Penultimate Draft – to appear in *Handbook of Abductive Cognition*, (ed. L. Magnani.), Springer.

Inference to the Best Explanation – An Overview

Frank Cabrera

Abstract: In this article, I will provide a critical overview of the form of non-deductive reasoning commonly known as "Inference to the Best Explanation" (IBE). Roughly speaking, according to IBE, we ought to infer the hypothesis that provides the best explanation of our evidence. In section 2, I survey some contemporary formulations of IBE and highlight some of its putative applications. In section 3, I distinguish IBE from C.S. Peirce's notion of abduction. After underlining some of the essential elements of IBE, the rest of the entry is organized around an examination of various problems that IBE confronts, along with some extant attempts to address these problems. In section 4, I consider the question of when a fact requires an explanation, since presumably IBE applies only in cases where some explanation is called for. In section 5, I consider the difficult question of how we ought to understand IBE in light of the fact that among philosophers, there is significant disagreement about what constitutes an explanation. In section 6, I consider different strategies for justifying the truth-conduciveness of the explanatory virtues, e.g., *simplicity*, *unification*, *scope*, etc., criteria which play an indispensable role in any given application of IBE. In section 7, I survey some of the most recent literature on IBE, much of which consists of investigations of the status of IBE from the standpoint of the Bayesian philosophy of science.

Keywords: Abduction, Inference to the Best Explanation, Explanatory Virtues, Bayesianism, Scientific Inference

1. Introduction

In this entry, I will provide a critical overview of the form of non-deductive reasoning commonly known as "Inference to the Best Explanation" (IBE). Roughly speaking, according to IBE, we ought to infer the hypothesis that provides the best explanation of our evidence. In section 2, I survey some contemporary formulations of IBE and highlight some of its putative applications. In section 3, I distinguish IBE from C.S. Peirce's notion of abduction, the primary difference being that abduction is a method of generating hypotheses, whereas IBE is a method of evaluating rival hypotheses. After underlining some of the essential elements of IBE, the rest of the entry is organized around an examination of various problems that IBE confronts, along with some extant attempts to address these problems. In section 4, I consider the question of when a fact requires an explanation, since presumably IBE applies only in cases where some explanation is called for. In section 5, I consider the difficult question of how we ought to understand IBE in light of the fact that among philosophers, there is significant disagreement about what constitutes an explanation. In section 6, I consider different strategies for justifying the truth-conduciveness of the explanatory virtues, e.g., *simplicity*, *unification*, *scope*, etc., criteria which play an indispensable role in any given application of IBE.

In section 7, I survey some of the most recent literature on IBE, much of which consists of investigations of the status of IBE from the standpoint of the Bayesian philosophy of science. As I conclude, this recent work, and the more general application of formal tools in the study of explanatory reasoning, will no doubt continue to push discussions of IBE in promising new directions.

2. What Is Inference to the Best Explanation?

2.1 IBE as Four-Step Argument Schema

Textbook accounts of scientific method give pride and place to *prediction*. We start with a hypothesis H, which when conjoined with some background assumptions entails some novel fact F. When we observe F, perhaps after conducting an experiment, we regard this observation as evidence for, or confirmation of, H. Accordingly, many prominent philosophical models of scientific reasoning emphasize the importance of prediction, especially of novel facts (e.g., Whewell 1840/1968; Popper 1959; Lakatos 1978). However, sometimes we think that evidential support accrues to a hypothesis, not because of its ability to successfully predict novel facts, but because of its ability to *explain*, in an extraordinarily satisfying way, some particular fact or set of facts that we have already observed. In short, sometimes inference is mediated by explanation.

As an illustration, consider an oft-cited passage from the end of the 6th edition of *On the Origin of Species,* in which Darwin reflects upon the nature of his "one long argument":

> I have now recapitulated the chief facts and considerations which have thoroughly convinced me that species have been modified, during a long course of descent, by the preservation or the natural selection of many successive slight favourable variations. I cannot believe that a false theory would explain, as it seems to me that the theory of natural selection does explain, the several large classes of facts above specified (1872: 421).

The "several large classes of facts" to which Darwin refers include a) the existence of manifest anatomical similarities in the bone structure of the hand of a human, the wing of a bat, and the fin of a porpoise, and b) the existence of vestigial structures "organs bearing the stamp of inutility"—such as the useless teeth of the embryonic calf and shriveled wings of some beetles (1872: 420).

As a second illustration, consider another case of a more everyday sort. Suppose I notice a full six-pack of beer in my refrigerator on a Thursday morning, and then return late that evening to notice that one bottle, and nothing else, is unexpectedly missing. Although there are many possible beliefs that I might form in light of this datum—including, for example, that my home was burglarized by

beer-guzzling aliens—in this case, I merely draw the inference that my wife—the only other occupant of my house— had a beer while I was away. To be sure, were I in fact the victim of an intergalactic heist, that would explain the missing beer bottle. But this far-fetched hypothesis is obviously a poor way of explaining the relevant facts. Realizing this, I do not infer that any aliens visited my home.

According to one prominent view among philosophers of science and epistemologists, how I go about inferring that my wife consumed the bottle of beer and how Darwin summarized his argument for evolution by natural selection have something important in common. Although the subject matters are quite different, both inferences are instances of the same argument form, at least at a very abstract level of description. This argument form is commonly called "Inference to the Best Explanation" (IBE). As Lipton (2004): 56) puts it, "the core idea of Inference to the Best Explanation is that explanatory considerations are a guide to inference." In cases in which we rely on IBE, we infer some hypothesis only after first taking an "explanatory detour" (Lipton 2004: 65). Contemporary discussion of IBE traces back to Harman (1965), who appears to have coined the term. Harman's initial characterization of IBE is as follows:

> 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).

Following Harman's informal characterization, proponents of IBE tend to formulate the inference method as a four-step argument schema. Existing schematizations of IBE differ in some minor details but share their essential elements in common Consider, for instance, Lycan's (2002: 413) formulation:

> (1) F_1, F_2, …, F_n are facts in need of explanation.
> (2) Hypothesis H explains the F_i.
> (3) No available competing hypotheses would explain the F_i as well as H does.
> Therefore, (4) H is true.

Here, Lycan (2002)'s formulation agrees with Harman (1965)'s characterization, according to which, a successful application of IBE permits us to believe that the conclusion is true, full-stop. Elsewhere, Lycan (1988: 129) gives a slightly different formulation of IBE, where the conclusion of the argument is not that "H is true", but rather, more modestly "[probably] H is true".

On the basis of Lycan's formulation, Josephson and Josephson (1994: 5) provide a very similar schematization of IBE:

(1) D is a collection of data (facts, observations, givens).
(2) H explains D (would, if true, explain D).
(3) No other hypothesis can explain D as well as H does.
Therefore, (4) H is probably true.

This formulation by Josephson and Josephson (1994) is endorsed by Psillos (2002), another advocate of IBE. Importantly though, Psillos clarifies that the word "probably" does not involve a commitment to any particular interpretation of the probability calculus; rather, the inclusion of the word "probably" is intended to convey only that the conclusion of IBE is non-demonstrative (2002: 614, fn. 17).

While "IBE is typically seen as a rule of acceptance" (Psillos 2004: 83), one might also adopt an even more modest view, according to which the conclusion of a successful instance of IBE is just that E provides inductive support for H, but not necessarily conclusive justification to believe H.[1] In most accounts of IBE, explanatory goodness is cashed out in terms of a list of intuitive, albeit imprecise, "explanatory virtues", including such criteria as *simplicity*, *unification*, *precision*, *conservativeness*, *scope*, etc. The hypothesis that does best, overall, with respect to the explanatory virtues is the one that constitutes the best explanation and is thus the one that we ought to prefer.

2.2 The Importance of Background Knowledge

Although standard formulations of IBE counsel us to consider multiple different explanations for the phenomena, it is of course impossible to consider every conceivable explanation for the facts. We can only compare explanations that are available to us. Some explanations doubtless will fail to be considered. As is often discussed in accounts of the "underdetermination problem", an infinite number of theories are logically compatible with any given set of data (e.g., Duhem 1906/1954: 151-3). This problem is easiest to see in cases of curve-fitting, since an infinite number different of mathematical equations can go through any finite set of data points (Ladyman 2002: 164-5).

In light of our inability to consider every possible explanation for the phenomena, van Fraassen articulates a famous objection to IBE known as the "argument from the bad lot" (1980: 142-3). If we have a such a subset X of potential hypotheses $H_1 \ldots, H_n$ that each offer an explanation of the evidence E, and if we want to say that H_1 is the best explanation of E, and so, in accordance with IBE, that H_1 is probably true, or should be inferred, etc., then we must assume that the true explanation is

[1] While Lipton often writes as though IBE is a rule of acceptance, which results in "full belief", he also expresses sympathy with the idea that IBE is a theory of confirmation, which deals with "degrees of belief" (2004: 63). Much of the most recent literature on IBE, which considers the relationship between IBE and Bayesianism, assumes that IBE is a theory of confirmation, not a rule of acceptance. I discuss this literature in section 7.

more likely than not to be found in X. However, many potential explanations haven't even been conceived of yet, and so they won't lie in X.[2] According to van Fraassen, IBE requires that we assume that it is typically more likely than not that the true hypothesis lies within the subset of explanations being considered, but it is not clear how we could be justified in making this assumption.

One plausible response to the bad lot objection is to first highlight that any account of scientific methodology should take seriously the point that scientific inference is always conducted against a particular set of background knowledge (or assumptions). There's little hope in trying to characterize the nature of scientific inference in purely logical or syntactic terms, without taking into account the fact that the practice of science always takes place in a particular context in which certain facts are known or taken for granted. The project of trying to characterize the confirmation relation relative to an empty background set suggested by Hempel (1965: 20) has not had a fruitful track-record. In light of criticism by Good (1967) and others, it seems a plausible conjecture, as Sober puts it (1988: 59) that, "confirmation is a three-place relationship between hypotheses, observations, and background assumptions."

