

## Can There Be a Bayesian Explanationism?: On the Prospects of a Productive Partnership

Abstract: In this paper, I consider the relationship between Inference to the Best Explanation (IBE) and Bayesianism, both of which are well-known accounts of the nature of scientific inference. In section 1, I give a brief overview of Bayesianism and IBE. In section 2, I argue that IBE in its most prominently defended forms is difficult to reconcile with Bayesianism because not all of the items that feature on popular lists of “explanatory virtues”—by means of which IBE ranks competing explanations—have confirmational import. Rather, some of the items that feature on these lists are “informational virtues”—properties that do not make a hypothesis  $H_1$  more probable than some competitor  $H_2$  given evidence  $E$ , but that, roughly-speaking, give that hypothesis greater informative content. In section 3, I consider as a response to my argument a recent version of compatibilism which argues that IBE can provide further normative constraints on the objectively correct probability function. I argue that this response does not succeed, owing to the difficulty of defending with any generality such further normative constraints. Lastly, in section 4, I propose that IBE should be regarded, not as a theory of scientific inference, but rather as a theory of when we ought to “accept”  $H$ , where the acceptability of  $H$  is fixed by the goals of science and concerns whether  $H$  is worthy of commitment as research program. In this way, IBE and Bayesianism, as I will show, can be made compatible, and thus the Bayesian and the proponent of IBE can be friends.

**Key words:** Inference to the Best Explanation; IBE; Bayesianism; Explanatory Virtues; Values in Science

## Introduction

Despite the popularity of both Inference to the Best Explanation (IBE) and Bayesianism as accounts of the nature of scientific inference, up until recently there has been remarkably little dialogue between proponents of IBE—i.e. “IBE-ists”—and Bayesians. Perhaps this lack of serious engagement between IBE-ists and Bayesians can be attributed to the fact that prominent supporters of IBE (Psillos 2004, 2007) and Bayesianism (van Fraassen 1989) have urged that the two views are incompatible. However, since the Bayesian approach and IBE are both *prima facie* plausible ways in which to understand the nature of scientific inference, it would be undesirable if IBE and Bayesianism were, upon final analysis, fundamentally incompatible. While Bayesianism is no doubt the dominant theory of confirmation, IBE plays a central role in debates in general philosophy of science, such as the realism/anti-realism debate and the problem of under-determination.<sup>1</sup> Thus, since philosophers of science have appealed to IBE as a way to tackle important problems, we should want to know how IBE connects up with Bayesianism. Given the apparent virtues of the two accounts, it would be most attractive if IBE and Bayesianism were not only compatible, but, in the end, mutually reinforcing. Ideally, IBE-ist ideas would be of use to Bayesians in solving some of the problems raised by critics and vice versa. As I will argue, however, certain features of IBE, as the view is articulated by its prominent proponents, renders a compatibilist view difficult to defend. In broad outline, this is because not all the virtues that feature on the lists of explanatory virtues upheld by IBE-ists have confirmational import. In light of this, I suggest that we ought not to view IBE as a theory of scientific inference at all, but rather as a theory of when a hypothesis H is “acceptable”, where, roughly-speaking, this concerns whether H is worthy of commitment as research program.

---

<sup>1</sup> Traditionally, defenders of scientific realism who rely on the so-called “No-Miracles Argument” formulate their defense of realism as an inference to the best explanation, e.g. Putnam (1975), Boyd (1983), Psillos (1999). Moreover, appeals to explanatory virtues such as simplicity are often made as a kind of tie-breaker in cases in which theory is under-determined by data. See, for instance, Kukla (1994) for a critical discussion of the appeal to “non-empirical” virtues as a possible solution to the under-determination problem.

## 1. Bayesianism and IBE

Bayesianism has a central place in contemporary discussions of theory confirmation, rationality, evidence, etc., and so I will only briefly review some of its key features. First, Bayesians claim that rational agents can be thought of as having “degrees of belief” or “credences” with respect to propositions, a notion which corresponds intuitively to how strongly the agent believes a proposition. Second, these degrees of belief can be modeled as a probability distribution, satisfying the standard axioms of the mathematical theory of probability.<sup>2</sup> Third, for an agent to be rational, at any given time, her personal probability function (Pr-function) must be constrained so as to obey the axioms of probability.<sup>3</sup> Lastly, in addition to the synchronic requirement of obeying the axioms, most varieties of Bayesianism assert that a rational agent must also update her degrees of belief over time in light of new data by the rule of subjective conditionalization:<sup>4</sup>

For any hypothesis  $H$ , any body of evidence  $E$ , and any times  $t_i$  and  $t_j$ , where  $i < j$ , upon learning  $E$  at  $t_j$ ,  $\Pr_j(H|E) = \Pr_j(H)$ , where  $\Pr_i(H|E)$  is the probability of  $H$  given  $E$  at  $t_i$  and  $\Pr_i(H)$  is the probability of  $H$  at  $t_i$ .

One handy way of calculating the quantity  $\Pr(H|E)$  is by way of Bayes’ theorem, which, when considering two hypotheses  $H$  and  $\sim H$  is given by the equation:

$$\Pr(H|E) = \frac{\Pr(H)\Pr(E|H)}{\Pr(H)\Pr(E|H) + \Pr(\sim H)\Pr(E|\sim H)}^5$$

---

<sup>2</sup> Credences are tracked by and for some Bayesians conceptually identified with betting behavior. For instance, if the agent believes with certainty, i.e. with a credence of 1, that some proposition  $A$  is true, then she would be willing to purchase a bet that pays \$1 if  $A$  is true at a price  $\leq$  \$1, if she has a credence of 1/2 that  $A$  is true, then she would be willing to purchase a bet that pays \$1 if  $A$  is true at a price  $\leq$  \$.5, etc., if she has a credence of 1/4, then she would be willing to purchase such a bet at a price  $\leq$  \$.25, etc.

<sup>3</sup> Formally, the agent’s personal credences can be represented as a function from a set of propositions  $S$  to the set of real numbers  $R$ , such that for each proposition  $A \in S$ ,  $A$  satisfies the three axioms of probability theory: 1)  $\Pr(A) \geq 0$  (Non-Negativity), 2) If  $A$  is a tautology, then  $\Pr(A) = 1$  (Normality), and 3) If  $A$  is incompatible with some proposition  $B \in S$ , then  $\Pr(A \vee B) = \Pr(A) + \Pr(B)$  (Finite Additivity). Additionally the conditional probability of  $A$  given  $B$ , i.e.  $\Pr(A|B)$  satisfies, and for some Bayesians is defined by, the “Ratio Formula”:  $\Pr(A|B) = \frac{\Pr(A \& B)}{\Pr(B)}$ .

<sup>4</sup> Subjective conditionalization assumes that when some piece of evidence  $E$  is learned,  $\Pr(E) = 1$ . In cases in which the agent increases her credence in  $E$  to some value  $x$ , where  $x < 1$ , then the agent may update by a more general rule, called “Jeffrey conditionalization”:  $P_i(H) = P_i(H|E)P_i(E) + P_i(H|\sim E)P_i(\sim E)$ . See, Jeffrey (1983). Henceforth, I will neglect the issue of uncertain evidence and assume that whenever  $E$  is learned,  $\Pr(E) = 1$ .

<sup>5</sup> For a finite partition of hypotheses  $H_1, H_2, \dots, H_n$ , Bayes’ theorem can be generalized as:  $\Pr(H_i|E) = \frac{\Pr(H_i)\Pr(E|H_i)}{\sum_j \Pr(H_j)\Pr(E|H_j)}$

As a theory of scientific inference, Bayesianism includes an account of when the evidence  $E$  confirms a hypothesis  $H$ . The standard qualitative Bayesian account of confirmation is simple. For any  $H$  and any  $E$ ,  $E$  confirms  $H$  if and only if  $\Pr(H|E) > \Pr(H)$ .<sup>6</sup>

The view that the above two normative constraints are necessary and sufficient for a rational  $\Pr$ -function is often called “Subjective Bayesianism”. Some philosophers insist in addition that a rational  $\Pr$ -function must satisfy further constraints, such as, for instance, the Principal Principle (Lewis 1981), according to which one ought to set one’s credences to conform with known objective chances (with some caveats about admissible evidence). The more normative constraints placed on what counts as a rational  $\Pr$ -function, the more “objective” Bayesianism becomes. In the limit, “Objective Bayesianism” (e.g. Carnap 1962) is the name sometimes used for the position according to which there is only one rationally permissible  $\Pr$ -function.

Discussions of IBE trace back to Harman (1965), who appears to have coined the term.<sup>7</sup>

Harman characterizes IBE in the following way:

In making this inference one infers, from the fact that a certain hypothesis would explain the evidence, to the truth of that hypothesis. In general, there will be several hypotheses which might explain the evidence, so one must be able to reject all such alternative hypotheses before one is warranted in making the inference. Thus one infers, from the premise that a given hypothesis would provide a “better” explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true (1965: 89).

IBE is typically formalized in a four-step inference pattern (e.g. Psillos 2002; Lycan 1988), such as:

- (1)  $F$  is some fact or collection of facts
- (2) Hypothesis  $H_1$ , if true, would explain  $F$  sufficiently well.

---

<sup>6</sup> As an account of the nature of scientific inference, Bayesianism also includes some quantitative account of *how much*  $E$  confirms  $H$ . There are many options on offer. One example which is not without its problems is the “Difference Measure” (Carnap 1962: 21-22), according to which the degree to which  $E$  confirms  $H$  is given by  $\Pr(H|E) - \Pr(H)$ . A full account of scientific inference would include a quantitative measure as well, but we need not take any stand on this question here. See Fitelson (1999) for a discussion of the plurality of measures.

<sup>7</sup> As Harman (1965: 88-9) notes, something like IBE often goes by many names one of which is “abduction”, a term coined by C.S. Peirce. But despite some similarities between Peirce’s abduction and IBE, it has been argued that viewing abduction as the intellectual forbearer of IBE is not historically accurate, e.g. Minnameier (2004), Campos (2011). One key difference between Peirce’s abduction and IBE is that the former belongs to the context of discovery, while the latter belongs to the context of justification. For Peirce, it seems abduction is just the way one invents a new hypothesis, and so it is not warrant-conferring. For instance, Peirce writes: “[a]bduction is the process of forming explanatory hypotheses. It is the only logical operation which introduces any new idea” (CP 5.172).

- (3) No competing explanations  $H_2, H_3, \dots, H_n$  would explain  $F$  better than  $H_1$ .
- (4) Therefore, one is justified in believing that  $H_1$  is true over its competitors.

The most important step for the purpose at hand is the third. Most IBE-ists propose that we ought to rank the success of an explanation by reference to various so-called “explanatory virtues”. Accordingly,  $H_1$  is a good enough explanation to the extent that  $H_1$  possesses these virtues to a sufficiently high degree, and  $H_1$  is a better explanation than its competitors  $H_2, H_3, \dots, H_n$  to the extent that  $H_1$  exhibits these virtues to a higher degree than its competitors. Proponents of IBE differ with respect to what belongs on the list of explanatory virtues, although there is much overlap in their lists, and at times the differences are merely terminological.

An early source for understanding theory choice by way of a list of informal and imprecise, albeit intuitive, criteria can be found in the writings of Thomas Kuhn. Despite his not explicitly defending IBE, it is striking how much articulations of IBE have in common with the picture of theory choice sketched by Kuhn in the 1969 post-script to *The Structure of Scientific Revolutions* (1962) and in later works (e.g. his 1977). In rejecting the radical relativism that eager postmodernists had imputed to him, Kuhn insisted that genuine, scientific progress is possible. For Kuhn, such progress is mediated in a messy, non-linear fashion by a non-exhaustive list of virtues, including: *accuracy*, *consistency*, *scope*, *simplicity*, and *fruitfulness*, all of which constitute “standard criteria for evaluating the adequacy of a theory” and “...provide a shared basis for theory choice” (1977: 321-22).

