Confirmation by Explanation: A Bayesian Justification of IBE

Marko Tešić* Ben Eva† Stephan Hartmann‡

August 14, 2017

Abstract

We provide a novel Bayesian justification of inference to the best explanation (IBE). More specifically, we present conditions under which explanatory considerations can provide a significant confirmatory boost for hypotheses that provide the best explanation of the relevant evidence. Furthermore, we show that the proposed Bayesian model of IBE is able to deal naturally with the best known criticisms of IBE such as van Fraassen's 'bad lot' argument.

1 Introduction

Inference to the best explanation (IBE) is a form of non-deductive reasoning that, it has been widely argued, plays a crucial role in both scientific and everyday reasoning contexts. To illustrate, suppose that you leave a piece of cheese on the kitchen table in the evening. The next morning, you find that the cheese is gone (except for a few crumbs), and you see that there is a small hole in the bottom of the wall. The best explanation for these observations is that a mouse visited the kitchen in the night, and you subsequently infer the truth of this hypothesis on the basis of its explanatory power

*Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://lmu-munich.academia.edu/MarkoTesic – marko.tesic375@gmail.com.

†Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://www.mcmp.philosophie.uni-muenchen.de/people/faculty/eva_benjamin/index.html – benedgareva@icloud.com.

‡Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://www.stephanhartmann.org – s.hartmann@lmu.de.

(the example is due to van Fraassen 1980: 19–20). Similarly, Edmund Halley (1752) argued that the best explanation of the observed comets of 1531, 1607, and 1682 was that these observations were all due to a single comet (later named 'Halley's comet') that made three revolutions in an elliptical orbit around the sun with a period of 75–76 years. That the one-comet hypothesis best explains the evidence raises our confidence in that hypothesis.[1] The prevalence of IBE in science has led some to suggest that IBE is *the* quintessential way of arguing for theories in science (e.g. Lipton (2004); Psillos (1999); Williamson (2016)).

Despite the apparent omnipresence of IBE in scientific reasoning, there has been no broad agreement on its normative status, or even on its exact formulation. Consider the two examples just mentioned: in the cheese example one infers the truth of the hypothesis from the fact that the hypothesis best explains the evidence (see Harman 1965), whereas in the comet example the inference results only in an increase in the probability of the conclusion.[2] Regardless of how one formulates IBE, the general idea is the following: explanatory considerations are truth-conducive; that a hypothesis is the best explanation is a mark of the truth of that hypothesis. More specifically, all formulations agree that explanation has a confirmatory role: "explanatory considerations contribute to making some hypotheses more credible, and others less so" (Douven 2011: 22). The normative problem then is to show under what conditions (if any) the fact that a hypothesis is the best explanation makes that hypothesis more likely to be true than if it had not been the best explanation.

Given the lack of any consensus regarding the conditions under which IBE can be legitimately employed as a sound form of ampliative inference, it is perhaps unsurprising that many authors have argued that the inference scheme can never be given a genuine normative vindication. For example, consider the so called 'bad lot' argument (originally due to van Fraassen 1989). The gist of the argument is the following. The value of any instance of IBE is constrained by the set of hypotheses under consideration. If the set does not contain a true hypothesis, then IBE can only ever return a false conclusion. For, IBE takes as a premise only that some hypothesis provides a better explanation than all those explanations that have hitherto been considered. IBE does not provide us with

[1] More examples of IBE can be found in Douven (2011); Glymour (1984); Lipton (2004); Thagard (1978). For an extensive overview and a critical discussion of examples of IBE see Norton (2016a; 2016b).

[2] For more on different formulations of IBE see Douven (2011; 2002).

any guarantee that we are not starting with a bad lot, i.e. a set of false hypotheses that does not contain the true one. Therefore, IBE can hardly be an inferential scheme for attaining true beliefs (for responses to the bad lot objection, see e.g. Lipton (2004); Day and Kincaid (1994); Schupbach (2013); and Brössel (2015)[3]).

A second well known objection (also due to van Fraassen 1989: 160–170) is that IBE, when formulated in probabilistic terms, is incoherent. Imagine a Bayesian agent who also considers IBE to be a legitimate inference scheme and agrees to add 'bonus points' to the posterior probabilities of a hypothesis after conditionalizing on new evidence, on the basis of how well the hypothesis explains the evidence: the hypothesis which best explains the evidence receives the bonus points and all other hypotheses receive no bonus points. Van Fraassen contends that this updating strategy is liable to a dynamic Dutch-Book, for the simple reason that it departs from standard Bayesian conditionalization (see e.g. Teller 1973). Therefore, this probabilistic version of IBE violates the demands of Bayesian rationality.[4]

Despite these (and many other) fundamental criticisms of the soundness of IBE-style inferences, a number of authors have nevertheless attempted to provide normative foundations that legitimate the use of IBE in scientific reasoning (see e.g. Harman 1967, Douven 2002, and Psillos 2002). Here, we will be interested specifically in those defences of IBE that explicitly attempt to render the inference scheme compatible with Bayesian approaches to inductive reasoning in science. Perhaps the most influential defence of this kind is due originally to Lipton (2001, 2004). Lipton argues that IBE and (subjective) Bayesianism can be made compatible once one allows for the possibility that explanatory considerations can be used to inform the prior probabilities and likelihoods that play a role in Bayesian updating. Famously, subjective Bayesianism (in its standard formulation) does not place any definite restrictions on the assignment of prior probabilities.[5] Thus, it seems natural to suggest that explanatory considerations can play a significant role in determining the prior probabilities and likelihoods that are underspecified by standard subjective Bayesianism. If this is true, then IBE, far from being incompatible with Bayesian reasoning, actually plays an important role in determining the subjective prob-

[3]See Dellsén (2017b) for a view on why some of these responses do not succeed.

[4]A number of authors (e.g. Okasha 2000 and Lipton 2004) have criticized this argument claiming that van Fraassen's representation of a probabilistic version of IBE—imagining a Bayesian agent who adds extra bonus points to the best explanatory hypothesis after conditionalization—is idiosyncratic.