The key importance of background knowledge helps to defuse the bad lot objection, for the proponent of IBE can argue that background knowledge ensures that only plausible potential explanations are considered (e.g., Psillos 1996, Lipton 2004: 151-63). For instance, as Psillos argues, "background knowledge can drastically narrow down the space for hypotheses that provide a potential explanation of the evidence at hand" (1996: 39). The crucial role that background knowledge plays is made manifest in Lipton's (2004: 59-61) "two-filter" account of IBE, according to which the pool of all possible explanations is first winnowed considerably by considerations of initial plausibility. Relatedly, Lipton adds the important caveat to his account of IBE that "the best explanation must be good enough to merit inference" (2004: 64).[3] Thus, by appealing to background knowledge, only serious candidates make it to the stage where the explanatory virtues are considered, after which, we infer the best explanation, provided that explanation is "sufficiently good" (Lipton 2004: 154).[4]

To be sure, sometimes the background set is defective. Einstein's photon hypothesis that light must (in some sense) be *particulate*, posited in order to explain the photoelectric effect, was in deep

[2] This line of reasoning is not merely hypothetical. As Stanford (2006) argues, an honest assessment of the history of science will demonstrates a worrisome "problem of unconceived alternatives."

[3] See Dellsén (2021) for a sophisticated account of what it means for an explanation to be "good enough."

[4] Bird (2020: 49) also mentions an important corollary of Lipton's caveat, namely that the best explanation should also be "significantly better than its nearest rival."

conflict with the dominant view of light at the time. The received then view was the theory that light is a wave, which was thought to be the only way to explain facts about refraction and interference. For about next 20 years that followed the publication of his paper "Concerning an Heuristic Point of View Toward the Emission and Transformation of Light" (1905), Einstein's hypothesis was met with almost widespread rejection.[5] Here, the set of background theories accepted by the physics community at the time was having an undesirable influence, leading us further away from the truth.

All this incident goes to show, however, is that inductive inference mediated by background assumptions can sometimes lead us astray. But this is something that we should not be surprised by. Inductive arguments, as everyone knows, are not intended to be infallible. Of course, if the proponent of the bad lot objection doubts the legitimacy of the background set, then this would undercut the possibility of employing IBE. What if the background set is just another bad lot? But this is a skeptical worry that could be raised against any account of scientific inference. If the set of background beliefs is hopelessly defective, then any non-deductive method of inference will lead one astray.[6] This is true for IBE just as much as it is true for competing accounts of scientific inference. However, doubting the legitimacy of all background theories, by which we narrow the array of possible explanations, is not really a live option, for that amounts to acquiescing to inductive skepticism.

2.3 Some Further Applications of IBE

Problems posed by underconsideration notwithstanding, it would be easy to multiply examples of plausible instances of IBE. According to Douven (2021), our use of IBE is "so routine and automatic that it easily goes unnoticed."[7] For example, I might infer that my dog chewed up my pillow upon observing the pillow stuffing in the dog's bed; or I might infer that there is a mouse in the house after hearing squeaks and scratching behind the stove; or I might infer that there was a power outage

---

[5] For example, in 1913 Max Planck wrote a letter defending Einstein's nomination to the Prussian Academy of Sciences, writing of Einstein that, although "he may have gone overboard in his speculations", this "should not be held too much against him" (quoted in Fölsing 1997: 147). Harsher was the critique in 1916 by experimental physicist Robert Millikan, who called Einstein's view a "bold, not to say reckless, hypothesis...which flies in the face of thoroughly established facts of interference" (Millikan 1916: 355). Even as late as 1922, Niels Bohr remarked, during his Nobel prize acceptance speech, that "[t]he hypothesis of light quanta...is not able to throw light on the nature of radiation" (Bohr 1922: 14).

[6] See Schupbach (2014), who explores something akin to this line of reasoning in greater depth. For a recent critique of some prominent responses to the bad lot objection, see Dellsén (2017)

[7] See Douven and Mirabile (2018) and the references contained therein for empirical evidence suggesting that the use of IBE in everyday life is quite common.

after observing all the lights in my house suddenly turning off, etc. Furthermore, some philosophers have argued that we gain testimonial knowledge by way of an inference to the best explanation (e.g., Fricker 2018). If it is true that testimonial knowledge requires an inference to the best explanation, then that would mean that we gain a lot of knowledge by using IBE, since we gain much of our knowledge of the world ultimately from the testimony of those in our social group (Fricker 2004). So too, others argue that language acquisition requires the use of IBE (Hobbs 2004; Fricker 2017).

Furthermore, many philosophers have maintained that prominent episodes from the history of science can be best explicated as application of IBE. In addition to the Darwin case (Okasha 2000) which we have already considered, Lipton (2004: 74-90) argues at length that the research conducted by Ignaz Semmelweis in the 1840s on the causes of childbed fever provide a superb case of IBE in action. So too, in a detailed case study of the reasoning of John Snow, a mid-19th century physician who studied the causes and transmission of cholera, Tulodziecki (2011: 315) argues that Snow's reasoning makes "appeals to explanatory power at virtually every juncture." In addition to these extended case studies, there is a long list of cases from the history of science in which philosophers have argued or asserted that IBE was used by practicing scientists. This list includes the Copernican argument for the heliocentric model of the solar system (Gauch 2012); Couch and Leverrier's inference to the existence of a hitherto unknown planet, i.e. Neptune, as the best explanation of the aberrant orbit of Uranus (Douven 2021); Thomson's discovery of the electron (Achinstein 2001); Priestley's discovery of photosynthesis (Matthews 2015); Einstein's argument that the theory of special relativity is preferable to Lorentz's modification of the then-reigning ether theory (Janssen 2002); Lavoisier's rejection of phlogiston and defense of the oxygen theory of combustion (Thagard 1978); and Huygens' argument for the wave theory of light over Newton's corpuscular theory (Thagard 1978). Finally, some practicing scientists have even recommended a more self-conscious employment of IBE, for example, in psychology (Haig 2009) and in archaeology Campanaro (2021).

Of importance also is the fact that philosophers themselves have explicitly relied on IBE in a number of diverse contexts. Again, here it would be easy to multiply examples, especially since it has become increasingly common in meta-philosophical discussions to endorse IBE as an indispensable tool for doing philosophy (Biggs and Wilson 2017). For instance, some philosophers have employed IBE to defend scientific realism (Psillos 1999), to combat external world skepticism (Vogel 1990), to defend mathematical Platonism (Baker 2009), to defend moral anti-realism (Leiter 2014), etc. Oftentimes, an argument for the use of IBE to support distinctively philosophical conclusions is made on the basis of the use of IBE in the sciences and in ordinary life, (e.g., Swoyer 1999; Hawley 2006;

Sider 2009; Paul 2012; Williamson 2016). Since science is a paradigmatically rational enterprise, if it's permissible for the scientist to use IBE to draw conclusions, then we should be able to use that same method in philosophy—or so the argument goes.[8] Some philosophers even go so far as to argue that all epistemic justification requires explanatory reasoning in some form or another (Lycan 1988; Conee and Feldman 2008; Poston, 2014). One need not accept this quite bold view about the scope of epistemic justification in order to think that IBE is a common and useful way of reasoning, of course.[9]

To be sure, IBE is an intuitive idea and has the weight of many different examples behind it; still, the inference pattern raises a host of difficult questions. As McCain and Poston note in an edited volume comprising new essays on IBE, "despite its widespread use in science and everyday life, the nature and use of IBE are hotly contested" (2017: 1). Indeed, each step in the argument schema demands clarification and defense, as we will soon see. Clarifying the various steps of the argument schema will be necessary if only to push back against the dismissive charge, raised by some critics, that "'inference to the best explanation' is nothing but a slogan" (Salmon 2005: 58). [10]

3. Peirce's Abduction and IBE: An Aside

Before doing that, though, it is worth mentioning that the schematizations of IBE considered din section 2.1 bear a striking resemblance to the pattern of reasoning that C.S. Pierce called "abduction"—a word which he coined from the Greek "apagoge", a term Aristotle uses in the *Prior Analytics* to denote a similar sort of reasoning. It is common in contemporary discussions of IBE to link abduction to IBE in some way, either to directly equate IBE with abduction, or to trace IBE back to Peirce, or at least to use "IBE" and "abduction" interchangeably (e.g., Harman 1965; Thagard 1978; Lycan 1988; Barnes 1995; Lipton 2004; Douven 2021). According to Peirce, "[r]easoning is of three types, Deduction, Induction, and Abduction" (1934, CP 5.161). Although sometimes "induction" is used more broadly to refer to any form of non-deductive reasoning, for Peirce, however, induction and abduction are irreducibly distinct modes of inference. In his *Harvard Lectures on Pragmatism*, Peirce puts forward his most precise formulation of abduction (1934, CP 5.189):

---

[8] Sober (2009) discusses this meta-philosophical view, which he calls "Methodological Naturalism$_P$".

[9] Lycan (2002: 417) helpfully delineates four different versions of "explanationism" of varying degrees of strength and universality.

[10] Similarly, in a chapter critical of IBE, Norton (2021: 270) concludes "inference to the best explanation is not a self-contained rule of inductive inference. It is at best a loose guide in urgent need of development."

(1) The surprising fact, *C*, is observed.
(2) But if *A* were true, *C* would be a matter of course.
Hence, (3) there is reason to suspect that *A* is true.

While Peirce's notion of abduction is often cited as the intellectual forbearer of IBE, this view has been rejected by a great number of commentators (e.g., Minnameier 2004; McKaughan 2008; Campos 2011; McAuliffe 2015; Park 2015).[11] Indeed, we can already see some differences between the schemas put forth by contemporary proponents of IBE and the abductive schema that Peirce puts forth.

For one thing, although Peirce thinks a good abduced hypothesis "must explain the facts" (1934, CP 5.1917)—demonstrating that he has in mind by his locution "*C* would be a matter of course" simply that *A* explains *C*—his formulation of abduction lacks the comparative component that features prominently in the four-step schematizations of IBE. Nowhere in Peirce's formulation is it said that *A* is the best explanation of *C*. Pierce does not tell us, moreover, to examine various competing hypotheses in light of criteria for explanatory goodness. For another thing, and perhaps owing to the previous point, the conclusion of Peirce's abductive inference is much weaker than that which is represented by most formulations of IBE. Peirce doesn't commit himself to the truth of the hypothesis *A* abduced on the basis of *C*. Instead, Peirce states that "the conclusion of an abduction is problematic or conjectural" (1934, CP 5.192). From Peirce's formal statement of abduction, we cannot even conclude that *A* is probably true or even approximately true. Rather, we only have "reason to suspect" that *A* is true. This conclusion is compatible with the claim that we have reason to suspect *B* is true, where *B* is an incompatible, rival hypothesis, which is also the result of abduction. Perhaps, it is also the case that if *B* were true, then *C* would be a matter of course. Indeed, Peirce clearly states that although induction determines what is "actually operative", abduction "merely suggests that something may be" (1934, CP 5.171). Thus, it is evident the conclusion of a Peircean abduction is rather weak, certainly much weaker than standard formulations of IBE.