Another early source for understanding theory choice in this manner is Quine and Ullian (1970), who despite also not explicitly defending IBE, endorse the following virtues: *conservatism*, which consists in not conflicting with already held beliefs; *modesty*, which corresponds to logical weakness; *simplicity*, which when dealing with mathematical equations, corresponds to the degree, or the order of an equation (e.g. Jeffreys 1931), but in other cases is harder to define; *generality*, which corresponds to the applicability of a hypothesis to wide-ranging circumstances; and, *refutability*, which corresponds to the familiar Popperian notion that a hypothesis should be in principle falsifiable.

As for self-avowed IBE-ists, Thagard (1978) defends three virtues: *consilience*, which corresponds to how much a theory explains and the extent to which it unifies disparate domains of evidence; *simplicity*, which corresponds to the size of the set of auxiliary assumptions needed by a hypothesis in order to explain the data, and perhaps the number of entities postulated; and *analogy*, which corresponds to a preference for old, already accepted mechanisms and processes. Similarly, Psillos (2002: 614-6) discusses the virtues of *consilience*, which corresponds to fit with background data; *completeness*, which corresponds to explaining all the relevant data; *importance*, which corresponds to explaining “salient” phenomena; *parsimony*, which corresponds to the size of the set of auxiliary assumptions needed by a hypothesis to explain the data; *unification*, which corresponds to bringing together disparate bodies of data under one explanatory hypothesis; and *precision*, which corresponds to articulating a precise causal-nomological mechanism to explain the data. Finally, Lipton mentions many of the same virtues, such as “mechanism, precision, scope, simplicity, fertility or fruitfulness, and fit with background belief” (2004: 122). In what follows, I will primarily focus on the virtues put forward by Lipton, since his book-length defense of IBE is the most comprehensive. For the sake of organization, below is a table connecting the terminology that I will use to discuss the virtues with the approximately equivalent terms used by other IBE-ists and philosophers given to virtue-talk:

<b>My Terminology</b>	<b>Other Terminology</b>
Scope	Generality (Quine and Ullian); Consilience (Thagard)
Explanatory power	Completeness (Psillos); Consilience (Thagard)
Mechanism	Precision (Psillos)
Simplicity	Parsimony (Psillos); Simplicity (Thagard)
Unification	Consilience (Thagard)
Fit with background data	Conservatism (Quine and Ullian); Consilience (Psillos); Analogy (Thagard)
Precision	Precision (Psillos)
Empirical Adequacy	Accuracy (Kuhn)

It is not my purpose here to discuss the intricacies of IBE, but one common objection to the view should be mentioned. Given that what constitutes an explanation is a matter of long-standing

debate,<sup>8</sup> one might think, as Salmon (2001a: 68) does, that IBE can really have “no clear meaning” until the disputes regarding the nature of explanation are resolved.<sup>9</sup> However, in my view, whether  $H_1$  constitutes an explanation according to one of the extant philosophical models— e.g. the Deductive-Nomological model (Hempel and Oppenheim 1948), the Statistical Relevance model (Salmon 1971), the Unificationist model (Friedman; 1974; Kitcher 1989), the Causal model (Salmon 1984; Woodward 2003), etc.— does not seem to do any real justificatory work. Rather, the feature that justifies any application of IBE is that the hypothesis does well with respect to the various virtues listed above. Presumably, any hypothesis that does well with respect to those virtues will be confirmed in accordance with IBE. Indeed, it seems right to say that a hypothesis counts as an explanation in this context just in case it does well with respect to the explanatory virtues. Whether that hypothesis constitutes an explanation according to one of the extant models of explanation is beside the point. Thus, it is no problem that proponents of IBE (e.g. Lipton 2004; Psillos 2002) want to stay with respect to the debates surrounding the various models of explanation.

## 2. The Central Difficulty for Compatibilism

Contemporary discussion of the relationship between IBE and Bayesianism begins with van Fraassen’s critique of IBE in his *Laws and Symmetry* (1989). There, van Fraassen argues that if a Bayesian agent also uses IBE, then the agent will be subject to a Dutch book, which is supposed to indicate that the agent is irrational.<sup>10</sup> To make his argument, van Fraassen combines IBE and Bayesianism by assigning a post-conditionalization probabilistic boost to hypotheses that are more explanatory. Clearly, if adopted as a rule of updating, this will lead to diachronic incoherence. Van Fraassen’s Dutch book argument against IBE has been criticized elsewhere. Generally, proponents

---

<sup>8</sup> See Salmon (1989) for an important critical survey of analyses of the concept of explanation in the 20th century.

<sup>9</sup> In a direct response, Lipton (2001: 100) resists this criticism, arguing that our grasp of the concept of an explanation is good enough, even in the absence of a widely agreed-upon philosophical account. It should be noted, Salmon (2001b: 122) concedes this point in a further response to Lipton.

<sup>10</sup> That is to say, if the agent adopts IBE as a rule of updating, then there exists a series of bets, each of which the agent, will individually sanction as fair, but which together logically guarantee a loss. See Roche and Sober (2013) who take themselves to be making a similar argument against IBE, although one that does not rely on pragmatic features.

of compatibilism with respect to IBE and Bayesianism (e.g. Okasha 2000; Lipton 2004; Weisberg 2009, etc.) take issue with the ad hoc way in which van Fraassen incorporates IBE into the Bayesian framework. Compatibilists have argued that there are other, more plausible ways to reconcile IBE with Bayesianism, and these views are not subject to a Dutch book argument.<sup>11</sup>

In this section, however, I will raise problems for the compatibilist project. One lacuna in contemporary discussions of the relationship between IBE and Bayesianism is that the indispensable role that the virtues are supposed to play in justifying any application of IBE is all but universally neglected.<sup>12</sup> If IBE is to be compatible with Bayesianism, then at the very least, the explanatory virtues that are upheld by IBE-ists must be truth-conducive, which is to say that the virtues must have some bearing on the probable truth of hypotheses that manifest them. As I will argue below, it is not the case that all the virtues on IBE-ists' lists of explanatory virtues meet this standard.

## 2.1 Reformulating the Question

According to the standard Bayesian account of confirmation, a hypothesis  $H$  is confirmed by a body of evidence  $E$  if and only if  $\Pr(H | E) > \Pr(H)$ . Presumably then, if we want to show that IBE is compatible with the Bayesian account of confirmation, we would need to show that, for each particular virtue  $V$  of some version of IBE, if a hypothesis  $H_1$  has  $V$  and some rival  $H_2$  lacks  $V$ , then, other things being equal,  $H_1$  will not only be confirmed by  $E$ , i.e.  $\Pr(H_1 | E) > \Pr(H_1)$ , but in addition, the degree of confirmation (DOC), according to some measure, of  $H_1$  will be higher than that of  $H_2$ .

However, one problem with attempting to show that the DOC of  $H_1$  is higher than the DOC for  $H_2$ , provided  $H_1$  has the virtues and  $H_2$  lacks them, is that often it is difficult to know what to say about the prior probability of the evidence. When we expand  $\Pr(H | E)$  in accordance with

---

<sup>11</sup> Here, I will pass over those discussions of the relationship between IBE and Bayesianism that explicitly reject the key assumptions of Bayesianism, e.g. Betz (2013: 3555), along with those that propose or defend versions of IBE that are straightforwardly incompatible with Bayesianism, e.g. Douven and Wenmackers (2015), Douven and Schupbach (2015).

<sup>12</sup> Hitchcock (2007) is an exception, rightly noting that a full reconciliation between Bayesianism and IBE "can only be achieved (if at all) in a piecemeal fashion, through the identification of factors that contribute to explanatory loveliness, and an exploration of the ways in which these factors might plausibly influence the probability assignments of a Bayesian agent" (439). As I argue below, the prospects for fulfilling this project look grim.



Bayes' theorem,  $\Pr(E)$  shows up in the denominator. We can further analyze  $\Pr(E)$ , since it follows from the Law of Total Probability that  $\Pr(E) = \Pr(H)\Pr(E|H) + \Pr(\sim H)\Pr(E|\sim H)$ . But it is difficult to make sense of the value  $\Pr(E|\sim H)$  in many cases. For example, if  $H =$  Newton's theory of gravity (NTG) and  $\sim H =$  Not-Newton's theory of gravity (Not-NTG), then  $\sim H$  is a long disjunction of every other alternative theory of gravity, including those that have yet to be thought of. If one has a value for  $\Pr(H)$ , then of course,  $\Pr(\sim H) = 1 - \Pr(H)$ . However, the value of  $\Pr(E|\sim H)$  is the probability that the disjunction of every alternative possible theory of gravity confers on the evidence. Without knowing what those possible alternative theories of gravity are, it is difficult to say anything about the value of  $\Pr(E|\sim H)$ . But, if that's true, then we can't determine the value of  $\Pr(E)$ . In this case, it will be quite difficult, if not impossible, to decide if  $H_1$  will have a higher DOC than  $H_2$ , provided that  $H_1$  has the virtues and  $H_2$  lacks them.

One way around this problem is to compare the posterior probabilities of  $H_1$  and  $H_2$  where  $H_1$  and  $H_2$  are incompatible hypotheses but not together exhaustive. By Bayes' theorem and some algebra it follows that  $\Pr(H_1|E) > \Pr(H_2|E)$  if and only if  $\Pr(H_1)\Pr(E|H_1) > \Pr(H_2)\Pr(E|H_2)$ . That is,  $H_1$  is better supported by the evidence  $E$  if and only if the product of the prior and the likelihood of  $H_1$  is greater than the product of the prior and likelihood of  $H_2$ . So, instead of trying to determine if it is true that for each virtue  $V$ , if  $H_1$  has  $V$  and  $H_2$  lacks  $V$ , then, other things being equal, the DOC of  $H_1$  is greater than  $H_2$ , we can try to answer the contrastive question concerning the relationship between  $\Pr(H_1|E)$  and  $\Pr(H_2|E)$ . Now we want to know: for each virtue  $V$ , if  $H_1$  has  $V$  and  $H_2$  lacks  $V$ , then, other things being equal, is it true that  $\Pr(H_1|E) > \Pr(H_2|E)$ ? This question is much more tractable than trying to decide if having the virtues will make  $H_1$  confirmed by  $E$  and afford  $H_1$  a higher degree of confirmation than  $H_2$ , which lacks those virtues.

With respect to the contrastive question, however, I argue that we should answer in the negative. In a discussion of Kuhn's view of theory choice in the context of Bayesianism, Salmon

(1990: 197) notes in passing that, while some of the criteria put forth by Kuhn, such as consistency and simplicity, seem to be confirmationally relevant, some of Kuhn's other criteria are not confirmational. Rather, some criteria are what Salmon calls "informational virtues", properties that do not make a hypothesis H more probable, but that, roughly-speaking, give H greater content.<sup>13</sup> Hypotheses that excel with respect to informational virtues say more about the world, but they are not made more probable by virtue of having these virtues. I claim that this feature that is of true Kuhn's list of virtues is likewise true of IBE-ist lists of virtues: only some of the IBE-ist virtues are confirmational virtues, whereas others are informational virtues. Since IBE mixes informational and confirmational virtues, it is difficult to see how IBE could be made compatible with Bayesianism.