[5]This fact is commonly referred to as the 'problem of the priors'.

ability distributions that underlie the Bayesian formalism. This implies that IBE inherits its normative justification from the role it plays in Bayesian inference.[6] However, this way of justifying IBE would render IBE only an auxiliary device and not an autonomous mode of inference, since it is Bayesianism that provides us with normatively correct answers (see Farmakis & Hartmann 2005). Indeed, the sense in which explanatory considerations can be used to inform prior probabilities and likelihoods has never been systematically explicated, and it has been argued that the proposal is too vague to count as a genuine justification of IBE (see Weisberg 2009).[7]

Overall then, the current situation is clearly unsatisfactory. On the one hand, IBE is arguably one of the most important methods for arguing in science and, on the other hand, no extant justification of IBE is able to provide us with precise conditions for the soundness of IBE-style inferences. The goal of this paper is to specify such conditions. More specifically, our aim is to provide a (subjective) Bayesian justification of IBE without simply stipulating that explanatory considerations inform the priors and the likelihoods. We attempt to show, contra van Fraassen, that explanatory considerations directly affect the confirmation that hypotheses receive from novel inductive evidence, and that they do so in a way that is perfectly compatible with Bayesian conditionalization.

The article proceeds as follows. In the next section we motivate and present a novel Bayesian model of IBE, arguing that this model allows us to treat explanatory considerations as evidentially significant without departing from the standard Bayesian framework (Section 2). We then go on to discuss some prominent criticisms of IBE and address them in light of our model (Section 3). Lastly, we present conclusions (Section 4). Throughout the article, we work in the framework of Bayesian epistemology.[8]

[6]This way of justifying IBE can also be attributed to a number of authors who argue for the compatibility of IBE and Bayesianism: see Okasha (2000) and Henderson (2015) for instance. Weisberg (2009) also argues for the compatibility claim between IBE and *objective* Bayesianism and, at least implicitly, the reliability of IBE. For a criticism of Weisberg's proposal see Cabrera (2017).

[7]For further criticisms of the idea that IBE and Bayesianism can be rendered compatible by allowing explanatory considerations to 'inform' the priors and likelihoods, see Henderson (2015).

[8]For surveys on Bayesianism see Háyek & Hartmann (2010) and Hartmann & Sprenger (2011). For a critical discussion of Bayesianism see Earman (1992). Applications of Bayesian epistemology to scientific reasoning can be found in Bovens & Hartmann (2003). Throughout the article, we follow the convention that propositional variables are denoted by italicized letters (A) and the values of these variables are denoted by non-italicized letters (A or ¬A).

2 A Bayesian Model of IBE

2.1 IBE and Novel Evidence

In standard formulations of IBE, it is commonly assumed that the evidence to be explained is *old evidence*. We know that the evidence obtains and we try to explain it. The hypothesis that offers the best explanation is subsequently confirmed. Thus, in the cheese example, the evidence leads us to formulate the hypothesis that a mouse ate the cheese, and the fact that this offers the best available explanation of the evidence leads us to regard the hypothesis as (probably) true. On closer inspection, however, one finds that in the literature on IBE it is *new evidence* that typically provides confirmation. There are two ways to understand this. First, van Fraassen (1989) and Douven (2013) take IBE to operate on a pre-existing set of hypotheses: IBE selects the best among existing explanations that have already been formulated irrespective of evidence. For instance, both van Fraassen and Douven consider a set of hypotheses where each hypothesis expresses a different coin bias before the coin was tossed, and it is only after a coin was tossed that they ask which of the hypotheses best explains the evidence (which in this case is a sequence of heads and tails). Thus, IBE only takes effect once new evidence is obtained.

Another way of understanding the role of novel evidence in IBE is to say that although hypotheses are often formulated in order to explain an existing body of old evidence, the actual confirmation of those hypotheses only happens later, once new evidence is obtained. Thus, we read:

> Although a hypothesis might be reasonably accepted as the most plausible hypothesis based on explanatory considerations (abduction), the *degree of confidence* in this hypothesis is tied to its subsequent confirmation. (Psillos 2000: 67, original emphasis)

Indeed, it would seem rather circular to say that, following our example, we first form the hypotheses in order to explain the existing evidence, and then use the very same evidence to confirm them. Sentiments of this type have also been forwarded by e.g. Norton (2016a; 2016b), who argues that (new) evidential import always plays a significant role in prospective examples of IBE. Lipton (2004: 113) argues that both 'loveliness' (i.e. how explanatory a hypothesis is relative to evidence) and Bayesian likelihoods are relative to new evidence. Henderson (2015: 696) points out that unification (often cited as a

paradigmatic explanatory virtue) depends on both old and new evidence, and that "[w]e assess the best explanation with respect to all the data, past and present." Okasha (2000: 703) argues that explanatory considerations figure simultaneously with confirmation, once new evidence is obtained. To illustrate, consider the following example (due to Okasha 2000: 702–203). A mother takes her child, who is obviously in pain, to see a doctor. Based on the mother's description, the doctor forms two hypotheses concerning what's wrong with the child, H_1 and H_2. The doctor then further examines the child, observes some new symptoms, and decides that H_2 is a better explanation of the observed symptoms than H_1. The doctor concludes that H_2 is a more plausible hypothesis. So, the doctor first formulated the two hypotheses based on the mother's description (old evidence). But it is only after the doctor has further examined the child (i.e. after the new evidence came in) that she decides that the hypothesis H_2 is the better explanation of the symptoms and rejects H_1 as implausible.

Another example that illustrates how new evidence plays a role in IBE is the case of Halley's comet. Halley (1752: Oooo3) recounts: "...I suspected, from the like situation of their Planes and Perihelion, that the Comets which appeared in the years 1531, 1607, and 1682, were one and the same Comet that had made three Revolutions in its Elliptic Orbit." After establishing the orbit of the hypothesized comet more precisely, Halley went on to show that the observational consequences of his hypothesis cohered well with the existing data:

> You see therefore an agreement of all the Elements in these three, which would be next to a miracle if they were three different Comets; or if it was not the approach of the same Comet towards the Sun and Earth, in three different revolutions in an Ellipsis around them. Wherefore if according to what we have already said it should return again about the year 1758, candid posterity will not refuse to acknowledge that this was first discovered by an *Englishman*. (Halley 1752: Ssss, original emphasis)

Halley's reasoning seems very much in line with IBE. He formulated a hypothesis that nicely explained the existing evidence, and argued that other explanations postulating more than one comet seemed unlikely, though they might have fit the evidence equally well. However, it was not until the next observation of the comet that Halley's hypothesis was actually confirmed (see also Laplace 1995/1825: 3 and Salmon 2001: 123–124). So it

is not the old evidence (i.e. the evidence that the hypothesis was formulated to explain) that confers confirmation to the hypothesis; rather, the hypothesis is confirmed by the future, yet unobserved evidence. Although the fact that a hypothesis offers the best available explanation of the evidence it was designed to explain may well be good reason for us to entertain it as a serious possibility, it seems strange to claim that this kind of reasoning leads to genuine confirmation of the hypothesis. Thus, we follow Norton and others in claiming that in prospective examples of IBE, the actual confirmation only ever takes place after some novel evidence has been obtained. But unlike Norton and other critics of IBE, we contend that explanatory considerations can make a significant difference to the confirmatory import of that novel evidence.