What is more, it is doubtful that Peircean abduction is even a species of evaluative inference at all. According to Peirce, "abduction is the process of forming an explanatory hypothesis"; and, moreover, abduction "is the only logical operation which introduces any new idea" (1934 CP 5.171). Owing to remarks such as these, those who reject the identification of abduction with IBE, including some those dissenting commentators cited above, have argued that Peircean abduction

[11] However, see Mohammadian (2021) for a defense of a middle ground position, according to which "abduction and IBE have important similarities as well as differences" (4205).

belongs not to the "context of justification", but rather to the "context of discovery."[12] For instance, Minnameier (2004, p. 75) argues, "while abduction marks the process of generating theories—or, more generally, concepts—IBE concerns their evaluation."[13] Similarly, McKaughan (2008) argues that Perice's abduction belongs, not to the context of justification, but to what Laudan (1977) once called the "context of pursuit." That is, the point of abduction is to help us determine when a theory is worthy of pursuit, which is related to, but not the same thing as a theory being justified, as pursuit-worthiness involves pragmatic factors, such as the "economy of research." By contrast, proponents of IBE typically claim that a strong, successful inference to the best explanation can rationally justify full-fledged belief in a hypothesis. In other words, it is standard to regard IBE as belonging to the context of justification, rather than the context of discovery or context of pursuit.[14]

Part of the reason for the interpretative confusion is that Peirce developed his notion of abduction over several decades, admitting toward the end of his life "in almost everything I printed before the beginning of this century I more or less mixed up hypothesis [i.e. abduction] and induction" (1934, CP 8.221). As a result, the concept was both expressed differently and evolved significantly throughout his career, eventually culminating in something quite different from the contemporary notion of IBE. Although there are certainly some similarities between Peirce's abduction and contemporary accounts of IBE, my focus will be on the later to the exclusion of the former. For this reason, I will avoid using "IBE" and "abduction" as synonyms whenever possible.

4. When Do Facts Require an Explanation?

Recall that in the first step of IBE, we must first start with some data, phenomena, facts, etc., which are known in advance, or at least provisionally accepted, before applying IBE, and which serve as a necessary springboard for the inference. These data might be acquired by way of sense-perception (e.g., I observe that a bottle of beer is missing), by way of a combination of observation and inductive inference (e.g., we infer that there is a correlation between smoking cigarettes and getting lung cancer on the basis of sample frequency data), or through some other form of reasoning. For convenience,

[12] See Reichenbach (1935) for the original distinction between the context of justification and the context of discovery. For critical discussion of this distinction, see the various entries in edited volume by Schickore and Steinle (2006).

[13] The idea that Peircean abduction is a method of hypothesis generation has been extensively explored by Hanson (1958, 1961), and more recently by (Magnani 2000, 2009) and Aliseda (2006, 2007).

[14] See Cabrera (2021a) for further discussion of the context of pursuit, both in general, and with respect to debates over the scientific legitimacy of string theory.

let's refer to those facts or data that are plugged into the first step of the argument schema as the "explanatory trigger." This helpful concept derives from the term "abductive trigger", which Aliseda (2006) coins to refer to the "surprising fact C" in Peirce's formulation of abduction.

Crucially, the reason that something can be used to initiate an act of explanatory reasoning is that those facts which constitute the explanatory trigger *require an explanation*. This subtle point is acknowledged in the formulation of IBE provided by Lycan (2002: 413), according to which the facts that begin the process of IBE are ones that are "in need of an explanation." Although Harman (1965) makes no mention of this "explanatory requirement" condition, and indeed some other schematizations of IBE exclude it (e.g., Psillos 2002), as Cabrera (2021b) argues, this qualification included by Lycan should not be neglected in our account of IBE. If some fact or set of facts F did not require an explanation, then to infer that H is true, or probably true because H explains F better than its competitors would amount to a non-sequitur. An intuition underlying IBE is that H ought to be awarded some justificatory "credit" for alleviating the epistemic "burden" imposed upon our web of beliefs by some unexplained fact. The epistemic standing of our web of beliefs is to some degree sub-optimal until we arrive at the best explanation of that fact that needs to be explained. By contrast, if we mistakenly believe that F requires an explanation, then of course it would be bizarre and unwarranted to "reward" H for providing some unwarranted explanation.

While Lycan (2002) is one of the few proponents of IBE to explicitly acknowledge the important point that those facts that initiate the inference process must first require an explanation, he does not provide an account of explanatory requirement. According to Lycan, doing so is not necessary when trying to give a philosophical account of IBE because "substantive disagreements about what does or does not need explaining are rare" (2002: 413). The implicit argument here seems to be that since people almost always agree on the extension of the concept, we can get by in our philosophical theorizing by leaving the concept of explanatory requirement intuitive. However, this argument appears quite weak once we consider analogous cases. For example, there is widespread agreement about the extension of the concept "knowledge." From a commonsense, non-skeptical point of view, we know quite a lot of things, e.g., Paris is the capital of France, there is something rather than nothing, Canada borders the United States, etc. Yet, even though there is a lot of agreement about the extension of the concept "knowledge", this consensus does not obviate the need to provide a philosophical analysis of "knowledge", as contemporary epistemologists engaged in this project would agree. Thus, it does not seem that widespread agreement about which facts require an explanation should dissuade us from exploring the conditions governing explanatory requirement.

What's more, Lycan underestimates the extent to which people disagree about whether a fact require an explanation. To be sure, debates about whether some application of IBE succeeds are usually located at a different stage of the argument schema. Often two parties that disagree agree about some application of IBE will at least agree that the fact being used as the explanatory trigger does indeed require an explanation; however, one party favors some hypothesis H_1 as the best explanation, whereas the other party favors some incompatible hypothesis H_2 as the best explanation. Still, disputes about which facts require an explanation are far from uncommon. In addition to cases in both philosophy and in science,[15] we can find many disputes about over whether some fact requires an explanation in ordinary life. In general, these debates arise whenever one person claims that an explanation for some series of events is required—such as the fact that Evelyn Adams won the New Jersey Lottery in 1985 and then again in 1986 (Sober 2012)—and the other party claims that this series of events is a coincidence, and thus requires no explanation.[16] We could easily multiply ordinary examples of *prima facie* "meaningful connections", where some people think that these connections require an explanation, whereas others prefer to think that such connections are best dismissed as a fluke or a mere coincidence.[17] The ease with which we can find or think of cases from ordinary life that fit this mold suggests that debates about whether some fact or set of facts requires an explanation are an salient element of our ordinary inductive and explanatory practices. Consequently, a complete account of IBE should have something to say about when a fact can serve as an explanatory trigger. Without some such account, one might worry that it is difficult to know when IBE ought to be applied.

Ideally, we would like to find an informative, general principle that provides the necessary and sufficient conditions for when a fact requires an explanation. There have been some attempts to formulate such a principle, which have proceeded independently of discussions of IBE. Unfortunately,

---

[15] In debates about the metaphysics of laws of nature, disagreements between Humeans (e.g., Lewis 1973) and non-Humeans (e.g., Armstrong 1983) partly turn on whether the fundamental, universal patterns found in nature *require* an explanation. Non-Humeans claim that an explanation is needed and are thus motivated to posit ontologically robust laws of nature to explain the patterns that we observe, whereas Humeans deny that this is an explanatory requirement (e.g., Swartz 1985: 204). In science, sometimes physicists disagree with each other over whether the "Past Hypothesis", i.e., the fact that shortly after the Big Bang the universe was in a state of very low entropy, requires an explanation (Baras and Shenker 2020).

[16] Here we must commit to the claim that "the joint occurrence of X and Y is a coincidence" implies that "the joint occurrence of X and Y has no explanation." Although this entailment is consonant with ordinary language usage—indeed, Lando (2016: 152) calls it a "truism"—and is sometimes made explicit, in for instance Lange's (2016) account of "mathematical coincidence", it is not entirely uncontroversial (e.g., Sober 2012).

[17] For example, twins, Jim Springer and Jim Lewis, separated at birth, twice married different women with the same name (Chen 1979).

however, as Baras (2019, 2020) has argued, all attempts to formulate a substantive principle of explanatory requirement have thus far have proved unsatisfactory. The most significant difficulty is that it is "unclear that there is any plausible non-trivial analysis" (Baras 2020: 1504) of the concept of "explanatory requirement", or related concepts such as "crying out for an explanation." Often, in efforts to articulate what it means for some fact F to require an explanation, philosophers propose analyses such as F requires an explanation if and only if F is "surprising", or "unexpected", or "striking" given our background knowledge (e.g., Horwich 1982; White 2005; Jakobsen 2020). However, such analyses do not seem to be of much help in telling us when IBE applies.

A natural way to interpret existing proposals is that they are just a "potentially confusing way" (Isaacs & Hawthorne 2018: 143) of expressing some claim about how probable some fact F is given our background beliefs. One might think then that we should regard the facts that require an explanation as simply those facts that are sufficiently improbable given our background knowledge, and we should regard the facts that do *not* require an explanation as those facts that are sufficiently probable given our background knowledge. But this simple and elegant analysis quickly runs afoul of counterexamples. For example, my winning the lottery is improbable given my background knowledge, but *somebody* must win, and so the mere fact that I won does not entail that this fact requires an explanation.[18] Because of cases such as these, "the improbability thesis is a non-starter" (Baras & Shenker 2020, p. 36, fn. 26). Other attempts to specify a more restricted principle governing the conditions of explanatory requirement suffer from similar problems. For instance, White (2000: 270) recognizes that the simple improbability analysis fails and claims that those facts that require an explanation are ones that "challenge our assumptions." However, it is unclear how to understand this proposal in a way that is not simply identical to the analysis in terms of improbability given our background knowledge. If facts that challenge our assumptions aren't ones that are improbable given our background beliefs, then what else could they be?