## 2.2 The Confirmational Virtues

Of course, some of the virtues upheld by IBE-ists are relevant to whether  $\Pr(H_1 | E) > \Pr(H_2 | E)$ . These virtues are thus what I call "confirmational virtues". If the confirmational virtues are to make a difference in the comparative posteriors of  $H_1$  and  $H_2$ , then possession of any confirmational virtue must make it the case that, other things being equal,  $\Pr(H_1) > \Pr(H_2)$  or  $\Pr(E | H_1) > \Pr(E | H_2)$ .

As for the confirmational virtues, *simplicity* may be one of them, depending on what IBE-ists mean by simplicity. One notion of simplicity defended by IBE-ists relates to the number of auxiliary assumptions needed to derive a prediction by a theory (e.g. Thagard 1978; Psillos 2002). Other things being equal, this notion of simplicity has confirmational import. If  $\Pr(T_1) = \Pr(T_2)$ , and both  $T_1$  and  $T_2$  entail some evidence E, but if  $T_1$  requires a multitude of auxiliary assumptions  $A_1$ , which we have no good independent reasons to believe, then this will make  $\Pr(T_1 \& A_1)$  quite low. On the other hand, if  $T_2$  can derive E with a more modest set of auxiliaries  $A_2$ , then this fact will make it the case that  $\Pr(T_1 \& A_1) < \Pr(T_2 \& A_2)$ . Of course, whether  $\Pr(T_1 \& A_1) < \Pr(T_2 \& A_2)$  will depend on substantive background assumptions, in particular the nature of  $A_1$  and  $A_2$ . If  $A_2$  is a proper subset

---

<sup>13</sup> Salmon also claims that some of Kuhn's criteria are "economic" virtues—which concern how useful is to employ the theory for. While some of the virtues put forth by IBE-ists may be economic, I will not address this issue further.

of  $A_1$ , then certainly  $\Pr(T_1 \& A_1) < \Pr(T_2 \& A_2)$ . However, this is seldom the case. Normally,  $A_2$  and  $A_1$  are only partly over-lapping sets. Still, this notion of simplicity tracks something which often makes a difference in the comparative posteriors of two competing theories.

A form of *unification* is defensible as confirmationally relevant as well. As McGrew (2003: 562) notes, if we assume that for two competing hypotheses  $H_1$  and  $H_2$  that  $\Pr(H_1) = \Pr(H_2)$ , and that for some pieces of evidence  $E_1, E_2, \dots, E_n$ ,  $\Pr(E_n | H_1) = \Pr(E_n | H_2)$  for all  $n$ , then if  $H_1$  unifies the disparate pieces of evidence in the sense of making  $E_1, E_2, \dots, E_n$ , positively relevant to each other, i.e.  $\Pr(E_1 \& E_2 \& \dots \& E_n | H_1) > \Pr(E_1 | H_1) \times \Pr(E_2 | H_1) \times \dots \times \Pr(E_n | H_1)$  and  $H_2$  does not unify  $E_1, E_2, \dots, E_n$ , i.e.  $\Pr(E_1 \& E_2 \& \dots \& E_n | H_2) = \Pr(E_1 | H_2) \times \Pr(E_2 | H_2) \times \dots \times \Pr(E_n | H_2)$ , then it follows by Bayes' theorem that  $\Pr(H_1 | E_1 \& E_2 \& \dots \& E_n) > \Pr(H_2 | E_1 \& E_2 \& \dots \& E_n)$ . Thus, it seems that the probability calculus straightforwardly vindicates one particular notion of unification.<sup>14</sup>

Furthermore, *empirical adequacy* is sometimes cited as an explanatory virtue (e.g. Mackonis 2013), which if placed on one's list is clearly a confirmational virtue. If a hypothesis  $H$  entails some observational consequence  $O$ , then, provided  $0 < \Pr(H) < 1$  and  $0 < \Pr(O) < 1$ , it follows again from the probability calculus that  $\Pr(H | O) > \Pr(H)$ . Since Bayesianism vindicates hypothetico-deductive confirmation given a few caveats, to the extent that  $H$  is found to be empirically adequate, i.e. its observational consequences turn up true, then  $H$  will be confirmed. If we suppose, for two competing hypotheses  $H_1$  and  $H_2$ , that  $H_1$  entails  $O$ , and  $H_2$  does not, then this too will make a difference in the comparative posteriors of  $H_1$  and  $H_2$ , as then  $\Pr(O | H_1) > \Pr(O | H_2)$ .

Another virtue upheld by IBE-ists that is clearly confirmational is *fit with background data*. If  $H_1$  better fits with background data than  $H_2$ , then one way of making sense of this claim is that  $\Pr(H_1 | K) > \Pr(H_2 | K)$ , where  $K$  is one's set of background knowledge. Hitherto, for the sake of

---

<sup>14</sup> For a more extensive and technical treatment of this idea, see Myrvold (2003) who also agrees that "the ability of a theory to unify phenomena consists in its ability to render what, on prior grounds, appear to be independent phenomena informationally relevant to each other" (400). See Lange (2003) for a critique and Schupbach (2005) for a reply.

readability, I have suppressed the background K. But, of course, any probability statement should be relativized to some K. To the extent that  $H_1$  is better supported by the background data than  $H_2$ , the (relativized) prior probability of  $H_1$  will be higher than that of  $H_2$ .

Finally, one last virtue cited by IBE-ists (e.g. Okasha 2000) is *explanatory power*.<sup>15</sup> One way to make sense of explanatory power construed as a particular virtue is in terms of likelihoods (Sober 1990). Often, when we say that  $H_1$  better explains E than  $H_2$ , we mean that, although  $H_2$  may be compatible with E, still,  $H_1$  makes E more likely than  $H_2$ .<sup>16</sup> Consider the oft-cited passage by Darwin:

It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing; but it is a method used in judging of the common events of life, and has often been used by the greatest natural philosophers (1962: 476).

A plausible way to interpret what Darwin means when he says that the theory of natural selection explains several “large classes of facts” is that his theory has a higher likelihood than creationism:  $\Pr(\text{large classes of facts} \mid \text{Darwin's theory}) > \Pr(\text{large classes of facts} \mid \text{creationism})$ . If  $H_1$  has a higher likelihood than  $H_2$ , then this fact will make a difference in the comparative posteriors of  $H_1$  and  $H_2$ .

### 2.3 The (Non-confirmational) Informational Virtues

However, not all of the virtues cited by IBE-ists are confirmational virtues; rather, as I have suggested, some of them are informational virtues. Although informational virtues give a hypothesis greater informative content, these virtues are not properties that necessarily have a bearing on the probabilistic inequalities that are relevant to assessing the relationship between the posterior probabilities of  $H_1$  and  $H_2$ . For instance, consider again the list put forth by Lipton “mechanism, precision, scope, simplicity, fertility or fruitfulness, and fit with background belief” (2004: 122). At

---

<sup>15</sup> Of course, conceptually identifying explanatory power with likelihoods is problematic because  $\Pr(E \mid E) = 1 \geq \Pr(E)$ . Just because some hypothesis entails the phenomena, it does not follow that the hypothesis explains the phenomena.

<sup>16</sup> Sometimes “explanatory power” is used as a catch-all term to refer to how well a hypothesis does with respect to all the virtues (e.g. Thagard 1993; Psillos 1999; Mackonis 2013). But it is also common to use “explanatory power” to designate one virtue in particular (e.g. Okasha 2000; McGrew 2003; Glymour 2015).

least the first three are informational virtues and thus lack confirmational import.<sup>17</sup> Let's consider first the virtue of *mechanism*. According to Lipton's IBE, other things being equal, to the extent that  $H_1$  provides a causal-nomological mechanism to account for  $E$ , whereas  $H_2$  does not, then  $H_1$  does better with respect to the virtues than  $H_2$ . In this case then,  $H_1$  is a better explanation than  $H_2$ , and so, in accordance with IBE,  $H_1$  is more likely to be true. But if we understand "more likely to be true" to mean "has a higher posterior probability", i.e.  $\Pr(H_1 | E) > \Pr(H_2 | E)$ , then this claim is false.

Let  $H_1$  be the hypothesis that specifies the physio-chemical details of salt's dissolving in water, and let  $H_2$  be the hypothesis that salt dissolves in water because water has the solubility virtue. Now, it is not clear how to interpret  $H_2$ , but one way is to say that  $H_2$  posits that there is some property or set of properties had by water which disposes water to dissolve salt. And this is true. The solubility virtue is water's dipolar nature. For any given water molecule, its hydrogen atoms are partially positively charged and its oxygen atom is partially negatively charged. The positively charged hydrogen atoms of water molecules are attracted to the negatively charged chlorine ions and the negatively charged oxygen atoms are attracted to the positively charged sodium ions. The water molecules by way of this electromagnetic attraction surround the sodium and chlorine atoms that make up salt crystals and dissolve the table salt. Let  $P_1$  be this physio-chemical explanation, (or rather the complete explanation which includes those details I've left out).

Now, clearly  $H_1$  is a better explanation than  $H_2$  because  $H_1$  cites  $P_1$  in explaining the phenomena, but, contrary to what IBE-ists want to say,  $H_1$  is not more likely to be true because  $H_1$  entails  $H_2$ . It seems that  $H_2$  is really at bottom a statement of the form "There exists some causal mechanism by which water dissolves salt" and  $H_1$  is a statement of the form "There exists some causal mechanism by which water dissolves salt & that mechanism is  $P_1$ ". Clearly then,  $H_2$  cannot be less probable than  $H_1$  because  $H_2$  is entailed by  $H_1$ . In the future, however unlikely,  $P_1$  might turn

---

<sup>17</sup> Lipton (2004) seems aware of this (116), but not that it is in tension with his view of IBE as theory of confirmation (62), nor with his claim that his version of IBE can be reconciled with Bayesianism (103-120).

out false, and thus  $H_1$  will be falsified. But  $H_2$  will most likely survive, since it says less about the world than  $H_1$ . No matter what happens in the future of chemistry, water will probably continue to have the solubility virtue. To be sure though,  $H_1$  does better than  $H_2$  with respect to the virtue of mechanism, and so, if we take IBE seriously, then IBE should say that  $\Pr(H_1 | E) > \Pr(H_2 | E)$ . But this claim runs afoul of a theorem of probability theory according to which the logically stronger proposition necessarily has a probability less than or equal to the logically weaker hypothesis.

*Precision* suffers from the same problem as mechanism. The virtue of precision says that, other things being equal, if  $H_1$  gives more details than  $H_2$ , then  $H_1$  is a better explanation of the evidence  $E$  better than  $H_2$ . Thus, IBE-ists such as Lipton and Psillos who admit precision as a virtue in one form or another, should say that  $\Pr(H_1 | E) > \Pr(H_2 | E)$ . But let  $H_1$  be “Fields  $F_1$  and  $F_2$  differ in the average height of corn stalks by .5 meters” and let  $H_2$  be “There is some difference in the average height of corn stalks in  $F_1$  and  $F_2$ ”. Here,  $H_1$  is more precise than  $H_2$ , and so according to IBE, since  $H_1$  is a better explanation than  $H_2$ , other things being equal,  $H_1$  is more probable than  $H_2$ . This can’t be true, however, for the same reason above:  $H_1$  entails  $H_2$ , and so  $\Pr(H_1) \leq \Pr(H_2)$ .