Thus, the conception of IBE considered here can be found in the writings of both advocates and critics of IBE. Van Fraassen (1989) criticised this conception IBE as being liable to a dynamic Dutch book. Specifically, he argued that any attempt to include explanatory considerations in one's evidential updating rule will lead to a necessary deviation from Bayesian conditionalization, and will thereby render one susceptible to a dynamic Dutch book. However, Douven (2013), one of the advocates of IBE, showed that there exist many scenarios in which augmenting standard Bayesian conditionalization by awarding 'bonus points' to hypotheses that provide the best explanation of the novel evidence will lead to a genuine increase in performance (as measured by proper scoring rules). So, there are situations in which agents who update in accordance with IBE will consistently outperform their Bayesian counterparts. Furthermore, Douven and Schupbach (2015) report experiments which appear to demonstrate that explanatory considerations play a crucial role in the way that people actually go about updating their beliefs in everyday reasoning contexts. In what follows, we attempt to show that (i) it is possible for the Bayesian to take the confirmatory significance of explanatory considerations into account without surrendering or amending standard conditionalization, and (ii) it is possible for the advocate of IBE style inferences to avoid dynamic Dutch books.

2.2 The Model

This brings us to our Bayesian model of IBE. As mentioned above, we want to show that the fact that a hypothesis is the best explanation confers confirmatory support to the hypothesis *in addition* to the confirmatory support conferred by (new) evidence. We begin

by making some important conceptual clarifications and outlining some fundamental assumptions of our model.

First, note that the property of 'being a good (best) explanation' is not an intrinsic property of hypotheses. In particular, whether or not a hypothesis H counts as a good explanation is always determined relative to *a particular body of evidence*. H can exhibit a wealth of explanatory virtues when considered in the light of one body of evidence, but be found severely lacking in alternative evidential contexts. A nice way of seeing this is to consider, for example, the Bayesian Information Criterion (BIC), which evaluates a hypothesis based on both its fit to a fixed body of evidence and its internal complexity. If H accounts well for the evidence without sacrificing too much in the way of simplicity, it will have a good (low) BIC score, and will be considered a good explanation *of the relevant evidence*. In what follows, we assume that the property of 'being a good (best) explanation' is always a binary one that applies not to individual hypotheses, but rather to hypothesis-evidence pairs.

Secondly, we contend that explanatory considerations can only ever have confirmational import in situations where the evidence being explained is *known to obtain*. Clearly, the fact that H provides the best explanation of E should not be taken as indicative of H's truth in situations where we think that E is likely to be false. We might think that Creationism would provide an excellent explanation of the fossil record if it were the case that there were no fossils older than 10,000 years. However, since this is not the evidential situation we find ourselves in, it would be extremely strange to use this observation as an argument for the truth of Creationism. Thus, we assume that the fact that H provides the best explanation of E can only ever be confirmationally relevant to the truth of H in cases where we know that E obtains. So if we fix a potential piece of evidence E and a hypothesis H and let X be the proposition 'H is the best available explanation of E', we require that H and X should be probabilistically independent when we do not know whether or not E in fact obtains, since otherwise H and X would be probabilistically dependent on each other even if the evidence E does not obtain, which leads to strange consequences as shown in the example above. However, we allow for the possibility that X and H can be probabilistically dependent, once we know whether the evidence E obtains. To illustrate, imagine that H is a hypothesis that would provide the best available explanation of E, were E to obtain. We do not yet take this to count in H's favour, since

E might never be observed. However, if we subsequently go on to perform an experiment that produces the evidence E, the fact that H provides the best available explanation of E may well be taken as a sign of its truth. To summarise then, we have at least three propositional variables H, E, and X with corresponding values:

H: The hypothesis is true.

¬H: The hypothesis is false.

E: The evidence obtains.

¬E: The evidence does not obtain.

X: Of all the currently available hypotheses, H would be the best explanation of E, were E to obtain.

¬X: Of all the currently available hypotheses, H would not be the best explanation of E, were E to obtain.

We have argued that H and X should be probabilistically independent in the absence of knowledge about the value of E, i.e. (i) $H \perp\!\!\!\perp X$. We have also argued that H and X may be probabilistically dependent conditional on a known value of E, i.e. we want to allow for the possibility that (ii) $\neg(H \perp\!\!\!\perp X \mid E)$. Together, these conditions are sufficient to pick out the following Bayesian network representation (Figure 1) of the probabilistic relationships between H, E and X.

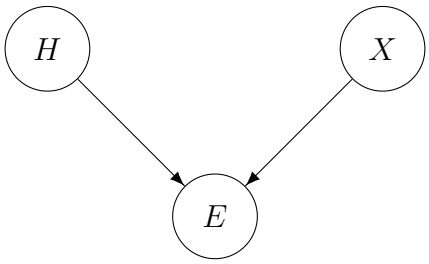

Figure 1: The Bayesian Network representation of IBE

Our basic aim in this paper is to show that explanatory considerations can make a difference to the confirmatory import of novel evidence. In this context, this amounts to proving the following inequality,

$$P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) > P(\mathrm{H} \mid \mathrm{E}) \tag{1}$$

If (1) holds, then upon learning the novel evidence E, the fact that H is the best available explanation of E will add to the confirmation that E confers upon H in the absence of explanatory considerations. Before proving Eq. (1) we need to specify the basic parameters of the network.

$$\begin{aligned} P(\mathrm{H}) = h \quad &, \quad P(\mathrm{X}) = x \\ P(\mathrm{E} \mid \mathrm{H}, \mathrm{X}) = \alpha \quad &, \quad P(\mathrm{E} \mid \mathrm{H}, \neg\mathrm{X}) = \beta \\ P(\mathrm{E} \mid \neg\mathrm{H}, \mathrm{X}) = \gamma \quad &, \quad P(\mathrm{E} \mid \neg\mathrm{H}, \neg\mathrm{X}) = \delta \end{aligned} \tag{2}$$

At this stage, we need to motivate one further constraint on the parameters of the network. This constraint is motivated by the idea that *we should expect to observe evidence that is well explained by the true hypothesis*. To illustrate, imagine that we are interested in describing house prices as a function of average income in the area of the property. Suppose further that we are certain that, in the long run, house prices are correctly described by one and only one of the three curves H_1, H_2 and H_3 (see Figure 2). Next, suppose that we are awaiting some relevant survey data regarding the relationship