The foregoing discussion has presupposed what Nozick calls an "inegalitarian theory" of explanation (1981: 121), according to which some facts require an explanation, whereas some facts do not. But, of course, the denial of this thesis, in the form of the Principle of Sufficient Reason (PSR) has been common throughout the history of philosophy, especially among early modern continental

[18] There are counterexamples in the other direction too. God might assure me of the antecedently surprising fact that all Fs are Gs. Once He assure me of this, the fact that all Fs are Gs is extremely probable given my background knowledge. But I still might want to know *what explains* the fact that all Fs are Gs. This, let's suppose, God has not revealed to me. Thus, even though this fact is extremely probable given my background knowledge, it does not follow that this fact does not require an explanation.

rationalists, such as Leibniz and Spinoza.[19] According to the PSR, for every fact F, there is an explanation, or reason why F obtains. The PSR is an "egalitarian theory" of explanation, since it says that all facts are to be regarded as equally in need of an explanation. Although proponents of IBE rarely ever link their epistemology to the PSR—perhaps owing to the dubious reputation that the PSR has these days (Della Rocca 2010: 1— it is worth pointing out that appealing to the PSR would immediately dissolve the problem of determining when a fact requires an explanation. Indeed, in a book-length defense of the PSR, Pruss has even argued that the use of IBE presupposes the PSR (2006: 280-1). Without a prior commitment to the PSR, any instance of IBE would be subject to the objection that the purported explanatory trigger has no explanation, since the "no-explanation" hypothesis is simpler than any other explanatory hypothesis (Pruss 2006: 281).

For those, however, who would prefer to not commit IBE to so controversial a principle as the PSR, perhaps there is some restricted principle out there still waiting to be discovered, which neatly delineates those facts that require an explanation from those facts that do not. Alternatively, perhaps the best that the proponent of IBE can do here is to say that knowledge of when a fact requires an explanation is highly context-specific and ultimately a "species of practical knowledge", which one develops over time by way of successfully identifying facts that do and facts that do not require an explanation (Cabrera 2021b: 19). In some cases, we might be able to justify the claim that a fact does not require an explanation through a kind of "pessimistic induction", that is, on the basis of long-term, pervasive inability to come up with a satisfactory explanation. Being able to distinguish facts that require an explanation is a kind of virtue that is cultivated through practice, akin to Duhem's notion of "good sense"—a skill that a scientist develops over time which aids in the process of theory choice, but which is not reducible to deductive or inductive logic.[20]

5. What Counts as an Explanation?

5.1 The Plentitude Problem

Like the first step, the second step of the IBE schema also raises some difficult questions. IBE counsels us to infer the best explanation of the facts; however, what counts as an explanation has been hotly debated in the philosophy of science for the last several decades. In 1965, Carl Hempel and Gilbert Harman each put out important works relating to explanation. Whereas Hempel published

[19] Leibniz states the PSR as follows: "no fact can be real or existing and no statement true unless it has a sufficient reason why it should be thus and not otherwise" (1714, § 32/Rescher 1991: 21).

[20] See Stump (2007) and Kidd (2011) for a virtue epistemological analysis of Duhem's notion of good sense.

Aspects of Scientific Explanation—which includes his most sophisticated formulation of the account of explanation first set forth in Hempel and Oppenheim (1948)—Harman published his classic article, titled simply "The Inference to the Best Explanation." As mentioned earlier, while discussions of IBE have advanced much since then, Harman's article is the *locus classicus* for work on IBE. Despite originating around the same time though these distinct research programs in philosophy of science have developed largely in isolation from one another. Until very recently, there has not been much interaction between the two research traditions. In general, proponents of IBE shy away from saying much about what constitutes an explanation and avoid extensive discussion of how IBE might fare once combined with one or another of the various models of explanation (e.g., Psillos 2002: 606).

No doubt, this is because the existing literature on the concept of "explanation" is quite vast and ever growing, with more than a half dozen models having been defended by philosophers of science over the years. This list includes the deductive-nomological model (Hempel 1965), the causal model (Salmon 1984; Lewis 1986; Woodward 2003), and the unificationist model (Friedman 1974; Kitcher 1981; Schurz 1999), among others.[21] To make matters worse, each model of explanation has spawned a literature of its own, in which different versions of the model are articulated, and attractions and objections are debated. Although Lipton (2004: 57) in his book-length defense of IBE favors a contrastive causal model of explanation, his version of IBE does not commit itself to the causal model or any model of explanation for that matter. Furthermore, Lipton does not discuss in much detail the other prominent models of explanation nor how they might cohere with IBE, acknowledging that his treatment of the connection between IBE and the explanation debates must be brief (2004: 28).

Even though the two research programs initiated by Harman and Hempel respectively have different philosophical aims, it is plausible that there is some substantive connection between the two, since both programs concern the concept of explanation (Salmon 1989: 8). One might reasonably think that if IBE is not to be a misnomer, then surely some concept of explanation must be "plugged into" the four-step IBE schema introduced in section 2.1. On this point Lipton (2001: 100) ultimately agrees, admitting that, in the final analysis, the two research programs must come together.

However, one might reasonably object—as Salmon (2001: 68) does—that until debates about the nature of explanation are resolved, IBE can really have "no clear meaning." On Salmon's view, the fact that what counts an explanation is a matter of immense dispute is a serious problem for IBE.

[21] See Salmon (1989) for a helpful critical survey of the attempts to analyze the concept of explanation in the twentieth century.

If what it means to explain some fact is unclear, then what IBE consists in will likewise be rendered indeterminate. If it is a matter of contention what it means to explain some phenomena, then one might worry that we cannot really use IBE as a method of inference. Alternatively, one might worry that without the correct account of "explanation" in hand, we won't know whether we have applied IBE correctly. For convenience, let's refer to the worrisome problem for IBE, whereby there exists a wide array of sophisticated, rival models of explanation, about which there is much disagreement, the "plentitude problem" (Cabrera 2020a). Without a solution to the plentitude problem, one might be inclined to look elsewhere for a philosophical model of scientific reasoning. So, it is imperative that proponents be able to give some response to this problem.

5.2 The Primitivist Response

One response to the plentitude problem is to the claim that we ought to treat the instances of "explain" and "explanation" in the IBE schema as primitive terms, at least until debates about the nature of explanation are settled. This "primitivist" response is similar to the position that Lipton (2001) articulates in his response to Salmon's challenge that IBE lacks a clear meaning, in light of the vigorous and long-standing explanation debates. According to Lipton (2001: 100)

> [w]hether or not explanatory considerations are a guide to inference does not depend on whether we have an adequate account of explanation, any more than our use of a grammar to understand our language depends on our ability to give an adequate explicit account of the structure of that grammar.

To be sure, a fully satisfactory theory of IBE would also include a substantive accounnt of the nature of explanation—a point which Lipton (2001: 100) admits. However, on Lipton's view, the proponent of IBE need not yet settle on a particular model of explanation. It is legitimate to carry on with the project of articulating a philosophical account of IBE without an account of explanation.

What lends plausibility to the primitivist response is that the terms "explain" and "explanation" are not technical terms, but rather belong to ordinary language. Thus, as Lipton suggests above, the concept should be familiar enough for us to be able to understand and apply IBE, even in the absence of a carefully worked-out account of explanation. As further evidence for Lipton's main premise, one might appeal to research in developmental psychology (e.g., Gopnik 1998), which suggests that explanation-seeking and -construction habits emerge at an early age, and thus prove to be an indispensable component of human cognition. In addition, the primitivist might exploit the fact that every philosophical account leaves certain concepts or terms unanalyzed. Thus, one might attempt

to construct a "companions-in-guilt" argument, according to which it is unfair to dismiss IBE as a failure for doing what every philosophical theory must, on pain of infinite regress, do at some point.

One problem with leaving the concept of "explanation" a primitive is that the prominent models of explanation defended by philosophers in the last several decades disagree drastically on the extension of the concept. So, those who use IBE while having in mind different concepts of explanation in mind will end up endorsing radically different methods of inference. To illustrate this difficulty, we can consider three prominent models of explanation: the deductive-nomological model, the causal model, and the unificationist model. According to the deductive-nomological (DN) model, an explanation is a sound deductive argument, one of whose premises makes essential reference to a law of nature. According to the causal model, a successful explanation cites information about the causal history of the fact that is to be explained. Finally, according to the unificationist model, an explanation unifies the fact to be explained with the rest of one's body of knowledge.[22]

Given these different concepts of explanation, consider now three different versions of IBE: "Inference to the Best Deductive-Nomological Explanation", or "IBE_{DN}" which employs the DN model as its preferred concept of explanation, "Inference to the Best Causal Explanation", or "IBE_{C}", which makes use of some version of the causal model of explanation, and finally "Inference to the Best Unificatory Explanation", or "IBE_{U}", which relies on some version of the unificationist model of explanation. The proponent of IBE who wishes to not take a firm stand on the question of explanation might claim that all three of these versions of IBE are equivalent, and so it does not matter which version of IBE we employ.[23] However, it clearly does matter. Since each of the three models of explanation places different constraints on what counts as an explanation, this entails limitations on what sorts of conclusions on may draw, *in principle*, using the corresponding version of IBE.

To establish this point, it will suffice to examine the DN version of IBE. Since the *explanans* of a DN explanation (i.e., that which does the explaining) is a conjunction of laws of nature and antecedent conditions, this restricts which sorts of hypotheses may be drawn by using IBE_{DN}. For example, one may not use IBE_{DN} to infer some singular causal or existence claim. However, often the kinds of examples put forward as paradigmatic applications of IBE have conclusions that are singular

[22] In Kitcher's unificationist account, an explanation consists in a derivation of the phenomenon from an argument pattern that is part of the best deductive systematization of our knowledge, where the best deductive system is that which contains arguments which 'instantiate a few, stringent patterns' (1981: 520).

[23] Cabrera (2020a) calls this response (which can be combined with primitivism) "accomodationism", according to which, one may simply plug in one's "favorite" account of explanation into the IBE schema.

causal or existence claims. Recall, proponents of IBE typically claim that the following are all good applications of explanation reasoning: i) Darwin infers that all organisms on Earth are descended from a common ancestor, ii) astronomers infer the existence of Neptune to explain the deviant orbit of Uranus, and iii) Semmelweis infers that cadaverous matter explains childbed fever in maternity wards. These conclusions do not look anything at all like the *explanans* of a DN explanation; they do not make essential reference to some law of nature. Thus, these conclusions will not count as explanations according to the DN model. Consequently, quite a large number of inferences frequently regarded as excellent instances of IBE in action will not be licensed by IBE_{DN}. Similar results follow for the other versions of IBE.[24] For instance, IBEC won't work for non-causal explanations. The upshot is that it matters greatly which model of explanation is plugged the IBE schema. Results are not equivalent when different models are combined with IBE since, the corresponding versions of IBE do not permit the same sorts of explanatory inferences to be drawn.