With respect to *scope*, as Salmon (1990: 196-7) in his discussion of Kuhn’s criteria of theory evaluation points out, Newton’s theory has greater scope than the conjunction of Galileo’s law of falling bodies and Kepler’s three laws of planetary motion. This means that Newton’s theory explains phenomena beyond the conjunction of the laws of Galileo and Kepler.<sup>18</sup> Newton’s theory entails the conjunction of Galileo’s law and Kepler’s laws, but of course the IBE-ist will want to say that Newton’s theory is rationally preferable because of its greater scope. But if Newton’s theory entails the laws of Galileo and Kepler, then Newton’s theory cannot be more probable than the conjunction of Galileo’s and Kepler’s laws. Scope is only an informational virtue. That some hypothesis  $H_1$  has greater scope than some rival  $H_2$  does not necessarily make  $H_1$  more probable.

---

<sup>18</sup> This is true only if we ignore the small corrections to the laws of Galileo and Kepler that Newton’s theory makes.

It is because all the intended justificatory work of IBE is done by the virtues and because IBE-ists often include informational virtues on their lists of explanatory virtues that a compatibilist view according to which Bayesianism provides a direct, formal explication of IBE is bound to fail. Consider, for instance, the “emergent-compatibilism” of Henderson (2014), which attempts to reconcile Lipton’s version of IBE with Bayesianism. On Henderson’s view, IBE and Bayesianism are compatible because “a Bayesian who adopts constraints on her probabilities, which are reasonable on her own terms, would end up favouring more explanatory theories”, and so “IBE would ‘emerge’ without the Bayesian doing anything differently from what she would anyway”(2014:12-13). While Henderson is right that IBE and Bayesianism license *some* of the same inferences, if IBE-ists include informational virtues on their lists of virtues, those which have no confirmational import, then in principle IBE will *not* emerge as a natural consequence from the Bayesian framework.<sup>19</sup>

### **3. The Compatibilist Response: An Appeal to Objective Bayesianism**

To the argument I make in the previous section, the proponent of reconciling IBE and Bayesianism might object on two counts. First, one might object that when discussing the virtues, we should only compare incompatible hypotheses. I argued that mechanism, precision, and scope cannot be confirmational virtues, because for each virtue, even if some hypothesis  $H_1$  excels with respect to that virtue, we could imagine some competitor  $H_2$  that lacked that virtue, but which nonetheless could not be less probable than  $H_1$ , because in each case  $H_1$  entailed  $H_2$ , and so according to the probability calculus, it is necessarily true that  $\Pr(H_1) \leq \Pr(H_2)$ . But, one might argue, if we are assessing hypotheses for their the explanatory virtues, we should compare, say, an  $H_1$  which claims there is a mechanism to explain  $E$  and some competitor  $H_2$  which denies there is such a mechanism, instead of one which commits to some mechanism but does not spell out its details. Second, one might object that the argument illicitly assumes a subjective form of Bayesianism. Indeed, one of the

---

<sup>19</sup> One case that Henderson (2014: 703-711) focuses on is the preference for Copernicus’ theory over Ptolemy’s theory.

reasons that Psillos (2007: 447), a dyed-in-the-wool IBE-ist, argues that IBE is incompatible with Bayesianism is that the latter allows for too much subjectivity in the setting of prior probabilities. The IBE-ist wants to say that one ought to infer the best explanation of the evidence, not merely that doing so is rationally permissible. Even if one has a higher conditional credence in the lower quality explanation, one still ought to infer the best explanation of the evidence. So, already it seems that IBE is offering more normative constraints than Subjective Bayesianism endorses.

Perhaps it's impossible to derive all of the IBE-ist explanatory virtues, such as scope and mechanism, purely from the probability calculus. Even so one might claim that such virtues still ought to function as normative constraints on what counts as a rational Pr-function. With this in mind, one might think that locating IBE explicitly in the Objective Bayesian framework will increase the prospects of a successful marriage between IBE and Bayesianism. A number of philosophers have endorsed versions of this sort of compatibilism, a view which I will call the "Objective Bayesian approach". Recent advocates of the Objective Bayesian approach (Huemer 2009; Weisberg 2009; Poston 2014) have argued that the best way to integrate IBE into the Bayesian framework is to think of IBE as providing extra normative constraints on the objectively correct Pr-function, or the range of permissible ones.<sup>20</sup> If additional constraints on what counts as a rational Pr-function can be defended by appealing to explanatory virtues, then this would go some way in overcoming the subjectivity of Subjective Bayesianism. It would also be attractive to the IBE-ist, since it would give the explanatory virtues independent normative status within Bayesianism.<sup>21</sup>

---

<sup>20</sup> It should be noted that although Psillos (2007: 447) claims that IBE and Bayesianism are incompatible, he expresses sympathy toward a marriage between IBE and Objective Bayesianism. Lipton (2007: 458) too expresses sympathy with the Objective Bayesian approach, holding out hope that there is a "normatively privileged way of assigning priors and likelihoods" that will vindicate the explanatory virtues of IBE.

<sup>21</sup> This is in contrast to other versions of compatibilism, endorsed by Okasha (2000), McGrew (2003), and Lipton (2004) which attempt to reconcile IBE from within the subjective Bayesian framework. This brand of compatibilism argues that explanatory factors can be used as a kind of heuristic approximation to Bayesian inference, at which we are by our nature psychologically deficient. In my view, the position that IBE simply provides a kind of heuristic aid, which we might use instead of performing a Bayesian inference directly, robs IBE of much of its philosophical interest. See Douven (2005), Weisberg (2009), and Henderson (2014) for extensive criticisms of this version of compatibilism.



Weisberg’s version of Bayesianism endorses what he calls “Objectivist Conditionalization”, which states that “[a]t any given time, your credence in an arbitrary proposition H ought to be  $p(H|E)$ , where  $p$  is the correct a priori probability distribution, and E is your total evidence at that time”(2009: 137).<sup>22</sup> Unlike most forms of Subjective Bayesianism, Weisberg’s Objective Bayesianism will count as irrational any Pr-function whose initial distribution of priors or whose likelihoods are not in accord with what  $p$  says. According to Weisberg and other proponents of the Objective Bayesian approach, we can appeal to IBE and explanatory considerations in general in order to help spell out the nature of  $p$ . In this way, IBE and Bayesianism might have a productive partnership.

Before evaluating the prospects of defending the virtues that don’t seem confirmationally relevant by the Bayesian’s lights as extra normative constraints, it would be helpful to look at some of Weisberg’s concrete proposals. Weisberg suggests three ways IBE could aid in the project of specifying the nature of  $p$ . Owing to considerations of space, I will only discuss two.<sup>23</sup>

First, the Objective Bayesian can incorporate into  $p$  constraints inspired by a response to the grue problem by White (2005), who argues that we ought to believe the all-green generalization over the all-grue generalization because the all-green generalization has the property of *stability*, where “[a]n explanation of a fact F is stable to the extent that according to this explanation, F couldn’t easily have failed to obtain”(2005: 4). Because “green” leads to a more stable hypothesis than “grue”,  $p$  should satisfy the following inequality:  $p(\text{All Green} | \text{All Observed Green}) > p(\text{All Grue} | \text{All Observed Green})$ . Since the all-green and the all-grue hypothesis both have the same likelihood with respect to the evidence, the difference in their posterior probabilities must lie in the prior probabilities, such that  $p(\text{All Green}) > p(\text{All Grue})$ . In general, according to this proposal, to the extent that a hypothesis H is more stable than a rival H\* given the evidence E,  $p(H|E) > p(H^*|E)$ .

---

<sup>22</sup> It should be noted that Williamson (2010) defends a version of Objective Bayesianism, but rejects Bayesian conditionalization, relying instead on the principle of maximum entropy as an updating rule.

<sup>23</sup> Weisberg’s third proposal is that explanatory factors can be used to patch up the Principle of Indifference (2009: 140), a canonical Objectivist principle that has fallen into disrepute (Sober 2002). See also Huemer (2009) for a similar view.

A second suggestion of Weisberg's takes its inspiration from the Best-Systems account of laws of nature defended by Lewis (1973). If worlds that are more simple and orderly will be better able to explain the facts—which is a central claim of the Best-Systems account—then Weisberg suggests that the Objective Bayesian can regard worlds that do not admit of a good deductive systemization as “explanatorily recalcitrant, and thus a priori improbable” (2009: 139). To avoid confusion with the notion of simplicity canvassed earlier, I will call simplicity in Weisberg's sense “orderliness”. When comparing two hypotheses  $H$  and  $H^*$ , if  $H$  describes a more ordered world than  $H^*$ , or if  $H$ , given our total evidence  $E$ , allows for a deductive system in which we can better situate—and thereby explain—particular, local matters of fact than  $H^*$ , then  $p(H|E) > p(H^*|E)$ .<sup>24</sup>

### 3.1 Criticism of the Objective Bayesian Response

To be sure, Weisberg's proposals are meant to be only suggestive, but it is not clear, even if further details were provided, that additional constraints of this sort could be defended in any satisfactory way. While the idea of eliminating the subjectivity from Subjective Bayesianism by appealing to further normative constraints inspired by explanatory considerations is tempting, as I will argue here, the Objective Bayesian approach to compatibilism fails. As a result, appealing to Objective Bayesianism as a response to the problems I raise in section 2—as way a way to defend the confirmational import of what I've deemed merely informational virtues—proves unsatisfactory.

For Weisberg, any hypothesis  $H$  that is more stable or more orderly than an unstable or disorderly competitor  $H^*$  has a higher prior probability. But what is it about stability, or orderliness that makes  $H$  more probable when  $H$  has these features? The invocation of Objective Bayesianism

---

<sup>24</sup> This second suggestion of Weisberg's is similar to Jeffreys' “Simplicity Postulate” (1931), according to which given a set of competing hypotheses, the simpler hypotheses ought to be assigned a higher prior probability. One way Jeffreys applies this idea is to interpret simplicity as the degree of a polynomial equation relating two variables to each other (e.g. temperature and pressure) and to distribute priors in the following way: the simplest hypothesis  $H_1$  of some partition is to be assigned a prior of  $1/2$ , the next simplest  $H_2$  a prior of  $1/4$ ,  $H_3$  a prior of  $1/8$ , and so on, where  $H_n$  receives a prior of  $(1/2)^n$ . Thus, for instance, the model that posits a linear relationship receives a prior probability of  $1/2$ , that which posits a quadratic relationship a prior probability of  $1/4$ , that which posits a cubic relationship a prior probability of  $1/8$ , etc. See Sober (2015: 87-93) for a critical discussion of the Simplicity Postulate.

suggests that a defense of the epistemic relevance of these constraints will be a priori. Indeed, Poston (2014: 83-4)—who endorses the Objective Bayesian approach as part of his extended defense of explanatory coherentism of a broadly Quinean sort—takes simplicity to be intrinsically epistemically relevant and unanalyzable.<sup>25</sup> Our justification for simplicity, if there is one at all, depends on reflectively considering our particular epistemic practices. This view is similar to that of Sider (2013: 3), who also considers simplicity—in the sense of positing few fundamental concepts or predicates—to be intrinsically epistemically relevant, our justification for which, Sider claims, is the a priori intuition that simplicity tracks truth. Of course, the problem with this sort of defense of orderliness, or stability, or any other explanatory virtue, is that it isn't much of a defense at all.

Rather than quixotically trying to justify a priori the truth-conduciveness of stability, orderliness, or any of the explanatory virtues, a more promising avenue is to appeal to the past success of theories that have had these properties. If it turns out that a large proportion of successful theories have had the explanatory virtues, then we might be justified a posteriori in employing them in our inferences about future phenomena. In this way, while a Pr-function that favors orderly hypotheses, or theories that are stable, or hypotheses that posit a mechanism would be rationally required in our world, such constraints would not apply to any conceivable Pr-function.