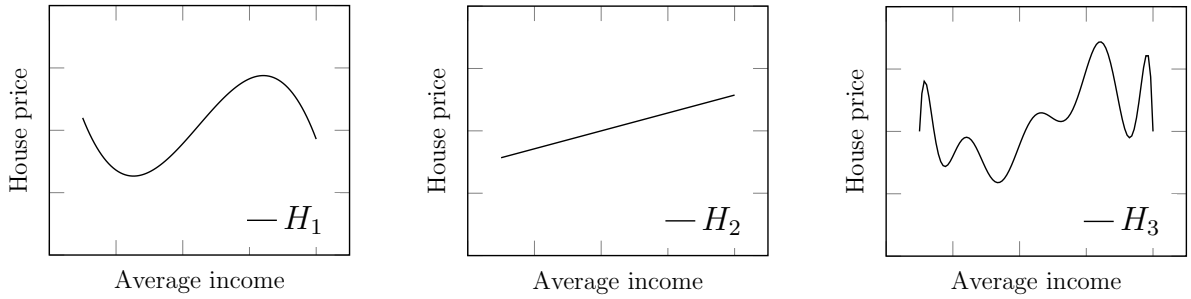

Figure 2: Three curves describing house prices as a function of average income

between house prices and average income. In particular, we are considering how likely it is that the survey produces the body of data, E (see Figure 3). Finally, suppose that another, more talented research team then inform us of (a) the results of a comprehensive study they conducted on the relationship between house prices and average income, and (b) the respective BIC scores of the curves H_1, H_2 and H_3 with respect to E. There are a number of possibilities. First, they could tell us that (i) it turns out that H_1 is actually the 'true curve' that accurately describes the relationship between house prices and average income, and (ii) H_1 has the best BIC score for E. Alternatively, they might tell us that (i) H_1 is the 'true curve' that accurately describes the relationship between

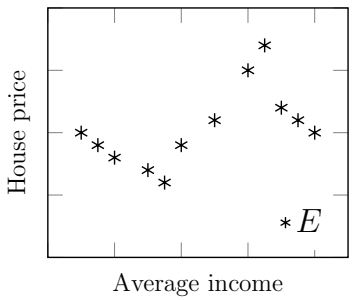

Figure 3: Data regarding house prices and average income

house prices and average income, and (ii') H_2 has the best BIC score for E. It seems clear that we should view E as a more likely outcome in the case where we learn (ii) than we should in the case where we learn (ii'). For, in the case where we learn (ii), we are told that E is best explained (equating explanatory virtue with a good BIC score) by what we know to be the true hypothesis, whereas in the case where we learn (ii') we are told that E is better accounted for by what we know to be a false hypothesis. Since we should expect to observe evidence that is well accounted for by what we know to be the true hypothesis, we should believe E to a higher degree in the former case than we do in the later case. This intuition is exactly what is captured by the following basic constraint on the parameters of the network:[9]

$$\alpha \geq \beta \quad , \quad \delta > \gamma \tag{3}$$

To reiterate, the inequalities in (3) simply state that we should view E as more likely to be true when it is best explained by what we know to be the true hypothesis compared to when it is best explained by what we know to be a false hypothesis. Overall, our model is characterised by the following basic constraints:

A1: The variable H is (unconditionally) independent of X:

$$H \perp\!\!\!\perp X \tag{4}$$

A2: The relation among the conditional probabilities is as follows:

$$\alpha \geq \beta \quad , \quad \delta > \gamma \tag{5}$$

[9]Note that the reasoning here can equally be taken to motivate the strict inequality $\alpha > \beta$, but it turns out that we only need the weaker condition $\alpha \geq \beta$.

Together, **A1** and **A2** are jointly sufficient to guarantee the following result (all proofs in Appendix):

Theorem 1 *If* **A1** *and* **A2** *hold, then* X *adds to* E*'s confirmation of* H*, i.e.* $P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) > P(\mathrm{H} \mid \mathrm{E})$.

And this is exactly what we wanted to show. Theorem 1 tells us that H's being the best explanation of the novel evidence E adds an additional confirmational 'boost' that would be absent if we neglected explanatory considerations. Furthermore, the confirmatory import of explanatory virtue has been explicated within a purely Bayesian framework. It is still assumed that agents update by standard Bayesian conditionalization. Thus, Theorem 1 allows us to respect the Bayesian's commitment to updating by conditionalization whilst also taking into account (i) the fact, demonstrated by Douven and Schupbach (2015), that explanatory considerations seem to play a crucial role in the way that people update their beliefs in everyday reasoning contexts, and (ii) Douven's (2013) observation that there are situations where ideal agents who take explanatory considerations into account will do better than their Bayesian counterparts. We interpret Theorem 1 as providing a demonstration that it is possible to take into account the significant confirmatory import of explanatory considerations without surrendering any fundamental features of the Bayesian framework.

2.3 Generalising the Model

Until now, we have dealt with only one hypothesis (H). We now extend our model to include n hypotheses. We aim at showing that $P(\mathrm{H_i} \mid \mathrm{E}, \mathrm{X_i}) > P(\mathrm{H_i} \mid \mathrm{E})$ for $1 \leq i \leq n$.[10] Apart from assumption **A1** from above (generalised in the obvious way), a different but related assumption to **A2** is needed to prove this inequality.

Let $\alpha_{ij} := P(\mathrm{E} \mid \mathrm{H_i}, \mathrm{X_j})$, $h_i := P(\mathrm{H_i})$, and $x_i := P(\mathrm{X_i})$, where $1 \leq i, j \leq n$.

A2': There is a least one pair (k, r) with $k \in \{1, \ldots, n\} \setminus \{i\}$ and $r \in \{1, \ldots, n\} \setminus \{i\}$ where $\alpha_{ii} \geq \alpha_{ir}$ and $\alpha_{kr} > \alpha_{ki}$. For all other pairs (l, m) with $l \in \{1, \ldots, n\} \setminus \{i, k\}$ and $m \in \{1, \ldots, n\} \setminus \{i, r\}$ it holds that $\alpha_{ii}\, \alpha_{lm} \geq \alpha_{im}\, \alpha_{li}$.

[10] Where $\mathrm{X_i}$ is the proposition 'Of all the currently available hypotheses, $\mathrm{H_i}$ would be the best explanation of E, were E to obtain'.

Notice that **A2'** collapses into **A2** when we have one hypothesis (i.e. H): $\alpha_{ii} \geq \alpha_{ir}$ and $\alpha_{kr} > \alpha_{ki}$ become $\alpha \geq \beta$ and $\delta > \gamma$. The motivation for **A2'** is directly analogous to the motivation for **A2** described above, i.e. that we should expect evidence which is well explained by the true hypothesis.