5.3 The Pluralist Response

Another response to the plentitude problem that one might pursue is that IBE should be viewed as a family of distinct, though related inference methods, all of which are united in their commitment to the evidential relevance of some concept of explanation. On this "pluralist" view, IBE_{DN}, IBE_{C}, IBE_{U}, and any other version of IBE that we might construct by plugging some plausible model of explanation into the IBE schema, are not to be regarded as competitors. Instead, we should think of them all as compatible. Thus, it is not a problem that IBE_{DN} will not permit inferences to singular casual or existence claims. Luckily, we have another version of IBE available, namely IBE_{C} that will allow us to draw such conclusions. According to Psillos (2002: 606), who adopts something akin to the pluralist view, IBE "is not usefully seen as a species of ampliative reasoning, but rather as a genus whose several species are distinguished by plugging assorted conceptions of explanation in the reasoning schema." Similarly, according to the "reductive pluralism" of Khalifa et al. (2017), IBE is not a fundamental form of inference, but instead reduces to a plurality of specific inference methods, each of which cites some "thick relation" (causal, nomological, mechanistic, etc.), along with some set of explanatory virtues that are peculiar to that thick relation.

A pluralist account of IBE gains plausibility from "long-budding consensus in the philosophy of science", according to which "scientific explanations come in many shapes and sizes" (Khalifa et

[24] See Ruetlinger (2017) for a discussion of non-causal explanations.

al. 2017: 82). Already at the end of his historical survey of the debates over the concept of explanation in the 20th century, Salmon (1989) suggested that pluralism about the concept of explanation may be necessary in order to make sense of the prevalence of explanations that appeal to *unification*, alongside those that appeal to *causation*. To illustrate the possibility of a rapprochement between the causal model and the unificationist model, Salmon (1989: 183) provides the following helpful example:

> Why did the balloon move toward the front of the cabin? Two explanations can be offered, both of which are correct. First, one can tell a story about the behavior of the molecules that made up the air in the cabin, explaining how the rear wall collided with nearby molecules when it began its forward motion, thus creating a pressure gradient from back to front of the cabin. This pressure gradient imposed an unbalanced force on the back side of the balloon, causing it to move forward with respect to the walls of the cabin. Second, one can cite an extremely general physical principle, Einstein's principle of equivalence, according to which an acceleration is physically equivalent to a gravitational field. Since helium-filled balloons tend to rise in the atmosphere in the earth's gravitational field, they will move forward when the airplane accelerates, reacting just as they would if a gravitational field were suddenly placed behind the rear wall.

On the basis of this example, Salmon considers the possibility that there are at least two distinct senses of explanation—explanation$_1$ in terms of causal-mechanical principles and explanation$_2$ in terms of theoretical unification. Perhaps then our concept of explanation is complex and multi-faceted, and so different kinds of explanation serve to illuminate various aspects of the same phenomenon. If there are multiple distinct, yet compatible notions of explanation, then, plausibly, one might think there ought to be multiple distinct, yet compatible notions of *inference* to the best explanation.

While the pluralist view of IBE is a promising response to the plentitude problem, there are some difficulties to which pluralism must attend. First, we cannot simply assume that IBE will be compatible with every concept of explanation that philosophers have articulated. Rather, the proponent of IBE will need to consider in greater detail what the inference method looks like upon plugging in some account of explanation into the IBE schema. This investigation should include some account of the criteria that ought to be used to determine the *best explanation,* which presumably will differ depending on the operative notion of explanation. Furthermore, it is certainly possible that upon conducting this more detailed investigation, we will discover that some concepts of explanation are not consistent with IBE. Recently, Prasetya (2021) has argued that, while some models of explanation are compatible with IBE, e.g., Salmon's (1984) causal-mechanical model and Kitcher's (1981) unificationist account, others such as Railton's (1978) deductive-nomological-probabilistic model, and van Fraassen's (1980) erotetic account are incompatible with IBE. A further problem for the pluralist view arises once we consider the possibility of applying different versions of IBE to the same fact. In

such cases, it is not hard to imagine that different versions of IBE will deliver conflicting verdicts about whether we ought to infer some hypothesis (Cabrera 2020a).[25]

5.4 Does IBE Require a "Model" of Explanation?

It is significant that both opponents and proponents of IBE assume that some account of explanation needs to be "plugged into" the four-step IBE argument schema. It is possible, however, to question this assumption. For instance, Cabrera (2020a) argues that IBE does not require a "model" of explanation. On this view, the primary reason that IBE does not need a model of explanation is that being categorizable as an explanation according to the correct philosophical account of explanation does not appear to do much, if any, justificatory work in the context of IBE. Rather, what seems to justify a conclusion in any successful application of IBE is precisely that the hypothesis is judged to be the *best* explanation. While IBE is sometimes criticized for not specifying what being the "best explanation" amounts to, proponents of IBE have not been silent on how competing explanations should be ranked. As mentioned in section 2.1, normally, proponents of IBE put forward an independent list of "explanatory virtues", e.g., *simplicity*, *unification*, *scope*, etc., which are supposed to help us compare different explanatory hypotheses.

Additionally, it is crucial to recognize that the explanatory virtues are properties that can be exhibited by hypotheses or theories of many different types: those that are causal or non-causal, those that logically entail or merely confer some statistical probability on the phenomena, those concerned with particular events or with nomic generalizations, and even theories that are philosophical rather than scientific. For example, consider that in physics, Newton's theory of gravitation was praised for its unification of disparate sets of physical facts, including facts about terrestrial and celestial motion, (e.g., Whewell 1840/1968: 153). By contrast, in philosophy of science, Armstrong (1983: 105) employs IBE to defend a conception of laws of nature as necessitation relations between universals by appealing to, among other things, the unifying power of this account of lawhood. What these two applications of IBE have most in common is their appeal to the virtue of *unification*.

Consequently, it is unclear whether a hypothesis even needs to satisfy the conditions of one the extent conceptions of explanation in order to be explanatorily virtuous. What appears to unite all of the many examples of explanatory inference, from science, philosophy, and ordinary life, is not

[25] See Climenhaga (2017a) who discusses a similar problem that arises once we consider different levels explanations. IBE, if formulated as an inference method for full beliefs (as opposed to degrees of belief), "licenses deductively inconsistent inferences from the same evidence" (251).

some or another conception of explanation, but rather a set of widely applicable explanatory virtues. This striking commonality suggests that, at the very least, what does most of the justificatory work in any particularly successful application of IBE is just that the hypothesis H does exceedingly well with respect to the explanatory virtues relative to some evidence E.[26] Hence, whether a hypothesis satisfies the conditions of any given model of explanation may be, epistemically, beside the point.

6. Do the Explanatory Virtues Track the Truth?

In light of their central importance in explanatory reasoning, it is necessary now to direct our discussion toward examining the nature and justification of the explanatory virtues. Proponents of IBE differ with respect to which items belong on the correct list of explanatory virtues, although as we will see there is much overlap in existing lists, and at times apparent differences are revealed to be merely terminological. What all proponents of IBE agree on, however, is that there is a plurality of explanatory virtues that are used to evaluate competing hypotheses. Accordingly, a hypothesis H_1 is a better explanation than its competitors H_2, H_3,...,H_n insofar as H_1 manifests these virtues to a higher degree, overall, than H_2, H_3,...H_n.

An early source for understanding theory choice in terms of a list of criteria can be found in the writings of Thomas Kuhn. Despite his not explicitly advocating IBE, it is striking how much contemporary formulations of IBE have in common with the account of theory choice laid out by Kuhn in the 1969 post-script to *The Structure of Scientific Revolutions* (1962) and in later works (e.g., Kuhn 1977).[27] Distancing himself from the radical relativism that enthusiastic postmodernists had attributed to him, Kuhn insisted that genuine, scientific progress occurs. For Kuhn, it is possible to compare the merits of competing scientific paradigms by reference to theoretical virtues such as: *accuracy*, *consistency*, *scope*, *simplicity*, and *fruitfulness*. There is no effectively decidable algorithm, of course, for determining how these different explanatory considerations ought to be applied in a given case, and certainly, different scientists will weight and apply such standards differently. Still, for Kuhn, these factors constitute "standard criteria for evaluating the adequacy of a theory" and thus "…provide a shared basis for theory choice" (1977: 321-22).

[26] See Iranzo (2001) who also highlights the crucial importance of the explanatory virtues in a response to van Fraassen's (1980) bad lot objection. According to Iranzo (2001: 95), "appealing to criteria of explanatory goodness makes some progress against the argument from the bad lot."

[27] See Walker (2012) for a Kuhnian account and defense of IBE.

Another early source for understanding theory choice in terms of a list of different virtues is Quine and Ullian (1978). As in the case of Kuhn, Quine and Ullian do not explicitly defend IBE, although they have something similar to IBE in mind[28] when discussing "five virtues that a hypothesis may enjoy" (1978: 66). In particular, according to Quine and Ullian, these virtues are:

i) *conservatism*, which consists in preferring hypotheses that do not conflict with already held beliefs

ii) *modesty*, which consists in preferring hypotheses that are logically weak, or ones that are of a "more usual or familiar sort" (1978: 68)

iii) *simplicity*, which when dealing with mathematical equations, corresponds to the degree, or the order of a mathematical equation (e.g., Jeffreys 1931), but sometimes is "not to be distinguished from modesty" and other times has to do with providing a "brief unified story" (1978: 71)

iv) *generality*, which consists in preferring hypotheses that apply to a wide range of phenomena

v) *refutability*, which refers to the familiar Popperian requirement that a hypothesis be falsifiable.

Quine and Ullian recognize that many of these virtues are related, and in some cases, such as between modesty and conservativism, "there is no call to draw a sharp line" (1978: 69). Presumably, the best hypothesis is the one that does best, overall, with respect to all of these criteria.

When we turn to the explanatory virtues that have been put forward by explicit proponents of IBE, we will find similar criteria. For instance, Thagard (1978) articulates three explanatory virtues: consilience, simplicity, and analogy. A theory is *consilient* to the extent that it explains the facts, and most important, to the extent that it unifies diverse domains of phenomena. *Simplicity* is determined by to the size of the set of auxiliary assumptions needed by a hypothesis to explain the data, and perhaps the number of entities posited by the theory. *Analogy* corresponds to a preference for familiar, well-established causal mechanisms and processes. Similarly, Psillos (2002: 614-6) discusses the virtues of consilience, completeness, importance, parsimony, unification, and precision. For Psillos, *consilience* simply refers to fit with background knowledge. *Completeness* consists in explaining all the relevant phenomena. *Importance* corresponds to the virtue of explaining "salient" phenomena. *Parsimony*, is more

[28] Specifically, Quine and Ullian (1978: 66) have in mind *explanatory* hypotheses, writing "People adopt or entertain a hypothesis because it would explain, if it were true, some things that they already believe."

or less identical to Thagard's notion of simplicity, having to do with the size of the set of auxiliary assumptions needed by a theory to explain the data. *Unification* consists in bringing together diverse bodies of data under one explanatory hypothesis. *Precision* refers to the virtue of stating a precise causal-nomological mechanism to explain the data. Finally, Lipton cites many of the same virtues just reviewed, including "mechanism, precision, scope, simplicity, fertility or fruitfulness, and fit with background belief" (2004: 122), but does not discuss them with much depth.