The idea that IBE and the explanatory virtues might be defended on empirical grounds has already been taken up by Psillos (1999) in his defense of scientific realism. According to Psillos (1999: 77), what matters for justifying a rule of inference is that using that rule on balance produces more truths than falsehoods.<sup>26</sup> To be justified in using some rule—in this case IBE—as an engine of inference, it is not necessary that we understand how that engine works, as it were. So, if appealing to the explanatory virtues upheld by IBE-ists has been historically successful in producing true

---

<sup>25</sup> While Poston may be right that there is no exhaustive analysis of simplicity that covers every plausible case, certain instances of simplicity *can* be given a deeper justification. See, for example, Forster and Sober (1994).

<sup>26</sup> Psillos (1999: 81) endorses externalism in general and reliabilism in particular, (e.g. Goldman 1986). On this set of views, it is not necessary to have some demonstration that an inference rule is reliable, as the internalist would demand.

beliefs, then—supposing that the reliability of the inference rule is all that’s needed for its justification—we would be justified in our belief that the explanatory virtues track truth. In this way, we would be justified in employing the virtues, including those whose confirmational import I doubted in section 2, even if they cannot be directly vindicated by the probability calculus.<sup>27</sup>

However, the problem remains that the historical track-record of the explanatory virtues will most likely not survive detailed empirical scrutiny.<sup>28</sup> If any theory ever did well in terms of its explanatory virtues, this theory was Newtonian mechanics. Indeed, the 19<sup>th</sup> century historian and philosopher of science William Whewell exalted some of the properties of Newton’s theory—such as its successful explanation of Kepler’s three laws—as good-making features of scientific theories in general (1968: 153). When a theory explained different classes of facts which before had seemed unconnected, Whewell called this a “consilience of inductions.” For Whewell, we should think consilience is a virtue because the historical record “offers no example in which a theory supported by such consiliences, had been afterwards proved to be false” (1968: 295).

Unfortunately for Whewell, the 20<sup>th</sup> century saw the overthrow of Newton’s theory of gravity in the form of Einstein’s. But even for those who would defend the more modest claim that the historical record more or less, though not perfectly, vindicates the explanatory virtues, the prospects look grim. As Sober (2015: 46-7) notes in a discussion of Whewell and this sort of a posteriori argument for a general criterion of simplicity, the track-record defense seems intuitively

---

<sup>27</sup> Interestingly, Salmon (1990), despite being a critic of IBE, offers this track-record defense of virtues such as simplicity and unification. According to Salmon (1990: 187), we can give the prior probabilities that feature in Bayes’ theorem an objective interpretation by understanding them “as our best estimates of the frequencies with which certain kinds of hypotheses succeed”.

<sup>28</sup> One might object, as Lycan (2002) does, that the track-record defense is *in principle* impossible because to argue, for instance, that simplicity has a successful track-record assumes that we can verify which hypotheses are true independently of our appeals to simplicity. But this is false. It is not as though we have access to all the true hypotheses and from there we can check to see that a high proportion of them have the virtue of simplicity. On the contrary, we believe that our hypotheses are true on the basis of simplicity. To respond to this worry, one might argue that, while we can’t inspect all the true theories from the view from nowhere and then see what proportion are simple, we can in principle investigate those theories which have been predictively successful and resisted disconfirmation for a long time and see what proportion of those theories have been simple, unifying, etc. If we have reason to believe that mature, predictively successful theories are true, and those theories are also simple, unifying, precise etc. then we would in principle be able to defend the track-record argument for the truth-conduciveness of explanatoriness.

plausible only because of our selective attention. We tend to remember those grand, elegant theories such as Einstein's theory that turned out to be (more or less) successful, and discount the probably thousands of theories that did well by the IBE-ist explanatory virtues, but which were not so successful. But for this a posteriori defense to be legitimate we must look at both the successful and the unsuccessful explanatory theories and compare them. No adequate investigation of this sort has been undertaken to my knowledge, and even if it is in principle possible, it seems unfeasible.

With regards to Weisberg's concrete proposals, if any defense can be given of why we should regard an instance of stability or orderliness as truth-conducive, it would seem to depend on substantive background beliefs. For instance, White writes: "On the all-grue hypothesis we have two properties, greenness and having been observed, which are co-instantiated by the same subclass of emeralds. This is a striking fact which seems to call for an explanation" (2005: 18-19). This remark suggests that the problem with the grue hypothesis stems from its positing a correlation that is surprising given our background knowledge. We should expect some causal mechanism linking these two properties, supposing there is some connection. The fact that the all-grue hypothesis does not propose one makes it implausible. Similarly, in the case of Weisberg's notion of orderliness, suppose we let  $H =$  "the gravitational force does not vary tremendously from location to location" and we let  $H^* =$  "the gravitational force varies tremendously from location to location." Intuitively,  $H$  is more orderly than  $H^*$ . Since  $H$  has the virtue of orderliness, we should say that  $p(H) > p(H^*)$ . Of course, if we have a pre-existing reason to believe that forces in general remain constant in their strength and don't vary will-nilly, then it would be right to think that  $\text{Pr}(H) > \text{Pr}(H^*)$ . However, it's not clear that this move can be defended as a general constraint on  $p$ , absent some pre-existing reason to believe that the world is orderly, such as Newton's belief that God made the world simple.

But if the defense of these proposed explanatory virtues depends on substantive background beliefs, then, for the IBE-ist who wishes to locate IBE in the Objective Bayesian camp, such virtues

will not be the kind of criteria to impose as general normative constraints on a rational Pr-function. A Pr-function might justifiably favor theories without scope, or ones that don't posit a mechanism, because of different justified background beliefs. Suppose, for instance, one already knows that positing a mechanism would be implausible. Suppose, for instance, that E is some spurious correlation, such as the positive correlation between the rise in British bread prices and Venice sea-levels over the past few centuries (Sober 2001). If  $H_1$  posits a mechanism to explain E—say, an international conspiracy involving the Illuminati—then  $H_1$  may explain E, in the sense of providing a common cause to screen off the correlation. But  $H_1$  would be quite implausible. Given our background knowledge, clearly  $H_2$ , which denies there is a mechanism, is preferable to  $H_1$ .

The central problem that I have raised here is something that has not escaped the notice of Bayesians. Consider a discussion of Myrvold (2015) concerning “Common Origin Inferences” (Jansen 2002), those inferences which posit some common cause or other type of explanation to account for the phenomena. As Myrvold argues, it is difficult to justify with any generality a preference for this type of unification within the Bayesian framework. Rather, it is only unification in the sense canvassed above, which Myrvold (2003) and McGrew (2003) defend, i.e. making pieces of evidence mutually positively relevant, that increases evidential support. To be sure, sometimes it is preferable to posit a unifying causal mechanism in light of the data, because otherwise one would be admitting an implausible coincidence. However, the sort of methodological rule that underlies this judgment is as Myrvold rightly notes, nothing deep (2015: 13). It is simply the requirement that we have a low prior credence in hypotheses that are improbable. But, as I suggested above, whether a coincidence is in fact improbable, and so whether positing a mechanism is indeed a virtue, will depend on our background knowledge and the nature of the phenomena being investigated.

In sum then, the Objective Bayesian response does not succeed in vindicating those virtues that I've argued are non-confirmational. In general, neither an a priori nor an a posteriori defense of

such virtues is plausible. This is likewise true for the further virtues proposed by Weisberg. So, even if we consider incompatible hypotheses instead of ones that are compatible, it is not clear how the virtues of mechanism, precision, or scope could be vindicated by the Bayesian's lights. Sometimes what I have been calling informational virtues will have a bearing on probability assignments, but when they do, that will turn on one's background knowledge and the nature of the problem at hand. Adopting these virtues as general constraints on a privileged Pr-function is thus unjustified.

#### **4. Reconceiving IBE as a Theory of Scientific "Acceptance"**

Despite the difficulty of vindicating the explanatory virtues upheld by IBE-ists from within the Bayesian framework—within both subjective and objective varieties—in this section, I will argue that this difficulty need not prevent the IBE-ist and the Bayesian from being friends. One way to avoid the problem that I have highlighted is to remove the informational virtues from the lists put forth by IBE-ists. But if we do that, and keep, say, only those items like empirical adequacy or fit with background data, then we risk conceiving of IBE as a kind of un-precisified Bayesianism. This would make IBE and Bayesianism easily compatible but only at the cost of trivializing IBE.<sup>29</sup> Indeed, proponents of IBE warn that we should not think of IBE as "inference to the likeliest explanation", that is, inference to the explanation with the highest posterior probability (Lipton 2004: 60; Psillos 2002: 617). Rather than rid IBE of its informational virtues to make it a more suitable, though uninteresting, theory of scientific inference, I propose that we keep those items on the list and reconceive the nature of IBE. Even if a theory is not made more likely than a rival to be true simply by having greater mechanism, scope, precision, etc., these are still properties that we want scientific theories to have. Informational virtues are still virtues; they are still good-making features of a scientific theory. This is an important point, which has been stressed especially by IBE-ists. If  $H_1$

---

<sup>29</sup> Some philosophers accept this deflationary view of IBE. For example, Ben-Menahem (1990: 324) equates explanatory merit with high probability, and happily admits that this trivializes IBE. See also, Cartwright (1983: 6), who remarks: "no inference to best explanation; only inference to most likely cause."

exhibits the informational virtues to a much greater degree than  $H_2$ , then other things being equal—where, suppose  $\Pr(H_1) = \Pr(H_2)$ —there is some sense in which  $H_1$  is *better* than  $H_2$ .

As I will argue, IBE should be thought of as an informal theory of *scientific acceptance*, where the degree of acceptability of  $H$  includes but is not exhausted by  $\Pr(H)$ .<sup>30</sup> In broad outline, the acceptability of  $H$  concerns whether  $H$  is worthy of commitment as a research program, which in turn depends on whether  $H$  meets the goals of science. I will advance this view for a number of reasons: 1) we should want some concept of acceptance to make sense of scientific practice; 2) IBE constitutes a plausible theory of what makes a hypothesis acceptable in one particular sense; and finally, 3) appeal to acceptance makes good sense of the relationship between IBE and Bayesianism.

#### **4.1 Acceptance, Non-epistemic Values, and the Goals of Science**

It's clear that scientists can take any number of attitudes toward a hypothesis  $H$ , of which belief is only one of them. Of course, some attitudes are more philosophically interesting than others. For instance, we shouldn't be too much concerned if a scientist loves or hates  $H$ . Instead, we should be interested in what have been called "cognitive attitudes", such as believing  $H$ , having a credence in  $H$  greater than .5, entertaining  $H$  for the purposes of pointing out its weakness, and accepting  $H$ , all of which consist in "an evaluative response to some content" (Elliott and Willmes 2013: 810-11). Crucial to understanding the proper relationship between IBE and Bayesianism is getting clear about the different sorts of cognitive attitudes scientists might take to a hypothesis or model, in particular this attitude of "acceptance". What does it mean to accept  $H_1$  but not to believe  $H_1$ ?