Using assumptions **A1** and **A2'** one can prove the following theorem:

Theorem 2 *If* **A1** *and* **A2'** *hold, then* $\mathrm{X_i}$ *adds to* E*'s confirmation of* $\mathrm{H_i}$*, that is* $P(\mathrm{H_i} \mid \mathrm{E}, \mathrm{X_i}) > P(\mathrm{H_i} \mid \mathrm{E})$.

Theorem 2 generalizes Theorem 1 and it says that in cases of n explanations, $\mathrm{X_i}$ provides additional confirmation to $\mathrm{H_i}$ if **A1** and **A2'** hold.

2.4 What We Haven't Done

At this stage, it is important to clarify exactly what we take the philosophical import of Theorem 1 (and Theorem 2) to be. At the fundamental level, the novel insight here is that it is possible to think of explanatory considerations as providing a confirmatory 'boost', as described by Douven (2013), without abandoning conditionalization and thereby becoming liable to a dynamic Dutch book. This allows the Bayesian to account for the empirical fact that people appear to rely on explanatory considerations when updating on novel evidence, and it also means that the Bayesian can respond to Douven's observation that there are situations where agents who reason in accordance with IBE outperform their Bayesian counterparts. For, according to the model described above, the difference between the two update strategies described by Douven (2013) is not a difference between a Bayesian agent on the one hand and a non-Bayesian agent on the other. Rather, it is a difference between a Bayesian who fails to take into account relevant explanatory considerations and another (possibly non-Bayesian) who does take those considerations into account. The problem is not with conditionalization as an update rule, but rather with the fact that the Bayesian agent is ignoring the explanatory virtues of the relevant hypotheses.

It is important to note that we do not take ourselves to have contributed to the debate concerning the nature of the explanatory virtues. In particular, we have said nothing about what makes a hypothesis a good explanation of some given evidence. In the justification of our model, we equated explanatory virtue with the Bayesian information

score because, *whatever one thinks about the nature of explanatory virtue*, it is natural to think that curves with lower information scores count as better explanations of the relevant evidence. We intend to remain ecumenical about the nature of explanatory virtue, and merely take ourselves to be providing conditions under which explanatory considerations can contribute to confirmation.

3 Explanation, Confirmation, and Bad Lots

In this section, we turn to addressing the relationship between the model described here and some well known criticisms of IBE from the literature. We start with van Fraassen's 'bad lot' objection.

3.1 Bad Lots and Bad Explanations

An implicit assumption of our model is that the hypotheses being considered are mutually exclusive and, more importantly, jointly exhaustive, i.e. their probabilities always sum to 1. As we mentioned in the introduction, this is a strong assumption that is not always justified. There is no guarantee that we are not starting with a bad lot. This problem can be resolved by assuming that one of the hypotheses being considered is a 'catch-all', i.e. the negation of the disjunction of all the other hypotheses (see Niiniluoto 1999: S447–S448). Thus, the variable H will have $n+1$ values, the first n of which correspond to the hypotheses under consideration. The $n+1$'th value, denoted 'H_C', corresponds to the proposition:

H_C: All of the considered hypotheses are false.

It is clear that this guarantees that the probability of the values of H will sum to 1. However, a new problem arises at this point. Specifically, if we let H_C denote the catch-all proposition, we need to provide a suitable interpretation of the corresponding value X_C of X. We cannot simply interpret X_C as the proposition 'H_C would provide the best available explanation of E', since this proposition will generally have zero probability. Catch-all hypotheses are paradigms of explanatory vice. The negation of some finite set of scientific hypotheses is generally going to fail to provide a satisfactory explanation of any non-trivial evidence. Also, note that X_C always having zero probability is incompatible

with the basic assumptions of our model, since we want to allow for the situation where conditioning on X_C raises the probability of H_C when we have already conditioned on E.

To solve this problem, we follow Lipton (2004: 59) (the two-filter process) and Musgrave (1988: 238–239) (the minimal adequacy condition) in arguing that IBE can only be applied in situations where, as well as knowing that the relevant hypothesis provides the best explanation of the evidence E, we also know that the hypothesis provides a *good enough* explanation of E. This requires that for each hypothesis $H_i \neq H_C$, we interpret the value X_i of X as the proposition:

X_i: Of all the currently available hypotheses, H_i would be the best explanation of E, were E to obtain, *and* H_i provides a sufficiently good explanation of E

Under this interpretation, we can then think of X_C as the proposition:

X_C: None of the currently available hypotheses provide a sufficiently good explanation of E.

It is easy to see that this slight shift in the interpretation of the variables does not interfere with the original philosophical motivations for the constraints on our model. However, another issue does arise here. Specifically, we now need to specify what counts as a 'sufficiently good explanation'. We leave the provision of a detailed answer to this problem for another day, and restrict ourselves to the following observation. Of course, the notion of a 'sufficiently good' explanation is fundamentally a vague one, and it seems unlikely that one can provide a principled specification of the threshold of explanatory virtue above which an explanation counts as 'sufficiently good'. However, the mere fact that the notion is a vague one does not mean that it cannot be the subject of probabilistic partial beliefs. One may not be sure whether or not string theory gives a good explanation of the isotropy of the cosmic microwave background, but one may be more confident that it does so than one is that non-relativistic quantum mechanics does. Similarly, one may be more confident that the person one sees in the distance is tall than one is that their T-shirt is red, even though both of the relevant concepts are inherently vague.

Once one reinterprets the values of the variables in the way described above, the bad lot objection ceases to be a problem. We now have a justification for the assumption that the hypothesis space is exhaustive. Moreover, we also obtain the following desirable result: learning that none of the hypotheses being considered provide a sufficiently good

explanation of the evidence E gives a confirmatory boost to the catch-all hypothesis H_C with respect to E. Thus, our model can capture the intuition that IBE plays an important role not just in the context of justification, but also in the context of discovery, as pointed out by Lipton (2004: 67) and Okasha (2000: 706–707). If H_C receives a confirmatory boost, then none of the hypotheses under consideration is sufficiently good, hence the best explanation of the evidence E lies outside of the space of considered hypotheses and one should look for a new, not yet considered, explanation of the evidence E. Thus, our Bayesian model of IBE not only provides a novel justification of IBE, it also captures the fact that an agent using IBE can respond to the new evidence by inventing a new hypothesis. This allows us to respond to Okasha (2000: 707), who argues that Bayesianism is silent when it comes to the context of discovery.