As I suggested in section 5.4, the explanatory virtues play an indispensable role in any given application of IBE, and so if IBE is to be an inductively valid form of inference, then we need to be sure that the explanatory virtues are appropriately connected to the truth. This is no doubt the most important epistemological question that IBE faces. Lipton (2004: 144) refers to this general problem as *Voltaire's objection*[29]: "Why should the explanation that would provide the most understanding if it were true be the explanation that is most likely to be true? Why should we live in the loveliest of all possible worlds?" Addressing this objection is especially pressing, given that the truth-conduciveness of some of the virtues that proponents of IBE cite has long been doubted. For example, van Fraassen regards the criteria that feature on lists of explanatory virtues, including mathematical elegance, simplicity, scope, unification, and explanatory power as *non-epistemic*: "Values of this sort…provide reasons for using a theory, or contemplating it, whether or not we think it true, and cannot rationally guide our epistemic attitudes and decisions" (1980: 87). For van Fraassen, the epistemic virtues are just "consistency, empirical adequacy, and empirical strength", whereas the other virtues mentioned above are "*pragmatic* virtues…they provide reasons to prefer the theory independently of questions of truth (1980: 88). Relatedly, perhaps the preference for some of the explanatory virtues in science is an *aesthetic* one that is not necessarily linked to truth (Elgin 2020: 23).[30]

One promising strategy is that the truth-conduciveness of the explanatory virtues is to be justified of empirical grounds. This approach has already been taken up by Psillos (1999) in his defense of scientific realism. According to Psillos (1999: 77), a rule of inference is justified in virtue of producing, on balance, more truths than falsehoods. Psillos (1999: 81) thus assumes epistemological externalism in general and reliabilism in particular, (e.g., Goldman 1986) in his defense of IBE. On

[29] This title is in reference to Voltaire's play *Candide*, which skewers the Leibnizian view that we live in the best of all possible worlds.

[30] Some have also contested traditional virtues such as *simplicity* on the grounds that they are "laden with socio-political values" (Longino 1996: 52), and thus crowd out other virtues such as *ontological heterogeneity* and *novelty*, which are more important in certain contexts.

this package of views, one does not need to be able to produce some rational demonstration that an inference rule is reliable in order to be justified in using that inference rule. Such a demand needs to be met only if one accepts epistemological internalism, which requires that the rational grounds for the agent's beliefs be, in some sense, accessible to the agent. According to externalist views though, one does not need to understand how the "engine" of inference works in order to be justified in relying upon it. All that needs to be the case is that the historical track record of the explanatory virtues is a good one. That is, if relying on the explanatory virtues has historically been successful in leading us to the truth, then—supposing that the reliability of the inference method is all that is required for its justification—we would then be justified in relying on the explanatory virtues while using IBE.

Now, one might object, as Lycan (2002: 421) does, that this empirical defense is doomed to fail because it presupposes that we can identify which hypotheses are true independently of our appeals to explanatory virtues. But, of course, we cannot do this. Consider the virtue of simplicity. It is not the case that we have access to all the true hypotheses, and from there we can check to see that a large proportion of them have the virtue of simplicity. On the contrary, the proponent of IBE says we ought to believe that our hypotheses are true *precisely* because of their simplicity. There is a response to this objection that the proponent of IBE can give, however. While we cannot examine all the true theories from God's point of view and then determine which proportion are simple, we can investigate those theories which have proved predictively successful, or which have resisted disconfirmation, and see which proportion of *those* have had the virtue of simplicity. If we have reason to believe that predictively successful theories are true, and if it turns out that these theories quite frequently manifest the explanatory virtues, then the empirical defense of IBE can at least get off the ground.

Nevertheless, one might still worry that this empirical defense of the explanatory virtues will not survive a detailed and honest assessment of the historical record. As an illustration, consider the case of Newtonian mechanics. As Lipton (2004: 60) admits, despite being supplanted in the 20th century, "Newtonian mechanics is one of the loveliest explanations in science." The 19th century historian and philosopher of science William Whewell extolled some of the key virtues of Newtonian mechanics, such as its ability to unify many different kinds of facts, arguing that these are features that any good scientific theory ought to have (1840/1968: 153). When a theory united in a common explanatory framework diverse classes of phenomena, which before had appeared to be disconnected, Whewell called this a "consilience of inductions." For Whewell, we should think that consilience, or unification, is truth-conductive because the historical record "offers no example in which a theory supported by such consiliences, had been afterwards proved to be false" (1840/1968: 295). Such was

the supposed triumph of classical physics that at the end of the 19th century the physicist Albert Michelson (of the famed "Michelson-Morley experiment") remarked:

> [I]t seems probable that most of the grand underlying principles have been firmly established and that further advances are to be sought chiefly in the rigorous application of these principles to all the phenomena which come under our notice...the future truths of physical science are to be looked for in the sixth place of decimals (1896: 159)

Unfortunately, the 20th century witnessed the overthrow of Newton's theory of gravity in the form of Einstein's and the replacement of Newtonian mechanics with quantum mechanics.

Of course, proponents of IBE do not want to say that the explanatory virtues are a guarantor of truth; still, even those who would put forth the modest claim that the historical record more or less points toward the truth-conduciveness of the explanatory virtues may find this argument difficult to make. As Sober (2015: 46-7) points out, this sort of the track-record defense seems plausible only because of our selective attention or confirmation bias. When trying to defend the explanatory virtues, we think of those grand, unifying theories such as general relativity that turned out to be (more or less) successful. However, we often fail to consider the many dozens of theories that might have exceled with respect to the explanatory virtues, but in the end were abandoned for being not so successful. [31] Plausibly, the reason that we do not consider these theories is precisely that they were abandoned. However, for the empirical track-record argument to be legitimate, we must pay attention not only to the explanatory virtuous and successful theories, but also to the explanatorily virtuous and unsuccessful theories. Perhaps the empirical record would vindicate the explanatory virtues; however, to my knowledge no such detailed investigation of this kind has been yet undertaken.[32]

In light of the difficulties encountered while trying to defend an optimistic meta-inductive argument on behalf of the explanatory virtues, one might adopt the view that the justification for some of the explanatory virtues is *a priori*. For example, in an extended defense of a version explanatory coherentism of a broadly Quinean sort, Poston (2014: 83-4) ultimately concludes that certain explanatory virtues, such as simplicity, are *intrinsically* epistemically relevant. Similarly, one might judge

---

[31] Ivanova (2020: 95) makes a similar point in a discussion of "aesthetically pleasing" theories, (e.g., simple, elegant, unifying, etc.), many of which are now regarded as false, or have been subsequently abandoned.

[32] However, see Bird (2020) for an empirical defense of IBE that appeals, not to a *quantitative* conception of past success, but to the Kuhnian notion of an *exemplar*. Perhaps this way of defending IBE can bypass the shortcomings of the standard track-record defense. As Bird (2020: 64) writes: "if the exemplars of a field are true or highly truthlike, then the standards of explanatory goodness they generate will be truth-conducive. And since it would not be an implausible coincidence that such exemplars are true, then it does not require an implausible coincidence that true theories have a high degree of explanatory goodness."

it mistaken to try to give an empirical argument on behalf of simplicity if "it is an ultimate a priori epistemic principle that simplicity is evidence of truth" (Swinburne 1997: 1). This rationalist view fits with the claim sometimes put forth that IBE is an absolutely basic, rational non-deductive belief-forming method (e.g., Enoch and Schechter 2009). Presumably, if this rationalist view is right, it is not necessary, nor possible, to give an a *posteriori* defense of simplicity, and so one need not worry about the fact that it is unclear whether an induction over the historical record can support simplicity or the other explanatory virtues. Perhaps the best that one can do here to defend the epistemic relevance of simplicity is to cite, as Sider (2013: 239) does, the "intuition" that simplicity tracks truth. Of course, such a defense will not prove convincing to those who are already antecedently skeptical of IBE, or skeptical of the truth-conduciveness of the explanatory virtues more specifically.

7. New Directions: IBE and Bayesianism

Much of the latest work on the nature and potential justification of IBE has taken place in dialogue with formal epistemology, making heavy use of mathematical tools such as the probability calculus and Bayes' Theorem.[33] For example, of the sixteen essays included in a recent edited volume (McCain and Poston 2017b), eight of those contain substantive discussion of the relationship between IBE and Bayesianism, such as whether the two accounts of scientific reasoning are compatible, whether the one can be justified in terms of the other, etc. This formal work on the nature of IBE, is in my view, a welcome development. It is hard to contest Douven's (2017: 9) contention that traditional accounts of IBE have "none of the precision—not even remotely of its current main contender", i.e., Bayesianism. Since the Bayesian framework is a well-studied, powerful, and precise formal framework, investigating IBE from the perspective of Bayesianism may help us get clear about the nature and epistemic status of IBE. If it could be shown that IBE is not only compatible, but in some sense can be vindicated from within the Bayesian framework, then IBE would inherit whatever plausibility Bayesianism possesses. So too, perhaps an appeal to explanatory considerations would help Bayesianism deal with some of the problems raised by its critics (e.g., Huemer 2009).[34]

[33] See: Okasha (2000), McGrew (2003), Lipton (2004, Ch. 7), Psillos (2004), Weisberg (2009), Henderson (2014), Glass (2012), Iranzo (2016), Cabrera (2017), Dellsén (2018), Douven and Wenmackers (2017), Bird (2017), Climenhaga (2017a), Schupbach (2017), and Blanchard (2018).

[34] See Earman (1992) for a book-length examination of Bayesianism and Sober (2002) for a shorter discussion of the scope and limitations of Bayesianism.