In ordinary English "to accept" is often used synonymously with "to believe". So, if "to believe  $H_1$ " just means "to take  $H_1$  to be true", and that is all that is meant by "to accept  $H_1$ ", then Bayesianism will have much to say about the matter. At the very least, if it is rational to believe  $H_1$ ,

---

<sup>30</sup> The views of Dawes (2013) and Nyrup (2014), both of whom sketch a deflated view of IBE, are similar to the one that I advocate below, although differing in the details. My approach, however, is unique in that I propose reconceiving the nature of IBE as a solution to the problem of sorting out IBE's relationship to Bayesianism.



then, absent some deviant account of the nature of belief, it ought to be the case that  $\Pr(H_1) > .5$ .<sup>31</sup> If this condition did not hold, then one could rationally believe that P even though one is more confident that  $\sim P$ . However, for one to utter “I believe that it will rain, but I am 60% confident that it won’t rain” seems irrational, if not incoherent. If anything, one ought to believe that it will not rain.<sup>32</sup> With this in mind, the success of the thesis that the IBE-ist informational virtues are relevant to what we ought to accept, but not to what we ought to believe, depends on a clear distinction between “belief” and “acceptance”, despite these terms often being used interchangeably.

The proposal that I make here concerning the informational virtues and acceptance of a hypothesis parallels a view that some philosophers take regarding the role of other non-epistemic values, e.g. political, moral, social, etc. in science.<sup>33</sup> Traditionally, the view regarding the relationship between science and (non-epistemic) values is that scientific reasoning ought to be “value-free” (e.g. Levi 1960, McMullin 1983; Lacey 1999). The value-free ideal is often expressed by the claim that scientists ought to be “objective”; they ought not to let their ethical, political, or personal values influence their judgments about the truth of a theory. Certainly, non-epistemic values play a role in the selection of research questions, the allocation of funding, or the treatment of human test subjects. But these roles are external to the practice of science itself. Non-epistemic values ought to play no internal role in science, i.e. no role in a scientist’s judgments about a theory *qua* scientist.

Recently, however, in one way or another, it has become increasingly common to challenge the traditional value-free ideal (e.g. Brown 2013; Douglas 2000, 2009; Elliott and McKaughan 2014; Elliott and Willmes 2013; Kourany 2003, etc.). To do so, it is not necessary to deny a distinction

---

<sup>31</sup> On Kaplan’s view, “You count as believing P just if, were your sole aim to assert the truth (as it pertains to P), and your only options were to assert that P, assert that  $\sim P$ , or make neither assertion, you would prefer to assert that P” (1996: 109). On this view, as Kaplan notes (1996: 111), it is possible to have very little confidence in P and still believe P.

<sup>32</sup> One might wish to assert the stronger principle that it is rational to believe that P *if and only if*  $\Pr(P) > .5$ ; however, while this principle seems *prima facie* plausible, the lottery paradox (Kyburg 1961) shows that imposing the threshold  $\Pr(P) > .5$  (or any threshold aside from 1) as necessary and sufficient for its being rational to believe that P, assuming some plausible rational constraints on binary beliefs, leads to objectionable results.

<sup>33</sup> I will use “virtue” synonymously with “value” and “non-epistemic” synonymously with “non-confirmational” below.

between epistemic and non-epistemic values (e.g. Rooney 1992, Longino 1996). Even if we reject the idea that a theory is more probable on the basis of its moral or political virtues, non-epistemic values might still play an indirect role in setting the standard of evidence for accepting or rejecting a null hypothesis (Rudner 1953; Douglas 2000), for determining what hypothesis or model we ought to accept for some regulatory policy, or for managing risk (Elliott and McKaughan 2014), etc.

Fortunately, philosophers who oppose the value-free ideal in this way have already sought to clarify this distinction between belief and acceptance (e.g. Elliott 2011; Elliott and Willmes 2013).

One useful account is that Elliott and Willmes (E&W) who define “acceptance” as follows:

A subject S accepts a hypothesis (or model, theory, etc.) H if and only if S presupposes H for specific reasons in her deliberation (2013: 811).<sup>34</sup>

On E&W’s account, acceptance is left deliberately vague so that there will be different, particularized notions of acceptance, such as acceptance for reasons of simplifying a task, acceptance for reasons of drafting regulatory policy, or acceptance for reasons of furthering future investigation. In this way, what one accepts will depend on contextual factors such as one’s chosen goals. From this, it follows that acceptance differs from belief, in that belief, plausibly, is not directly a matter of will.

So, just as philosophers who oppose the value-free ideal argue that non-epistemic values can be relevant with regards to accepting a hypothesis as a basis for action, I argue that the informational virtues upheld by IBE-ists, which are difficult to defend as confirmationally relevant, ought to be regarded in a similar fashion. The particularized notion of acceptance that I argue we ought to associate with IBE most closely matches the third variety mentioned by E&W, namely acceptance in the sense of making a commitment to a hypothesis or theory as a research program. In *The Scientific Image* (1980), van Fraassen, nicely articulates this notion of acceptance:

[A]cceptance involves not only belief but a certain commitment. Even for those of us who are not working scientists, the acceptance involves a commitment to confront any future

---

<sup>34</sup> As E&W note, this account is similar to that of Cohen (1992) and Bratman (1999), who also speak of acceptance as a matter of S’s presupposing some H in background deliberation, without necessarily believing H.

phenomena by means of the conceptual resources of this theory. It determines the terms in which we shall seek explanations. If the acceptance is at all strong, it is exhibited in the person's assumption of the role of explainer, in his willingness to answer questions *ex cathedra*. Even if you do not accept a theory, you can engage in discourse in a context in which language use is guided by that theory—but acceptance produces such contexts (12).

For van Fraassen, to accept a theory T goes beyond believing that T is empirically adequate. Rather, to accept T involves some commitment to T. And to be committed to T goes beyond simply entertaining the possibility of T's truth, but rather entails dispositions to pursue T as a research program. Certainly, if T is worth committing to, then T should have a sufficient degree of plausibility. It would be foolish, for instance, to commit to T if one knows that T is false or very improbable.<sup>35</sup> But it is certainly possible to commit to confront future phenomena in terms of the resources of T, to employ T in one's future inquiry into, deliberations about, and explanations of phenomena, without believing T is true.<sup>36</sup> The latter is all that I will mean by "acceptance" of a theory.<sup>37</sup> In fact, accepting a theory T in the sense of making a commitment to T in the manner described above—without believing T is true—seems to be something that scientists do all the time.

This fact is starkly evident in cases in which two well-confirmed theories from different domains turn out to be jointly inconsistent. Consider the controversy between biologists and physicists over the age of the Earth in the late 19<sup>th</sup> century. According to Lord Kelvin, one of the

---

<sup>35</sup> In accordance with E&W's broad account of acceptance, one might accept for predictive purposes a theory that one knows to be false, one which, for example, contains obviously false idealizations. But if the one the goals of science is truth, then a theory without such idealizations is ultimately desirable.

<sup>36</sup> Of course, van Fraassen thinks that we should never believe T is true if T makes claims about unobservables.

<sup>37</sup> This increasingly popular terminological choice of using "acceptance" in a way that is severed from belief, which I admittedly perpetuate, is in some ways unfortunate. For instance, Laudan (1977: 108-114) articulates a distinction similar to the one that I am urging here in the context of IBE, distinguishing between the "context of acceptance" and the "context of pursuit", where by "acceptability", he means, as in ordinary English, "worthy of belief" (110). My use of "acceptance" most closely matches Laudan's use of "pursuit" and its cognates. So, where Laudan would say, "scientists can have good reasons for *working on* theories they would not *accept*" (110), I would say, "scientists can have good reasons for *accepting* theories they would not *believe*." Similarly, where Laudan would say, "we cannot be accused of inconsistency or irrationality if we pursue (without accepting) some highly progressive research tradition" (112), I would say, "we cannot be accused of inconsistency or irrationality if we *accept* (without *believing*) some theory with a high degree of explanatory content." As I discuss below, I take Laudan's concerns about fecundity seriously by including fecundity as one of the IBE-ist informational virtues. Unlike Laudan, however, I urge that for a theory to be acceptable (or worthy of pursuit for Laudan) it must have some degree of plausibility. Still, on my view, we are justified in committing to (or pursuing for Laudan) theories that we would be irrational to believe, a point that Laudan takes pains to emphasize.

most respected physicists of his day, the second law of thermodynamics was inconsistent with the uniformitarian geology assumed by evolutionary biology. One of Kelvin's arguments was that the Earth as we know it can't be very old because the sun is "assumed to be an incandescent liquid now losing heat" (1862: 159). For the sun to be at its present energy state, it would have had to have been so hot only 50 million years ago as to make the Earth molten hot, and thus unable to support life. So, the Earth did not seem to have the geological history that Darwin needed for the theory of evolution to be true. The discovery that the sun is powered by nuclear fusion by Hans Bethe in 1939—for which he won the Nobel Prize—and that, moreover, this process emits an enormous amount of energy, allowed for older estimates of the age of the Earth and removed the tension that had existed between physics and biology. At the time, though, given that both thermodynamics and evolutionary biology were well-confirmed theories, certainly some scientists still accepted, i.e. were still committed to both Darwin's theory and thermodynamics as research programs—although Lord Kelvin was not among them. However, it would be manifestly uncharitable to say that scientists believed both theories, in the binary sense, given that the two theories were jointly inconsistent.<sup>38</sup>

But what determines in general whether a theory is acceptable in the relevant sense? Naturally, it would seem that the answer depends on the goals of science. A theory is worth committing to when it meets the goals of science. Put comparatively, a theory  $T_1$  is more acceptable than a rival  $T_2$ , when  $T_1$  meets the goals of science better than  $T_2$ . If the goals of science were simply to identify and have us believe theories that are highly probable, then the conditions under which  $T$  is acceptable, that is worth committing to and worth pursuing, would be no different than the conditions under which we are justified in believing  $T$ . I've mentioned that if  $T$  is acceptable, then  $T$

---

<sup>38</sup> At present one of the most well-known pairs of well-confirmed, though inconsistent theories, is the theory of general relativity and quantum mechanics. See the contributions in the edited volume of Callender and Huggett (2001) for contemporary approaches to reconciling general relativity and quantum mechanics.

should be sufficiently probable, and this makes sense given that truth is a goal of science. However, it is not plausible that the goals of science are simply to discover theories that are highly probable.

The idea that science cares about more than probable truth has been influentially voiced by Popper (1980: 416): “Science does not aim, primarily, at high probabilities. It aims at a high informative content, well backed by experience. But a hypothesis may be very probable simply because it tells us nothing, or very little.” For instance, a Moliere-esque hypothesis which attributes an unspecified dispositional property to account for the data will have a probability of unity or near unity. Of course, such hypotheses are not at all worthy of being taken seriously, not worthy of commitment, which shows that high probability is not the only goal of science.<sup>39</sup>

#### **4.2 Sketches of an IBE-ist Theory of Acceptance**

The foregoing discussion suggests that what makes for an acceptable scientific hypothesis then just is some appropriate balance of probability and informative content. But this approach will not be entirely satisfactory. It is true that scientists do prefer theories with high informative content, but it is not just any kind of content, since information comes cheap. One can trivially further the goal of informative content by artificially logically strengthening one’s hypothesis, e.g. by tacking on an irrelevant conjunct. In principle, tacking on an irrelevant conjunct could bring about a greater balance of probability and content. But this is counter-intuitive. Since it seems implausible that such an artificial maneuver could in principle make a hypothesis more scientifically respectable, it is not sufficient to say that an acceptable hypothesis consists in some appropriate balance of probability and informative content. Rather, the information has to have the right character.