3.2 Explanatoriness is not Confirmatory

Roche and Sober (2013) argue that explanatory considerations are incapable of adding to the confirmatory support of novel evidence. More specifically they argue that $P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}') = P(\mathrm{H} \mid \mathrm{E})$, where H and E are as in our model and X' expresses a counterfactual: if H and E were true, then H would explain E. Roche and Sober argue for that claim by considering the following example. Let H be the hypothesis that S smokes at least 10,000 cigarettes before age 50, and let E be the evidence that S gets lung cancer after age 50. X' then says that if S smoked 10,000 cigarettes before age 50 and S got lung cancer after age 50, then the smoking would explain the lung cancer. Observing a large sample of people that developed lung cancer after age 50 and counting how many of these individuals smoked at least 10,000 cigarettes before age 50, scientists can estimate $P(\mathrm{H} \mid \mathrm{E})$ and find it to be c. Roche and Sober then claim that conditioning additionally on X' clearly does not change that estimate c, which is determined purely by the observed frequencies. Therefore, $P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}') = P(\mathrm{H} \mid \mathrm{E})$.

However, this example is a very special case, where the likelihood $P(\mathrm{H} \mid \mathrm{E})$ is determined by clear and well defined frequencies. Generally, subjective Bayesian probabilities are not straightforwardly given by approximations to long run frequencies. Although we agree that explanatory considerations may be inert in the special cases where degrees of belief are fixed by observed frequencies, this does not speak against the applicability of our analysis in the much more general case where subjective degrees are not fixed by an

objective standard of this type. A similar response is developed in much more detail by Lange (2017).

3.3 Multiple Plausible Rivals

Dellsén (2017a) argues that IBE ignores the fact that an inference to a hypothesis H may be undermined by the availability of competing explanatory hypotheses. To illustrate, suppose that at time t_1 there were five hypothesis in our hypothesis space. Three of these are then refuted by new observations, so that at a later time t_2 there are only two remaining hypotheses under consideration. Intuitively, the two remaining hypotheses are more plausible at t_2 than they were at t_1, since they have less competitors. However, how well each hypothesis explains the evidence remains invariant between t_1 and t_2, and Dellsén argues that IBE is unable to incorporate this intuitive insight, since it admonishes us to infer the truth of hypotheses based on their explanatory virtues, which remain constant between t_1 and t_2.

Our model of IBE naturally bypasses criticisms of this type. For, on our approach, one does not simply infer the truth (or probable truth) of the hypothesis purely because it has the most explanatory virtue. Rather, explanatory virtues contribute to the extent to which hypotheses are confirmed by novel evidence. In the previous example, the probability of the two remaining hypotheses will generally increase between t_1 and t_2 in our model, regardless of the fact that their explanatory 'loveliness' remain constant. However, the extent to which the hypotheses explain novel evidence will still contribute to the degree to which they are confirmed by that evidence. There is no tension here, and the problem arises from an overly simplistic understanding of the role of IBE in scientific inference.

4 Conclusion

Overall then, we have presented a novel Bayesian justification of IBE style inferences. Specifically, we have argued that explanatory considerations can add to the confirmatory power of novel evidence in a way that is perfectly compatible with Bayesian conditionalization. This approach has a number of significant virtues. Firstly, it allows us to resolve a number of famous criticisms of explanatory reasoning including, for example,

van Fraassen's dynamic Dutch book and bad lot arguments. Secondly, it also accounts for the important role that IBE plays in the context of discovery, which many considered to be an advantage of IBE that Bayesianism cannot account for. The Bayesian model employed here can be thought of as 'probabilifying' explanatory considerations and providing a precise mechanism that explicates the role that (best) explanation plays in Bayesian updating. We conclude that explanatory considerations do indeed play a significant part in scientific reasoning, but they do so in a way that is perfectly compatible with standard Bayesian epistemology.

A Proofs

A.1 Theorem 1

We want to show that $P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) > P(\mathrm{H} \mid \mathrm{E})$, i.e. $P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) - P(\mathrm{H} \mid \mathrm{E}) > 0$. Applying the theory of Bayesian networks to Figure 1 and using assumption **A1** ($H \perp\!\!\!\perp X$), we calculate:

$$\begin{aligned} P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) &= \frac{P(\mathrm{H}, \mathrm{E}, \mathrm{X})}{P(\mathrm{E}, \mathrm{X})} \\ &= \frac{P(\mathrm{H})\, P(\mathrm{X})\, P(\mathrm{E} \mid \mathrm{H}, \mathrm{X})}{P(\mathrm{X}) \sum\limits_{H} P(H)\, P(\mathrm{E} \mid H, \mathrm{X})} \\ &= \frac{h\, x\, \alpha}{x\, (h\, \alpha + \overline{h}\, \gamma)} \\ &= \frac{h\, \alpha}{h\, \alpha + \overline{h}\, \gamma} \end{aligned}$$

$$\begin{aligned} P(\mathrm{H} \mid \mathrm{E}) &= \frac{P(\mathrm{H}, \mathrm{E})}{P(\mathrm{E})} \\ &= \frac{P(\mathrm{H}) \sum_X P(X) P(\mathrm{E} \mid \mathrm{H}, X)}{\sum_{H,X} P(H) P(X) P(\mathrm{E} \mid H, X)} \\ &= \frac{h\, (x\, \alpha + \overline{x}\, \beta)}{h\, (x\, \alpha + \overline{x}\, \beta) + \overline{h}\, (x\, \gamma + \overline{x}\, \delta)} \end{aligned}$$

Hence,

$$\begin{aligned}
P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) - P(\mathrm{H} \mid \mathrm{E}) &= \frac{h\,\alpha}{h\,\alpha + \overline{h}\,\gamma} - \frac{h\,(x\,\alpha + \overline{x}\,\beta)}{h\,(x\,\alpha + \overline{x}\,\beta) + \overline{h}\,(x\,\gamma + \overline{x}\,\delta)} \\
&= \frac{h\,\alpha\left[h\,(x\,\alpha + \overline{x}\,\beta) + \overline{h}\,(x\,\gamma + \overline{x}\delta)\right] - h\,(x\,\alpha + \overline{x}\,\beta)\,(h\,\alpha + \overline{h}\gamma)}{(h\,\alpha + \overline{h}\,\gamma)\left[h\,(x\,\alpha + \overline{x}\,\beta) + \overline{h}\,(x\,\gamma + \overline{x}\,\delta)\right]} \\
&= h\,\frac{\overline{h}\,\alpha\,(x\,\gamma + \overline{x}\,\delta) + (x\,\alpha + \overline{x}\,\beta)\,(h\,\alpha - h\,\alpha - \overline{h}\,\gamma)}{(h\,\alpha + \overline{h}\,\gamma)\left[h\,(x\,\alpha + \overline{x}\,\beta) + \overline{h}\,(x\,\gamma + \overline{x}\,\delta)\right]} \\
&= h\,\overline{h}\,\frac{x\,\alpha\,\gamma + \overline{x}\,\alpha\,\delta - x\,\alpha\,\gamma - \overline{x}\,\beta\,\gamma}{(h\,\alpha + \overline{h}\,\gamma)\left[h\,(x\,\alpha + \overline{x}\,\beta) + \overline{h}\,(x\,\gamma + \overline{x}\,\delta)\right]} \\
&= \overline{x}\,h\,\overline{h}\,\frac{\alpha\,\delta - \beta\,\gamma}{(h\,\alpha + \overline{h}\,\gamma)\left[h\,(x\,\alpha + \overline{x}\,\beta) + \overline{h}\,(x\,\gamma + \overline{x}\,\delta)\right]}.
\end{aligned}$$