According to Bayesianism, rational agents can be modelled as having "degrees of belief" or "credences" with respect to propositions, a notion which corresponds intuitively to the confidence that the agent has that a proposition is true. For an agent to be rational, her "credence function", or "personal probability function" (Pr-function) must obey the standard axioms of the probability calculus, and moreover, the agent must update her degrees of belief over time in light of new evidence by the following rule:

> *Conditionalization*: For any hypothesis H and any body of evidence E, upon learning that E is true, the agent's new probability in H, i.e., $pr_{new}(H)$ should be set equal to the agent's old probability in H conditional on E, i.e., $pr_{old}(H|E)$

Notably, van Fraassen (1989: 160-170) argued that Bayesianism is incompatible with IBE precisely because IBE is inconsistent with conditionalization. The reason that IBE is inconsistent with conditionalization is that, according to van Fraassen, IBE requires us to add post-conditionalization "bonus probabilities to certain hypotheses" (1989: 166) in light of their explanatory goodness. Clearly, this a proposal that is flatly inconsistent with the core tenets of Bayesianism.[35] However, most proponents of combining IBE with Bayesianism who have responded to van Fraassen's argument (e.g., Okasha 2000; Lipton 2004; Weisberg 2009) simply reject the ad hoc way in which van Fraassen merges IBE with Bayesianism[36], opting for other, more plausible ways of merging the two frameworks.

For example, in chapter 7 of his book *Inference to the Best Explanation*, Lipton suggests that IBE and Bayesianism are compatible because the explanatory virtues can be used to determine the values of the probabilities that are needed to transact a Bayesian inference. Crucially, Lipton (2004: 115) says the following about the relationship between the explanatory virtues and Bayesianism: "Explanatory considerations may also enter into the determination of priors in other ways. This is where considerations of unification, simplicity and their ilk would naturally come into play." In addition, in a later reply to critics, Lipton (2007), expresses further support for this general strategy of combining

[35] More specifically, if the agent adds post-conditionalization bonus points to explanatorily virtuous hypotheses, then the agent will be subject to a diachronic "Dutch book." That is to say, there will exist a series of gambling bets, each of which the agent, by her own lights, will individually sanction as fair, but which together logically guarantee a loss. Having a Pr-function that is susceptible to a Dutch book is supposed to indicate that the Pr-function is rationally defective. Thus, according to van Fraassen, IBE ought to be rejected. See Douven (1999) for a direct response to van Fraassen's Dutch book argument.

[36] Importantly, Douven and Wenmackers (2017) adopt van Fraassen's general approach to combining IBE and the Bayesian rule of conditionalization but argue that this non-Bayesian version of IBE is superior to standard Bayesianism. So too, Douven and Schupbach (2015: 6) cite some empirical research to argue that the aforementioned non-Bayesian form of "probabilistic abduction" "is descriptively superior to Bayesianism: by taking explanatory considerations into account, next to conditional probabilities, we arrive at more accurate predictions of people's updates than we would on the basis of the objective conditional probabilities alone."

the two frameworks, favoring the idea that "explanatory considerations are a guide to the assignment of probabilities", and resting the reconciliation of IBE and Bayesianism on "the hope that there might be a normatively privileged way of assigning priors and likelihoods" (458), once again referencing the proposal to use factors such as simplicity to determine probability values. Similarly, Weisberg (2009) suggests that virtues like *simplicity* or *stability* (a notion adapted from White 2005) should be used to constrain the prior probabilities on the objectively correct probability function.

While Lipton paints an attractive compatibilist picture here, much work needs to be done in order to carry it to fruition. Recall that in the previous two sections, we saw that the explanatory virtues play an indispensable role in any application of IBE, and what's more, that there are many different explanatory virtues. These two features of IBE suggest that Hitchcock (2007: 439) is correct in remarking that a full reconciliation of 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." One difficulty with the empirical track record defense that we saw in section 6 is that it is something of a blunt instrument. A more fruitful approach for the proponent of IBE is to consider whether each of the individual explanatory virtues is something that, other things being equal, increases the probability of a given hypothesis. The problem that we witnessed in the previous section, namely the difficulty of showing that the explanatory virtues are truth-conducive, may be better tackled by considering the explanatory virtues in the light of Bayesianism.

Indeed, and once we do, it is fairly straightforward that at least some of the explanatory virtues increase the probability of a hypothesis (Cabrera 2017: 1252-54). Consider, for instance, *conservatism* or *fit with background knowledge*. If a hypothesis H_1 more *conservative* than a rival hypothesis H_2, then one way of making sense of this claim is just that the prior probability of H_1 is higher than that of H_2 given one's background knowledge K, i.e., $pr(H_1|K)>pr(H_2|K)$. For example, the hypothesis that my wife drank the beer missing from the refrigerator is more *conservative* or *fits better with our background knowledge* than the rival hypothesis that aliens visited my house and stole the missing beer. Intuitively, hypotheses that do not fit well with background knowledge are just ones that are improbable given our background knowledge. Thus, to the extent that H_1 is more conservative or better fits with our background theories than H_2, the prior probability of H_1 will be higher than that of H_2. This feature, of course, makes the virtue of *conservativism*, clearly truth-conducive.

So too, recall that one notion of *simplicity* defended by proponents of IBE relates to the number of auxiliary assumptions required by a theory to logically entail some predicted observation (e.g.,

Thagard 1978; Psillos 2002).[37] Other things being equal, this notion of simplicity increases the probability of a theory. Take two rival theories T_1 and T_2. Suppose $pr(T_1)=pr(T_2)$ and both T_1 and T_2 entail some evidence E, but in each case only with the addition of some set of auxiliary assumptions. Now, if the auxiliary assumptions A_1 required by T_1 are a subset of the auxiliary assumptions A_2 required by T_2, then certainly, $pr(T_1\&A_1) \geq pr(T_2\&A_2)$. This result just follows from a theorem of the probability calculus, according to which the logically weaker hypothesis, necessarily, has a probability greater than or equal to the logically stronger hypothesis, e.g., $pr(A) \geq pr(A\&B)$.[38] Of course, this situation is seldom the case in practice. Normally, A_2 and A_1 are only partly over-lapping sets, and are not comparable in this way.[39] Still though, suppose that the auxiliary assumptions that T_2 requires are ones that have no good independent reasons to believe. On the other hand, suppose T_1 can predict E with a more plausible set of auxiliaries. By "plausible", I mean that more of A_1 can be independently verified, and is not postulated for the sole purpose that T_1 can predict E. Perhaps more of the set A_1 belongs to the background knowledge K than does A_2, or more of A_1 is directly observable than A_2. In any case, if A_1 is much more plausible than A_2, and if this is an adequate explication of the notion of "simplicity", then certainly simplicity will be a truth-conducive explanatory virtue.

Here, I have only scratched the surface of how *one* approach to investigating the truth-conduciveness of the explanatory virtues might proceed. Interestingly, Salmon (1990), despite being a critic of IBE, offers a kind of Bayesian version of the 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." Simple and unifying hypotheses succeed more frequently, according to Salmon, and so we are justified in regarding the prior probability of simple and unified hypotheses as higher. Proponents of IBE might also appeal to existing Bayesian accounts of explanatory virtues, such as unification (e.g., Myrvold 2017) or simplicity (e.g., Dowe, Gardner, and Oppy 2007), which have been articulated independently of attempts to understand and defend IBE.[40] Additionally, a number of philosophers have articulated and defended different

[37] This account of simplicity is similar to Quine and Ullian's (1978) notion of *modesty*.

[38] Sober (2015: 71) calls this application of simplicity, the "razor of silence", which should be distinguished from a different principle where this rationale does not apply, namely the "razor of denial."

[39] It should be noted that Thagard himself is aware that the question of simplicity cannot be neatly quantitative in this way but rather, "a qualitative comparison, application by application, must be made" (1978: 87).

[40] Proponents of IBE who are not motivated to combine IBE with Bayesianism may also avail themselves of non-Bayesian defenses of simplicity (e.g., Forster and Sober 1994).

probabilistic measures of "explanatory power" (e.g., Crupi and Tentori 2012; Schupbach and Sprenger 2011). While not all those who propose formal measures of explanatory power attempt to defend versions of IBE, some, such as Schupbach (2017) have done just that. If it can be shown that, all else being equal, doing well with respect to some formal measure of explanatory power makes a hypothesis more probable, then this may go some way toward solving the problem that the explanatory virtues are not clearly truth-conducive. On the other hand, given a piecemeal approach to this question, it is possible that some of the explanatory virtues won't end up being truth-conducive[41], in which case the rationality of employing IBE will depend very much on which explanatory virtues are being invoked.

8. Conclusion

Although Bayesianism is currently the "dominant" account of scientific reasoning (Douven 2021)—among philosophers of science at least— it is hard to shake the feeling that there is *something* right about IBE. It was not too long ago when IBE figured much more prominently in philosophical discussions of scientific method and other related problems in the philosophy of science. It has proved a common maneuver to appeal to explanatory virtues, such as simplicity or unification, in order to deal with the skeptical problem posed by the underdetermination problem (e.g., Kukla 1994). Furthermore, traditionally, proponents of scientific realism who endorse the so-called "No-Miracles Argument" formulate their defense of realism as an application of inference to the best explanation, (e.g., Putnam 1975; Boyd 1983), and recent work in philosophy of science arguing for realism still makes heavy use of the explanatory virtues (e.g., Schindler 2018). As I suggested in section 7, the application of formal tools to study of explanatory reasoning will likely better to serve to clarify and justify the core commitment of IBE, namely that explanation is a guide to inference. While much more work on the nature of IBE remains to be done, especially with respect to clarifying and rationally justifying the various explanatory virtues, IBE nonetheless remains an important philosophical model of scientific reasoning worthy of critical engagement.

[41] For example, Cabrera (2017: 1254) argues that some explanatory virtues, such as *precision*, *scope*, and *mechanism*, increase the informative content of a hypothesis, a fact which pulls against the probability that the hypothesis is true. Additionally, Roche (2018) argues that various formulations of parsimony cannot be justified as truth-conducive virtues. See also Roche and Sober (2013) for a more general argument that "explanatoriness is evidentially irrelevant"; however, see McCain and Poston (2017c), Climenhaga (2017b), Lange (2017) and Cabrera (2020b) for some responses to Roche and Sober's argument.

References

Achinstein, P. (2001). *The Book of Evidence*. New York: Oxford University Press.

Aliseda, A. (2006). *Abductive Reasoning: Logical Investigation into the Processes of Discovery and Evaluation*. Amsterdam: Springer

Aliseda, A. (2007). "Abductive Reasoning: Challenges Ahead," Theoria, 60: 261-270.

Armstrong, D. (1983). *What is a law of nature?*. Cambridge University Press.

Baker, A. (2009). "Mathematical Explanation in Science," *Brit. J. Phil. Sci.*, 60: 611-633.