Here is where IBE can help. Instead of informative content, we might say that another goal of science besides truth is to discover theories with a high degree of *explanatory content*, where a theory’s explanatory content is a function of how well it does with respect to the list of

---

<sup>39</sup> It should be noted, valuing informative content is not an aesthetic idiosyncrasy on the part of scientists; rather, we assume, and with good reason backed by experience, that whatever the truth is, that truth is a substantive truth.

informational virtues put forth by IBE-ists. Unlike Popper's conception of informative content which seems to be just a function of the logical strength of H, the explanatory content of H is a two-place relation, depending in part on the intrinsic nature of H, but also on how H is related to E. This is because some of the informational virtues put forth by IBE-ists depend on how well H serves as an explanation of E. For example, to know if  $H_1$  provides a more precise mechanism than  $H_2$ , and thus has a higher degree of explanatory content, we need to know what E is which is to be causally explained. For this reason, the acceptability of H must be relativized to E. A hypothesis might be acceptable in some contexts but not in others. For other virtues, such as scope, whether  $H_1$  has a higher degree of scope than  $H_2$  can be evaluated independently of what our evidence happens to be.

With this in mind, IBE construed as a theory of acceptance says that the extent to which  $T_1$  is an acceptable theory—in the sense of being worthy of commitment in the way described in above—depends on the extent to which  $T_1$  is probable and explanatorily contentful.  $T_1$  is acceptable just in case  $T_1$  has a sufficiently high degree of probable truth and explanatory content, and  $T_1$  is more acceptable than some rival  $T_2$  just in case  $T_1$  has a greater combination of probable truth and explanatory content than  $T_2$ . While  $\text{Pr}(T_1)$  is a component of the acceptability of  $T_1$ —not as an arbitrary stipulation, but because truth is a goal of science—IBE regarded as a theory of one particular notion of scientific acceptance does not say that if  $T_1$  does well with regards to the informational virtues, then that fact makes  $T_1$  more probable. Rendered in this way, IBE would be compatible with Bayesianism because the latter is silent on which theories are worth committing to for the purposes of deliberation, explanation, and further inquiry into phenomena.<sup>40</sup>

---

<sup>40</sup> This approach is similar to the epistemic utility theory of Isaac Levi (1967), according to which theory choice is a pragmatic, decision-theoretic problem mediated by the two distinct and competing goals of truth and informative content. Others who advocate a similar view, whereby truth and content combine to fix what it is that we ought to believe, include Hintikka and Pietarinen (1966), Maher (1993), and Huber (2007). One could locate IBE in this framework as well, but this involves radically re-thinking the connection between full belief and degrees of belief, as full belief would no longer be solely determined by degrees of belief, but would be partly determined by pragmatic factors.

On my proposal then, IBE-ists should not think of the compatibilist project as that of spelling out the constraints put on the a priori correct probability function  $p$ , as Weisberg (2009) claims; rather, if I am right that IBE in its most well-articulated forms cannot be defended as a theory of scientific inference, then IBE-ists should instead see their project as that of spelling out the nature of *explanatory content*. In addition to the informational virtues already discussed, we might add *fertility*, or the extent to which a theory opens up new avenues of research or points to new areas where problem-solving will be needed. Psillos (2002: 615) cites the virtue of *importance*, according to which, if  $H_1$  and  $H_2$  both fail to explain all the relevant data, but  $H_1$ , unlike  $H_2$  explains “the most salient phenomena”, then  $H_1$  is preferable to  $H_2$ . Like all the informational virtues, it is not clear why a hypothesis that explains the most salient phenomena—however “salience” is understood—should be more likely to be true. But explaining the most salient phenomena may be relevant to the explanatory content of the theory, and if a goal of science is to offer explanations by finding theories with a high degree of explanatory content, then insofar as a theory exhibits the virtue of *importance*, then that fact is relevant to whether we ought to commit to that theory.

Spelling out the nature of *explanatory content*, will of course be no easy task. For a fully satisfactory IBE-ist theory of acceptability, much more would need to be said by way of identifying and analyzing the particular informational virtues than can be said here. Once a clearer picture of the nature of scope, precision, mechanism, and the other explanatory virtues is provided, we will have a clearer picture of IBE construed here as a theory of acceptance. Luckily, some virtues such as mechanism are already the subject of extensive analysis (e.g. Machamer, Darden, and Craver 2000), and IBE-ists have provided some—albeit sketchy—descriptions of the other virtues.

Even without the details, we can see that the notion of acceptance that I have spelled out here makes good sense of how we ought to regard certain classes of hypotheses. First, it follows that vacuous explanations are not acceptable. Let  $H$  = “either it is raining or it is not raining” and let

$E$ ="salt dissolves in water". Clearly,  $\Pr(H) = 1$ . However,  $H$  does not explain  $E$  at all, and so the explanatory content of  $H$  with respect to  $E$  is little to none. In general, tautologies, such as  $H$ , should never be a candidate for scientific acceptance. Certainly, logical theorems and definitions can be part of a contentful theory worth pursuing. But one should not accept  $H$  in this context. As van Fraassen (1980: 4) nicely puts it, acceptance involves "continuing the dialogue with nature in the framework of one conceptual scheme rather than another", and tautologies such as "either it is raining or it is not raining" just seem like the wrong sort of entity to fill this role.

Second, given that the notion of acceptance I articulate here takes into account probable truth as well as explanatory content, no contradiction will be acceptable as worthy of pursuit, since obviously for any contradiction  $\perp$ ,  $\Pr(\perp)=0$ . On top of that, while contradictions may be maximally informative in the sense of entailing every proposition, contradictions are not capable of having explanatory content in the relevant sense. It's not clear how a contradiction could explain anything. This is another reason to prefer a notion of explanatory content over mere, informative content when determining which hypotheses are candidates for commitment as a research program.

Third, the notion of acceptance I defend here makes sense of how we ought to regard catch-all hypotheses. Let  $H$  = General Theory of Relativity (GTR) and  $\sim H$  = Not-General Theory of Relativity (Not-GTR). In this case,  $\sim H$  is equivalent to a large disjunction of every other alternative, incompatible theory of gravity, some of which haven't yet been thought of. Since the catch-all is equivalent to a disjunction of every other hypothesis in a partition—whose probabilities by Finite Additivity we sum—often the catch-all will have a greater probability than any specific hypothesis, at least at the beginning of inquiry. But, even though in our case  $\Pr(\sim H)$  may be high, the explanatory content of  $\sim H$  will not be high. It is not clear how a disjunction of disparate theories of gravity can be very explanatory, and so it is not clear how  $\sim H$  in our case can excel with respect to the IBE-ist informational virtues. While in some cases the catch-all may be an acceptable hypothesis with



explanatory content—e.g., when the hypothesis space can be enumerated—in cases of general theories such as GTR, the catch-all, despite its high probability, is not a hypothesis that scientists could commit to as a research program and in terms of which continue the dialogue with nature.

Finally, an appeal to informational virtues could play a role as a kind of “tie-breaker” in cases of under-determination. If, for two competing hypotheses  $H_1$  and  $H_2$ ,  $\Pr(H_1)=\Pr(H_2)$ , then if  $H_1$  better excels with respect to the informational virtues than  $H_2$ , then we ought to accept  $H_1$  as more worthy of commitment. Crucially, however, the fact that  $H_1$  exceeds  $H_2$  in explanatory content, i.e. better manifests the informational virtues, does not make  $H_1$  more likely to be true. It’s possible that both  $H_1$  and  $H_2$  are acceptable, if for example, the explanatory content of  $H_2$  is also sufficiently high. After all, although it is irrational to believe inconsistent theories, it is not irrational to accept inconsistent theories, as demonstrated by the case of the 19th conflict between physics and biology

On this point, however, one might object that since the notion of acceptance that I spell out includes probable truth as a component, it would seem to follow that no one can accept both  $H_1$  and  $H_2$ , since  $\Pr(H_1\&H_2)=0$ . Thus, it seems that it isn’t possible to commit to competing theories.<sup>41</sup> In my view, this problem can be avoided. Recall that contradictions don’t seem explanatory. If that’s true, then  $H_1\&H_2$  doesn’t have any explanatory content, and so its acceptability value is zero. Even though  $H_1$  is explanatory and  $H_2$  is explanatory, it does not follow that the pair is explanatory. The sense in which both  $H_1$  and  $H_2$  are acceptable is that scientists would be justified in committing either to  $H_1$  or  $H_2$ , or both individually, working on  $H_1$  and  $H_2$  at different times, perhaps. The conjunction  $H_1\&H_2$  is not acceptable, but this is no problem because  $H_1\&H_2$  is not explanatory

## **Conclusion**

As I have argued, if we distinguish between the probability and the acceptability of a hypothesis, then we can see why Bayesianism and IBE are compatible. IBE is a theory of when we should

---

<sup>41</sup> I owe this objection to an anonymous reviewer.

accept a hypothesis in one particular sense, and Bayesianism tells us what credence we ought to assign to a hypothesis. While the question of when we should accept a hypothesis is in part determined by how highly probable the hypothesis happens to be—since truth is a goal of science—IBE can, when cast as a theory of acceptance, respect all that Bayesianism has to say about the matter of truth. However, since Bayesianism is silent on the issue of acceptance, and probable truth does not exhaust the conditions under which a hypothesis is worthy of commitment, there is some substantive, interesting role for the explanatory virtues to play in our account of scientific practice.

In my view, it is reasonable to think of IBE as a theory of acceptance. While the Objective Bayesian approach to compatibilism discussed in section 3 amounts to a coherent position, the extra normative constraints proposed by the view are difficult to defend. This is ultimately, as I have argued, because IBE upholds as confirmational virtues certain properties that do not necessarily have confirmational import. In fact, IBE's mixing confirmational and informational virtues is what gives the illusion that we are warranted in *believing* the hypothesis that is the best explanation *in virtue of* its being the best explanation. To be sure, sometimes the best explanation is one that is highly probable. Supposing that IBE is a theory of acceptance, then the hypothesis we are warranted in accepting is sometimes the hypothesis that we are also warranted in believing. But when we are warranted in believing a hypothesis H, that is because the value of  $\text{Pr}(H)$  is sufficiently high. But the reason that  $\text{Pr}(H)$  is high is because of the confirmational virtues, most of which, it seems, can be explicated from within the Bayesian framework. Even though the non-confirmational explanatory virtues upheld by IBE-ists do not make H more likely to be true, they do make H a better hypothesis because they give H greater explanatory content, and thus make H more worthy of commitment as a research program.