Assumption **A2** ($\alpha \geq \beta$ and $\delta > \gamma$) entails that $\alpha\,\delta - \beta\,\gamma$ is strictly positive. Therefore, $P(\mathrm{H} \mid \mathrm{E}, \mathrm{X}) - P(\mathrm{H} \mid \mathrm{E})$ is strictly positive.

A.2 Theorem 2

Similarly as in the proof of Theorem 1, we use assumption **A1** and additionally allow that there are n hypotheses. We define $\alpha_{ij} := P(\mathrm{E} \mid \mathrm{H_i}, \mathrm{X_j})$.

$$\begin{aligned}
P(\mathrm{H_i} \mid \mathrm{E}, \mathrm{X_i}) &= \frac{P(\mathrm{H_i}, \mathrm{E}, \mathrm{X_i})}{P(\mathrm{E}, \mathrm{X_i})} \\
&= \frac{x_i\,h_i\,\alpha_{ii}}{x_i \sum\limits_j h_j\,\alpha_{ji}} \\
&= \frac{h_i\,\alpha_{ii}}{\sum\limits_j h_j\,\alpha_{ji}}
\end{aligned}$$

$$\begin{aligned}
P(\mathrm{H_i} \mid \mathrm{E}) &= \sum_j P(\mathrm{H_i} \mid \mathrm{E}, \mathrm{X_j}) P(\mathrm{X_j} \mid \mathrm{E}) \\
&= \frac{h_i\,\alpha_{ii}}{\sum\limits_j h_j\,\alpha_{ji}} \frac{P(\mathrm{X_i}, \mathrm{E})}{P(\mathrm{E})} + \left(\sum_{k \neq i} \frac{h_i\,\alpha_{ik}}{\sum\limits_j h_j\,\alpha_{jk}} \frac{P(\mathrm{X_k}, \mathrm{E})}{P(\mathrm{E})} \right) \\
&= \frac{h_i\,\alpha_{ii}}{\frac{P(\mathrm{X_i}, \mathrm{E})}{x_i}} \frac{P(\mathrm{X_i}, \mathrm{E})}{P(\mathrm{E})} + \frac{h_i}{P(\mathrm{E})} \left(\sum_{k \neq i} \alpha_{ik} \frac{P(\mathrm{X_k}, \mathrm{E})}{\frac{P(\mathrm{X_k}, \mathrm{E})}{x_k}} \right) \\
&= \frac{h_i}{P(\mathrm{E})} \left(x_i\,\alpha_{ii} + \sum_{k \neq i} x_k\,\alpha_{ik} \right)
\end{aligned}$$

$$= \frac{h_i}{\sum_{p,q} h_p\, x_q\, \alpha_{pq}} \left(\sum_j x_j\, \alpha_{ij} \right)$$

Let $\mathrm{K} := P(\mathrm{H_i} \mid \mathrm{E}, \mathrm{X_i}) - P(\mathrm{H_i} \mid \mathrm{E})$. We then calculate:

$$\begin{aligned}
\mathrm{K} &= \frac{h_i\, \alpha_{ii}}{\sum_j h_j\, \alpha_{ji}} - \frac{h_i}{\sum_{p,q} h_p\, x_q\, \alpha_{pq}} \left(\sum_j x_j\, \alpha_{ij} \right) \\
&= h_i \left[\alpha_{ii} \left(\sum_{p,q} h_p\, x_q\, \alpha_{pq} \right) - \left(\sum_j x_j\, \alpha_{ij} \right) \left(\sum_j h_j\, \alpha_{ji} \right) \right] \cdot G^{-1} \\
&= h_i \left[h_i\, \alpha_{ii} \left(\sum_j x_j\, \alpha_{ij} \right) + \alpha_{ii} \left(\sum_{k \neq i,j} h_k\, x_j \alpha_{kj} \right) - \left(\sum_j x_j\, \alpha_{ij} \right) \left(\sum_j h_j\, \alpha_{ji} \right) \right] \cdot G^{-1} \\
&= h_i \left[\left(\sum_j x_j\, \alpha_{ij} \right) \left(h_i\, \alpha_{ii} - h_i\, \alpha_{ii} - \left(\sum_{k \neq i} h_k\, \alpha_{ki} \right) \right) + \alpha_{ii} \left(\sum_{k \neq i,j} h_k\, x_j \alpha_{kj} \right) \right] \cdot G^{-1} \\
&= h_i \left[\alpha_{ii} \left(\sum_{k \neq i,j} h_k\, x_j \alpha_{kj} \right) - \left(\sum_j x_j\, \alpha_{ij} \right) \left(\sum_{k \neq i} h_k\, \alpha_{ki} \right) \right] \cdot G^{-1} \\
&= h_i \left[\alpha_{ii}\, x_i \left(\sum_{k \neq i} h_k\, \alpha_{ki} \right) + \alpha_{ii} \left(\sum_{k \neq i, r \neq i} h_k\, x_r\, \alpha_{kr} \right) - \alpha_{ii}\, x_i \left(\sum_{k \neq i} h_k\, \alpha_{ki} \right) \right. \\
&\qquad \left. - \left(\sum_{k \neq i} x_k\, \alpha_{ik} \right) \left(\sum_{k \neq i} h_k\, \alpha_{ki} \right) \right] \cdot G^{-1} \\
&= h_i \left[\alpha_{ii} \left(\sum_{k \neq i, r \neq i} h_k\, x_r\, \alpha_{kr} \right) - \left(\sum_{k \neq i} x_k\, \alpha_{ik} \right) \left(\sum_{k \neq i} h_k\, \alpha_{ki} \right) \right] \cdot G^{-1} \\
&= h_i \left(\sum_{k \neq i, r \neq i} h_k\, x_r\, \alpha_{ii}\, \alpha_{kr} - \sum_{k \neq i, r \neq i} h_k\, x_r\, \alpha_{ir} \alpha_{ki} \right) \cdot G^{-1} \\
&= h_i \left(\sum_{k \neq i, r \neq i} h_k\, x_r\, (\alpha_{ii}\, \alpha_{kr} - \alpha_{ir}\, \alpha_{ki}) \right) \cdot G^{-1},
\end{aligned}$$

where $G := \left(\sum_j h_j\, \alpha_{ji} \right) \left(\sum_{p,q} h_p\, x_q\, \alpha_{pq} \right)$.

