanatory capacity of H with respect to E should not make a difference to a Bayesian agent's credence in E. So Sprenger's solution is unsound in cases where the explanatory power of a hypothesis is independent of its truth value. Note here the work being done by the relaxation of the assumption that if X is true, then H deductively entails E. In the deductive case, we are able to show that the third premise is equivalent to $\mathfrak{p}(E|H, \neg X) < 1$, which is a much less suspect assumption than (26). This is because there is a deductive fact, rather than an explanatory one, that is able to do confirmatory work for the agent (see Sect. 7 for further elaboration of this point).

It might be objected at this point that true hypotheses are at least somewhat more likely than false hypotheses to explain some event E, so that $\mathfrak{p}(X|H) > \mathfrak{p}(X|\neg H)$. If this were true, then it would be the case that $\frac{1-\mathfrak{p}(X|\neg H)}{\mathfrak{p}(X|\neg H)} \frac{\mathfrak{p}(X|H)}{1-\mathfrak{p}(X|H)} > 1$, and Sprenger's axioms would be sound. However, I do not believe that we have sufficient motivation to assume that $\mathfrak{p}(X|H) > \mathfrak{p}(X|\neg H)$. Elsewhere, Sprenger and Schupach write that a hypothesis' explanatory power "has to do with a hypothesis' ability to decrease the degree to which we find the explanandum surprising" (Schupach and Sprenger 2011, p. 108). In many cases, it seems as though a false hypothesis or a highly idealized

---

[4] Indeed, Schupbach and Sprenger argue that one desideratum for an account of explanatory power is that it is not the case that "the plausibility of the hypothesis in itself would affect the degree of explanatory power that H lends to E" (2011, p. 111). This point is upheld by Crupi and Tentori in their critical evaluation of Schupbach and Sprenger (2012, p. 367).

explanatory model may render a given explanandum much less surprising than a true hypothesis.

A classic example of this positive relationship between falsehood and explanatory power is Hotelling's model in economics, which in spite of its many false assumptions—e.g. the assumption that agents move in one-dimensional space, or the assumption that consumer goods differ only with respect to the distance between the consumer and the point of consumption—is taken to render many social and economic phenomena less surprising. These phenomena include the clustering of electronics shops around Tottenham Court Road or the convergence of mainstream political platforms around the political centre (see Reiss 2012 for details). Thus, I conclude that the auxiliary assumption that a hypothesis' truth and its explanatory capacity are positively correlated lacks sufficient motivation, and hold that Sprenger's argument remains unsound.

7 General reasons for scepticism

At this point, the upshot of this discussion should be clear. I am not convinced that either H&F or Sprenger are able to solve the POE under a broader interpretation of the relation between H and E than the interpretation that Garber uses. As stated in the introduction, this is not a non-existence proof, but it is a good reason for maintaining a sceptical attitude as to the applicability of H&F's argument.

There are further reasons to be sceptical that such a solution is forthcoming. An argument to this effect proceeds as follows. Inductive explanations generally involve probabilities; the truth of the explanans ought to increase the probability of the truth of the explanandum. As Hájek (2012) argues, one necessary condition for probability assignments is that they be accountable to observed relative frequencies, when such observations can be made. In other words, it should be the case that more probable events generally occur more often than less probable events. With some exceptions, the targets of probabilistic explanations in science tend to be repeatable events. So a real-world Bayesian agent is likely to have background knowledge that includes the frequencies with which the types of events contained in a given explanandum occur. If we learn that some theory explains a class of observed phenomena in a probabilistic sense, then we learn some important new information. Namely, we learn that, given the truth of the hypothesis, the frequency with which explanandum-type events *ought* to occur is roughly equal to the observed frequency, where 'ought' is understood in a nomological or causal sense. The problem for Bayesian confirmation theory is that the numerical value of a conditional probability $P(E|H)$ is independent of whether or not E ought to have the probability that it does; this kind of nomological or causal information just cannot be represented by the probability function. Thus, it seems that the confirmatory power of probabilistic explanations of old evidence cannot be fit neatly into Bayesian confirmation theory.

In deductive cases of Garber–Style confirmation, it seems that it is the deductive relationship between H and E, rather than the explanatory relationship, that is doing the real confirmatory work. Roche and Sober (2014) argue this point as follows. Suppose that we introduce two separate extra-systemic propositions into our framework.

The first is the standard X, meaning 'If H were true, then H adequately explains E' and the second, I, is the proposition 'H deductively implies E'. It seems obviously true that $P(H|E, I, X) = P(H|E, I)$. In a context where we are uncertain as to whether or not H is true, the fact that H deductively implies E may increase an agent's degree of belief in H, but any facts about an explanatory relationship between H and E, over and above this deductive relationship, do not do any confirmatory work. Thus, it is deduction, and not explanation, that matters in these cases.

To be clear, I take this argument to be telling against Bayesian confirmation theory, rather than against inductive explanation. Suppose that we discover, surprisingly, that a broad probabilistic theory of some domain of events entails that one type of event in the domain ought to be observed about as often as it is, in fact, observed. This seems to speak in favor of the theory, everything else being equal. I concur with this intuition. It is just that standard Bayesian confirmation theory is not able to capture this intuition, for the reasons given above. The key culprit is the inability of Bayesian confirmation theory to distinguish between probabilities that just happen to hold, and probabilities that ought to hold, where the 'ought' is interpreted nomologically or causally.