## References

- Ben-Menahem, Y. (1990). "The Inference to the Best Explanation," *Erkenntnis*, 33 (3): 319-344.
- Betz, G. (2013). "Justifying inference to the best explanation as a practical meta-syllogism on dialectical structures," *Synthese*, 190: 3553-3578.
- Boyd, R. (1983). "On the Current Status of Scientific Realism," *Erkenntnis*, 19(3): 45-90.
- Bratman, M. (1999). *Faces of Intention*. Cambridge: Cambridge University Press.
- Brown, M. (2013). "Values in Science Beyond Underdetermination and Inductive Risk," *Philosophy of Science*, 80(5): 829-839.
- Callender, C. and Huggett, N. (2001). "Introduction," in C. Callender and N. Huggett (eds.), *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity*. Cambridge: Cambridge University Press.
- Campos, D. (2009). "On the Distinction between Peirce's Abduction and Lipton's Inference to the Best Explanation," *Synthese*, 180 (3): 419-442.
- Carnap, R. (1962). *Logical Foundations of Probability*, 2<sup>nd</sup> ed. Chicago: University of Chicago Press.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Cohen, J. (1992). *An Essay on Belief and Acceptance*. New York: Oxford University Press
- Darwin, C. (1962). *On the Origin of Species*. New York: Collier.
- Dawes, G. (2013). "Belief is Not the Issue: A Defence of Inference to the Best Explanation," *Ratio*, 26(1): 62-78.
- Douglas, H. (2000). "Inductive Risk and Values in Science," *Philosophy of Science*, 67(4): 559-579.
- Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
- Douven, I. (2005). "Wouldn't It Be Lovely: Explanation and Scientific Realism," *Metascience*, 14: 338-343.
- Douven, I. and Schupbach, J. (2015). "Probabilistic alternatives to Bayesianism: the case of explanationism," *Front. Psychol.*, (6): 459, doi: 10.3389/fpsyg.2015.00459.
- Douven, I. and Wenmackers, S. (2015). "Inference to the Best Explanation versus Bayes's Rule in a Social Setting," *Brit. J. Phil. Sci.*, in press.
- Elliott, K. (2011). "Direct and Indirect Roles for Values in Science," *Philosophy of Science*, 78: 303-328.
- Elliott, K.C. and McKaughan, D.J. (2014). "Non-Epistemic Values and the Multiple Goals of Science," *Philosophy of Science*, 81(1): 1-21.
- Elliott, K.C. and Willmes, D. (2013). "Cognitive Attitudes and Values in Science," *Philosophy of Science*, 80(5): 807-817.
- Fitelson, B. (1999). "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science*, 66: S362-78.
- Forster, M. and Sober, E. (1994). "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions," *Brit. J. Phil. Sci.*, 45(1): 1-35.
- Friedman, M. (1974). "Explanation and Scientific Understanding," *The Journal of Philosophy*, 71(1): 5-19.
- Glymour, C. (1980). "Explanations, Tests, Unity and Necessity," *Noûs*, 14(1): 31-50.
- Glymour, C. (2015). "Probability and the Explanatory Virtues," *Brit. J. Phil. Sci.*, 66(3): 591-604.
- Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Goodman, N. (1954). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.
- Harman, G. (1965). "The Inference to the Best Explanation," *Philosophical Review*, 74: 88-95.
- Hempel, C. and P. Oppenheim. (1948). "Studies in the Logic of Explanation," *Philosophy of Science*, 15: 135-175.
- Hempel, C., (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

- Henderson, L. (2014). "Bayesianism and Inference to the Best Explanation," *Brit. J. Phil. Sci.*, 65(4): 687-715.
- Hitchcock, C. (2007). "The lovely and the probable," *Philosophy and Phenomenological Research*, 74(2): 433-440.
- Hintikka, J. and Pietarinen, J. (1966). "Semantic Information and Inductive Logic," in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*. Amsterdam: North-Holland, 96-112.
- Huber, F. (2007). "Hempel's Logic of Confirmation," *Philosophical Studies*, 139(2): 181-9.
- Huemer, M. (2009). "Explanationist Aid for the Theory of Inductive Logic," *Brit. J. Phil. Sci.*, 60: 345-75.
- Janssen, M. (2002). "COI stories: Explanation and evidence from Copernicus to Hockney," *Perspectives on Science*, 10: 457-522.
- Jeffreys, W. and Berger, J. (1992). "Ockham's Razor and Bayesian Analysis," *American Scientist* 80: 64-72.
- Jeffrey, R. (1983). *The Logic of Decision*, 2<sup>nd</sup> ed. Chicago: University of Chicago Press.
- Jeffreys, H. (1931). *Scientific Inference*. London: Macmillan.
- Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.
- Kitcher, P. (1989). "Explanatory Unification and the Causal Structure of the World," in P. Kitcher and W. Salmon (eds.), *Scientific Explanation*, Vol. 13, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 410- 505.
- Kourany, J. A. (2003). "A Philosophy of Science for the Twenty-First Century," *Philosophy of Science*, 70(1): 1-14.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*, 1<sup>st</sup> ed. Chicago: University of Chicago Press (2<sup>nd</sup> ed. 1970).
- Kuhn, T. (1977). *The Essential Tension*. Chicago: University of Chicago Press, 320-339.
- Kukla, A. (1994). "Non-Empirical Theoretical Virtues and the Argument from Underdetermination," *Erkenntnis*, 41(2): 157-170.
- Kyburg, H. (1961). *Probability and the Logic of Rational Belief*. Middletown: Wesleyan University Press.
- Lacey, H. (1999). *Is Science Value-Free?: Values and Scientific Understanding*. London: Routledge.
- Lange, M. (2004). "Bayesianism and Unification: A Reply to Wayne Myrvold," *Philosophy of Science*, 71(2): 205-215.
- Laudan, L. (1977). *Progress and its Problems: Toward a Theory of Scientific Growth*. Berkeley, CA: University of California Press.
- Levi, I. (1960). "Must the Scientist Make Value Judgments?" *Journal of Philosophy*, 57: 345-57.
- Levi, I. (1967). *Gambling With Truth: An Essay on Induction and the Aims of Science*. Cambridge, MA: MIT Press.
- Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis, D. (1981). "A Subjectivist's Guide to Objective Chance," *The University of Western Ontario Series in Philosophy of Science*, 15: 267-297.
- Lipton, P. (2001). "Is Explanation a Guide to Inference?: A Reply to Wesley C. Salmon," in G. Hon and S. S. Rakover (eds.), *Explanation: Theoretical Approaches and Applications*, Dordrecht: Kluwer, 93-120.
- Lipton, P. (2004). *Inference to the Best Explanation*, 2<sup>nd</sup> ed. New York: Routledge.
- Lipton, P. (2007). "Replies," *Philosophy and Phenomenological Research*, 74(2): 433-440.
- Longino, H. (1996). "Cognitive and Non-cognitive Values in Science: Rethinking the Dichotomy," in Nelson and Nelson (eds.), *Feminism, Science, and the Philosophy of Science*. Dordrecht: Kluwer, 39-58.
- Lord Kelvin. (1862). "Physical Considerations Regarding the Possible Age of the Sun's Heat," *Philosophical Magazine and Journal of Science*, 23(4): 158-160.

- Lycan, W.G. (1988). *Judgement and justification*. Cambridge: Cambridge University Press.
- Lycan, W.G. (2002). "Explanation and epistemology," in P. Moser (ed.), *Oxford Handbook of Epistemology*. Oxford: Oxford University Press, 408–33.
- Machamer, P., Darden, L., and Craver, C. (2000). "Thinking About Mechanisms," *Philosophy of Science* 67: 1-25.
- Mackonis, A. (2013). "Inference to the best explanation, coherence and other explanatory virtues," *Synthese*, 190(6): 975-995.
- Maher, P.(1993). *Betting On Theories*. Cambridge: Cambridge University Press.
- McGrew, T. (2003). "Confirmation, heuristics, and explanatory reasoning," *Brit. J. Phil. Sci.*, 54 (4): 553-567.
- McMullin, E. (1983). "Values in Science," in Asquith and Nickles, (eds.), *Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association, Volume 1*. East Lansing: Philosophy of Science Association, 3-28.
- Minnameier, G. (2004). "Peirce-suit of truth: Why inference to the best explanation and abduction ought not to be confused," *Erkenntnis*, 60(1): 75–105.
- Myrvold, W. (2003). "A Bayesian Account of the Virtue of Unification," *Philosophy of Science* 70: 399-423.
- Myrvold, W. (2015). "On the Evidential Import of Unification," Unpublished MS. <http://philsci-archive.pitt.edu/11670/>
- Nyrup, R. (2014). "How Explanatory Reasoning Justifies Pursuit: A Peircean View of IBE," Unpublished MS. <http://philsci-archive.pitt.edu/11126/>.
- Okasha, S. (2000). "Van Fraassen's Critique of Inference to the Best Explanation," *Studies in the History and Philosophy of Science*, 31: 691-710.
- Peirce, C. S. (1934). *The Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.
- Popper, K. (1980). *The Logic of Scientific Discovery*. London: Routledge.
- Poston, T. (2014). *Reason & Explanation: A Defense of Explanatory Coherentism*. New York: Palgrave-MacMillan.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks the Truth*. London: Routledge.
- Psillos, S. (2002). "Simply the Best: A Case for Abduction," in A. C. Kakas and F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond*. Berlin: Springer-Verlag, 605-26
- Psillos, S. (2004). "Inference to the Best Explanation and Bayesianism," in Stadler (ed.), *Induction and Deduction in the Sciences*. Dordrecht: Kluwer, 83–91.
- Psillos, S. (2007). "The Fine Structure of Inference to the Best Explanation," *Philosophy and Phenomenological Research*, 74(2): 441–448.
- Putnam, H. (1975). "What is Mathematical Truth?" in *Mathematics, Matter and Method: Philosophical Papers, volume 1*. Cambridge: Cambridge University Press, 60-78.
- Quine, W. V. O. and Ullian, J. (1978). *The Web of Belief*. New York: Random House.
- Roche, W. and Sober, E. (2013). Explanatoriness is evidentially irrelevant, or inference to the best explanation meets Bayesian confirmation theory," *Analysis*, 73: 659-668.
- Rooney, P. (1992). "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" in Hull, Forbes, and Okruhlik (eds.), *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*. East Lansing, MI: Philosophy of Science Association, 3-28.
- Rott, H. (2004). "The Value of Truth and the Value of Information: On Isaac Levi's Epistemology," in E. Olsson (ed.), *Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi*. Cambridge: Cambridge University Press. 179-200.
- Rudner, R. (1953). "The Scientist Qua Scientist Makes Value Judgments," *Philosophy of Science*, 20(1): 1-6.

- Salmon, W. (1971). "Statistical Explanation", in W. Salmon, (ed.), *Statistical Explanation and Statistical Relevance*, Pittsburgh: University of Pittsburgh Press, 29–87.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. (1989). *Four Decades of Scientific Explanation*, Minneapolis: University of Minnesota Press.
- Salmon, W. (1990). "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes," In C. Wade Savage (ed.), *Scientific Theories*, vol. 14. Minneapolis: University of Minnesota Press, 175-204.
- Salmon, W. (2001a). "Explanation and Confirmation: A Bayesian Critique of Inference to the Best Explanation," in G. Hon and S. S. Rakover (eds), *Explanation: Theoretical Approaches and Applications*, Dordrecht: Kluwer, 61-91.
- Salmon, W. (2001b) "Reflections of a Bashful Bayesian: A Reply to Peter Lipton," in G. Hon and S. S. Rakover (eds.), *Explanation: Theoretical Approaches and Applications*, Dordrecht: Kluwer, 121-136.
- Schupbach, J. (2005). "On a Bayesian Analysis of the Virtue of Unification," *Philosophy of Science*, 72(4): 594-607.
- Sider, T. (2013). "Against Parthood," in K. Bennett and D. Zimmerman, (eds.), *Oxford Studies in Metaphysics*, vol. 8. Oxford: Oxford University Press, 237–93.
- Sober, E. (1990). "Explanation in Biology: Let's Razor Ockham's Razor," *Royal Institute of Philosophy Supplement*, 27: 73-93.
- Sober, E. (2001). "Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause," 52(2):331–346.
- Sober, E. (2002). "Bayesianism—Its Scope and Limits" in R. Swinburne, (ed.), *Bayes' Theorem, Proceedings of the British Academy Press*, 113, 21-38.
- Sober, E. (2015). *Ockham's Razors: A User's Manual*. Cambridge: Cambridge University Press.
- Thagard, P. (1978). The Best Explanation: Criteria for Theory Choice," *The Journal of Philosophy*, 75(2): 76-92.
- Thagard, P. (1993). *Computational philosophy of science*. Cambridge, MA: MIT Press.
- 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: 125–43.
- Whewell, W. (1968). *William Whewell's Theory of Scientific Method*. R. Butts (ed.), Pittsburgh, PA: University of Pittsburgh Press.
- White, R. (2005). "Explanation as a Guide to Induction," *Philosophers' Imprint*, 5(2): 1-29.
- Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.