Assumption **A2'** entails that the difference $\alpha_{ii}\, \alpha_{kr} - \alpha_{ir}\, \alpha_{ki}$ is non-negative and that it is strictly positive for at least one pair (k, r). Therefore, K is strictly positive.

References

Bovens, L., & S. Hartmann (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Brössel, P. (2015). On the Role of Explanatory and Systematic Power in Scientific Reasoning. *Synthese* 192: 3877–3913.

Cabrera, F. (2017). Can There be a Bayesian Explanationism? On the Prospects of a Productive Partnership. *Synthese* 194: 1245–1272.

Day, T. & H. Kincaid (1994). Putting Inference to the Best Explanation in its Place. *Synthese* 98: 271–295.

Dawid, R., S. Hartmann & J. Sprenger (2015). The No Alternatives Argument. *The British Journal for the Philosophy of Science* 66: 213–234.

Dellsén, F. (2017a). Abductively Robust Inference. *Analysis*, DOI: 10.1093/analys/anx049

Dellsén, F. (2017b). Reactionary Responses to the Bad Lot Objection. *Studies in History and Philosophy of Science* 61: 32–40.

Douven, I. (2002). Testing Inference to the Best Explanation. *Synthese* 130(3): 355–377.

Douven, I. (2011). Abduction. In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/abduction/.

Douven, I. (2013). Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation. *Philosophical Quarterly* 63(252): 428–444.

Douven, I. & J. Schupbach (2015). The Role of Explanatory Considerations in Updating. *Cognition* 142: 299–311.

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

Farmakis, L. & S. Hartmann (2005). Review of *Inference to the Best Explanation* by Peter Lipton. *Notre Dame Philosophical Reviews*.

Glymour, C. (1984). Explanation and Realism. In: J. Leplin (Ed.), *Scientific Realism*. Berkeley, CA: University of California Press, pp. 173–192.

Halley, E. (1752). *Astronomical Tables with Precepts, both in English and Latin, for Computing the Places of the Sun, Moon, Planets, and Comets*. London: Printed for William Innys.

Háyek, A., & S. Hartmann (2010). Bayesian Epistemology. In: J. Dancy, E. Sosa, & M. Steup (eds.): *A Companion to Epistemology*. Oxford: Wiley-Blackwell, pp. 93–105.

Harman, G. H. (1965). The Inference to the Best Explanation. *The Philosophical Review* 74(1): 88–95.

Harman, G. H. (1967). Detachment, Probability, and Maximum Likelihood. *Noûs* 1(4): 401–411.

Hartmann, S. & J. Sprenger (2011). Bayesian Epistemology. In: S. Bernecker, & D. Pritchard (eds.): *The Routledge Companion to Epistemology*. New York: Routledge, pp. 609–620.

Henderson, L. (2015). Bayesianism and Inference to the Best Explanation. *The British Journal for the Philosophy of Science* 65: 687–715.

Howson, C. (1991). The 'Old Evidence' Problem. *The British Journal for the Philosophy of Science* 42(4): 547–555.

Lange, M. (2017). The Evidential Relevance of Explanatoriness: A Reply to Roche and Sober. *Analysis*, DOI: 10.1093/analys/anx045

Laplace, P. S. (1995/1825). *Philosophical Essay on Probabilities*. New York, NY: Springer.

Lipton, P. (2001). Is Explanation a Guide to Inference? A Reply to Wesley C. Salmon. In: G. Hon & S. Rakover (eds.): *Explanation: Theoretical Approaches and Applications*. Dordrecht: Kluwer Academic Publishers, pp. 93–120.

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

Musgrave, A. (1988). The Ultimate Argument for Scientific Realism. In: R. Nola (ed.): *Relativism and Realism in Science*. Dordrecht: Kluwer, pp. 229–252.

Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.

Niiniluoto, I. (1999). Defending Abduction. *Philosophy of Science* 66: S436–S451.

Norton, J. (2016a). Inference to the Best Explanation: The General Account. In: J. Norton, *The Material Theory of Induction*. In preparation.

Norton, J. (2016b). Inference to the Best Explanation: Examples. In: J. Norton, *The Material Theory of Induction*. In preparation.

Okasha, S. (2000). van Fraassen's Critique of Inference to the Best Explanation. *Studies in History and Philosophy of Science* 31(4): 691–710.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Psillos, S. (2000). Abduction: Between Conceptual Richness and Computational Complexity. In: P. Flach & A. Kakas (eds.), *Abduction and Induction*. Dordrecht: Springer, pp. 59–74.

Psillos, S. (2002). Simply the Best: A Case for Abduction. In: A. Kakas & F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond*. Berlin: Springer, pp. 605–625.

Psillos, S. (2004). Inference to the Best Explanation and Bayesianism. In: F. Stadler (ed.), *Induction and Deduction in the Sciences*. Dordrecht: Kluwer, pp. 83–91.

Psillos, S. (2007). The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research* 74(2): 441–448.

Roche, W. & E. Sober (2013). Explanatoriness is Evidentially Irrelevant, or Inference to the Best Explanation Meets Bayesian Confirmation Theory. *Analysis* 73(4): 659–668.

Salmon, W. (2001). Reflections of a Bashful Bayesian: A Reply to Peter Lipton. In: G. Hon & S. Rakover (eds.), *Explanation: Theoretical Approaches and Applications*. Dordrecht: Kluwer Academic Publishers, pp. 93–120.

Schupbach, J. N. (2013). Is the Bad Lot Objection Just Misguided? *Erkenntnis* 79: 55–64.

Teller, P. (1973). Conditionalization and Observation. *Synthese* 26: 218–258.

Thagard, P. R. (1978). The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy* 75(2): 76–92.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

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

Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese* 167(1): 125–143.

Williamson, T. (2016). Abductive Philosophy. *Philosophical Forum* 47: 263–280.