One way that we could distinguish between probabilities that just happen to hold, and probabilities that ought to hold for nomological or causal reasons, is by measuring the degree of resiliency of a given probability assignment. Some probability assignments, such as the probability that the fourth ball drawn from a given urn is red, are not very resilient in the face of new evidence. If the first three pulls from the urn revealed a high or low frequency of red balls, then an agent's credence regarding the color of the fourth ball would quickly change. The same cannot be said for probabilities that hold in virtue of the truth of scientific hypotheses. Given the well-established connection between smoking and lung cancer, an agent who accepts that smokers are more likely than non-smokers to develop lung cancer will likely retain this belief, even if she is presented with many repeated cases of smokers who do not develop lung cancer and non-smokers who do develop lung cancer. However, this difference in resilience is not captured in the first-order probability assignments themselves, nor is it captured in the ordinal ranking of such assignments. Thus, I am sceptical that a model of confirmation using only first-order probability assignments will work in cases of inductive explanation.[5]

It might be suggested that the counterfactual probability function, which represents the credence that an agent with a sound understanding of theoretical principles would have in some event, aims to capture this degree of resiliency. The issue with this suggestion is that there may not be any difference between this scientifically informed agent's credence and the credence of an agent who observes frequencies but has no theoretical understanding of why those frequencies occur. Both agents might assign the same probability to the event that a given lung cancer patient smokes, but only one will have any understanding of why the probability has the value that it does.

---

[5] Though a full discussion of the resiliency of probability assignments and their role in confirmation is beyond the scope of this paper, Joyce (2005) gives an account of this notion of resilience, which he calls the "weight of evidence".

8 Conclusion

In this essay, I examined two recent Garber–Style solutions to the POE that relax the requirement that the new hypothesis H logically entails the old evidence E. In both cases, I found that these putative solutions move too quickly. The formal results, although valid, do not have a sound interpretation among the existing models of non-deductive scientific explanation. This fits with a more general scepticism regarding the prospects for such a solution. Although fully fleshing out this point is beyond the scope of this project, I suspect that a formal apparatus beyond first-order conditional probabilities is required to make sense of the way in which non-deductive probabilistic theories are confirmed by old evidence.

References

Climenhaga, N. (2017). How explanation guides confirmation. *Philosophy of Science*, *84*(2), 359–368.

Crupi, V., & Tentori, K. (2012). A second look at the logic of explanatory power (with two novel representation theorems). *Philosophy of Science*, *79*(3), 365–385.

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

Earman, J., Janssen, M., & Norton, J. D. (1993). *The attraction of gravitation: new studies in the history of general relativity*. Basel: Einstein studies Boston: Birkhäuser.

Garber, D. (1983). Old evidence and logical omniscience in bayesian confirmation theory. *Testing scientific theories*, *10*, 99–131.

Glymour, C. (1980). *Theory and evidence*. New Jersey: Princeton University Press.

Hájek, A.(2012). Interpretations of probability. In Zalta, E. N., (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University, winter 2012 edition.

Hartmann, S., & Fitelson, B. (2015). A new garber-style solution to the problem of old evidence. *Philosophy of Science*, *82*(4), 712–717.

Hempel, C. G.(1965). *Aspects of scientific explanation*. New York: The Free Press.

Howson, C. (1991). The 'old evidence' problem. *British Journal for the Philosophy of Science*, *42*(4), 547–555.

Jeffrey, R. C. (1983). Bayesianism with a human face. In J. Earman (Ed.), *Testing scientific theories* (pp. 133–156). Chicago: University of Minnesota Press.

Joyce, J. M. (2005). How probabilities reflect evidence. *Philosophical perspectives*, *19*(1), 153–178.

Lewis, D. (1994). Humean supervenience debugged. *Mind*, *103*(412), 473–490.

McCain, K., & Poston, T. (2014). Why explanatoriness is evidentially relevant. *Thought: A Journal of Philosophy*, *3*(2), 145–153.

Niiniluoto, I. (1983). Novel facts and bayesianism. *British Journal for the Philosophy of Science*, *34*(4), 375–379.

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, *19*(1), 43–62.

Roche, W., & Sober, E. (2013). Explanatoriness is evidentially irrelevant, or inference to the best explanation meets bayesian confirmation theory. *Analysis*, *73*(4), 659–668.

Roche, W., & Sober, E. (2014). Explanatoriness and evidence: A reply to mccain and poston. *Thought: A Journal of Philosophy*, *3*(3), 193–199.

Roche, W., & Sober, E. Is explanatoriness a guide to confirmation? a reply to climenhaga. *Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie*, forthcoming.

Salmon, W. C. (1971). *Statistical Explanation & Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Salmon, W. C. (1989). 4 decades of scientific explanation. *Minnesota Studies in the Philosophy of Science*, *13*, 3–219.

Schupbach, J. N., & Sprenger, J. (2011). The logic of explanatory power. *Philosophy of Science*, *78*(1), 105–127.

Sprenger, J. (2015). A novel solution to the problem of old evidence. *Philosophy of Science*, *82*(3), 383–401.