Baras, D. (2019). "Why do certain states of affairs call out for explanation? A critique of two Horwichian accounts," *Philosophia*, 47(5), 1405–1419.

Baras, D. (2020). "A Strike against a Striking Principle," *Philosophical Studies*, 177(6): 1501–1514.

Baras, D., & Shenker, O. (2020). "Calling for explanation: The case of the thermodynamic past state. *European Journal for Philosophy of Science*, 10(3), 1–20.

Barnes, E. (1995). "Inference to the loveliest explanation", *Synthese,* 103(2): 251–277.

Biggs, S. & Wilson, J. (2017). "The A Priority of Abduction," *Philosophical Studies*, 174(3): 735–58.

Bird, A. (2017). "Inference to the Best Explanation, Bayesianism, and Knowledge", in K. McCain and T. Poston (eds.), *Best Explanations: New Essays on Inference to the Best Explanation*. Oxford: Oxford University Press, pp. 97-120.

Bird, A. (2020). "Scientific Realism and Three Problems for Inference to the Best Explanation," in W. Gonzalez (ed.), *New Approaches to Scientific Realism*. De Gruyter, pp. 48-67.

Blanchard, T. (2018). "Bayesianism and Explanatory Unification: A Compatibilist Account.," *Philosophy of Science*, 85(4): 682-703.

Bohr, N. (1922). "Nobel Lecture: The Structure of the Atom", The Nobel Foundation. Available online at http://www.nobelprize.org/nobel_prizes/physics/laureates/1922/bohr-lecture.pdf, 1-33.

Boyd, R. (1983). "On the Current Status of the Issue of Scientific Realism," *Erkenntnis*,19(3): 45-90.

Cabrera, F. (2017). "Can there be a Bayesian Explanationism?: On the Prospects of a Productive Partnership," *Synthese*, 194(4): 1245–72.

Cabrera, F. (2020a). "Does IBE Require a 'Model' of Explanation?" *British Journal for the Philosophy of Science*, 71(2): 727-750.

Cabrera, F. (2020b). "Evidence and explanation in Cicero's On Divination*," Studies in History and Philosophy of Science Part A,* 82:34-43.

Cabrera, F. (2021a). "String theory, non-empirical theory assessment, and the context of pursuit," 198: 3671-3699.

Cabrera, F. (2021b). "Is epistemic anxiety an intellectual virtue? *Synthese*, 199: 13471–13495.

Campanaro, D.M. (2021). "Inference to the Best Explanation (IBE) and Archaeology: Old Tool, New Model," *European Journal of Archaeology*. 1-21. doi: 10.1017/eaa.2021.6.

Campos, D. (2009). "On the Distinction between Peirce's Abduction and Lipton's Inference to the Best Explanation," *Synthese*, 180 (3): 419-442.

Chen, E. (1979). Twins Reared Apart: A Living Lab. *New York Times*. https://www.nytimes.com/1979/12/09/archives/twins-reared-apart-a-living-lab.html.

Climenhaga, N. (2017a). "Inference to the Best Explanation Made Incoherent," *Journal of Philosophy*, 114 (5): 251-273.

Climenhaga, N. (2017b). "How Explanation Guides Confirmation", *Philosophy of Science*, 84*:* 359-368.

Conee, E. & Feldman, R. (2008). "Evidence," in Q. Smith (ed.), *Epistemology: New Essays.* Oxford University Press, 83-104.

Crupi, V. and Tentori, K. (2012). "A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems)", *Philosophy of Science*, 79: 365–85.

Darwin, C. (1872). *On the Origin of Species*. London: John Murray.
Della Roca, M. (2010). "PSR", *Philosophers' Imprint*, 10(7): 1-13.
Dellsén, F. (2017). "Reactionary responses to the Bad Lot Objection," *Studies in History and Philosophy of Science Part A*, 61: 32-40.
Dellsén, F. (2018). "The Heuristic Conception of Inference to the Best Explanation," *Philosophical Studies*, 175: 1745–1766.
Dellsén, F. (2021). "Explanatory Consolidation: From 'Best' to 'Good Enough'," *Philosophy and Phenomenological Research*, 103(1): 157-177.
Douven, I. (2017) "Inference to the Best Explanation: What Is It? And Why Should We Care?" in T. Poston and K. McCain (eds.) *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press, pp. 7-24.
Douven, I. (2021). "Abduction", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), ed. E. N. Zalta, URL = <http://plato.stanford.edu/archives/spr2011/entri,es/abduction/>
Douven, I. and Schupbach, J. (2015). "Probabilistic alternatives to Bayesianism: the case of explanationism," *Front. Psychol.*, (6): 459. 1-8. doi: 10.3389/fpsyg.2015.00459.
Douven, I. and Wenmackers, S. (2017). "Inference to the Best Explanation versus Bayes's Rule in a Social Setting," *British Journal for the Philosophy of Science*, 68: 535–570.
Douven, I., & Mirabile, P. (2018). "Best, second-best, and good-enough explanations: How they matter to reasoning," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(11), 1792–1813.
Douven, I., (1999). "Inference to the Best Explanation Made Coherent," *Philosophy of Science*, 66: S424–S435.
Dowe, D., Gardner, S. & Oppy, G. (2007). "Bayes Not Bust! Why Simplicity Is No Problem for Bayesians," *Brit. J. Phil. Sci.*, 58(4): 709–54.
Duhem, P. (1906/1954). *La Théorie Physique. Son Objet, sa Structure*. Paris: Chevalier & Riviére.
Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.
Einstein, A. (1905). "Concerning an heuristic point of view towards the emission and transformation of light", *American Journal of* Physics, 33(5): 132-148.
Elgin, C. (2020). "Epistemic Gatekeepers: The Role of Aesthetic Factors in Science," in French, S., & Ivanova, M. (eds.), *The Aesthetics of Science: Beauty, imagination and understanding*. New York: Routledge, 21-35.
Enoch, D. and Schechter, J. (2008). "How are basic belief-forming methods justified?" *Philosophy and Phenomenological Research*, 76: 547-579.
Fölsing, A. (1997). *Albert Einstein*: *A Biography*. New York: Viking.
Forster, M. & Sober, E. (1994). "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions", *British Journal for the Philosophy of Science*, 45(1): 1-35.
Fricker, E. (2004). "Testimony: Knowing through Being Told," In I. Niiiniluoto, M. Sintonen, & J. Wolensi (eds.), *Handbook of Epistemology* (pp. 109–130). Kluwer Academic Publishers.
Fricker, E. (2017). "Inference to the Best Explanation and the Receipt of Testimony: Testimonial Reductionism Vindicated," in K. McCain and T. Poston (eds.) *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford University Press, pp. 262–94.
Fumerton, R. (2017). "Reasoning to the Best Explanation," In K. McCain and T. Poston (eds.), *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press, pp. 65-79.
Gauch, H.G. (2012). *Scientific Method In Brief*. Cambridge, UK: Cambridge University Press.
Glass, D. H. (2012). 'Inference to the Best Explanation: Does It Track Truth?', *Synthese*, 185: 411–27

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
Good, I.J. (1967). "The white shoe is a red herring," *The British Journal for the Philosophy of Science*, 17: 322.
Gopnik, A. (1998). "Explanation as Orgasm," *Minds and Machines* 8 (1): 101–18.
Haig, B. (2009). "Inference to the best explanation: A neglected approach to theory appraisal in psychology," *The American Journal of Psychology*, 122 (2): 219–234.
Hanson, N. (1958). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge, MA: Harvard University Press.
Hanson, N. (1961). "Is There a Logic of Discovery?", in H. Feigl and G. Maxwell (eds.), *Current Issues in the Philosophy of Science*. New York: Holt, Rinehart, and Winston, 20-35.
Harman, G. (1965). "The Inference to the Best Explanation," *Philosophical Review*, 74: 88–95.
Hempel, C. & Oppenheim, P. (1948). "Studies in the Logic of Explanation," Philosophy *of Science*, 15: 135–175.
Hempel, C. (1965). *Aspects of Scientific Explanation*. 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.
Hobbs, J. R., (2004). "Abduction in Natural Language Understanding," in L. Horn and G. Ward (eds.), *The Handbook of Pragmatics*, Oxford: Blackwell, pp. 724–741.
Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press.
Huemer, M. (2009). "Explanationist Aid for the Theory of Inductive Logic," *Brit. J. Phil. Sci.*, 60: 345–75.
Iranzo V. (2001). "Bad Lots, Good Explanations," *Critica*, 3(98): 71-96.
Iranzo V. (2016)." Explanatory Reasoning: A Probabilistic Interpretation," In: Redmond J., Pombo Martins O., Nepomuceno Fernández Á. (eds.), *Epistemology, Knowledge and the Impact of Interaction*. Springer, Cham. https://doi.org/10.1007/978-3-319-26506-3_20.
Isaacs, Y., & Hawthorne, J. (2018). "Fine-tuning Fine-tuning," In J. Hawthorne, M.A. Benton, & D. Rabinowitz (eds.), *Knowledge, Belief, and God: New Insights in Religious Epistemology*. New York: Oxford University Press, 136-168.
Ivanova, M. (2020). "Beauty, Truth, and Understanding," in French, S., & Ivanova, M. (eds.), *The Aesthetics of Science: Beauty, imagination and understanding*. New York: Routledge, pp. 86-103.
Jakobsen, M. (2020). "Determining the Need for Explanation," *Faith and Philosophy, 37*(2), 230–241.
Janssen, M. (2002). "Reconsidering a scientific revolution: The case of Einstein versus Lorentz", *Physics in Perspective,* 4: 421–446.
Jeffreys, H. (1931). *Scientific Inference*. London: Macmillan.
Josephson, J. & Josephson, S. (1994). *Abductive Inference*. Cambridge: Cambridge University Press.
Khalifa, K., Millson, J., and Risjord, M. (2017). "Inference to the Best Explanation: Fundamentalism's Failures," in K. McCain and T. Poston (eds.) *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford University Press, pp. 80-96.
Kidd, I. J. (2011). "Pierre Duhem's epistemic aims and the intellectual virtue of humility: A reply to Ivanova," *Studies in History and Philosophy of Science Part A,* 42(1), 185–189.
Kitcher, P. (1981). "Explanatory Unification," *Philosophy of Science* 48(4): 507–31.
Kuhn, T. (1962). *The Structure of Scientific Revolutions*, *1st ed.* Chicago: University of Chicago Press (2nd 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.

Ladyman J. (2002). *Understanding Philosophy of Science*. New York: Routledge.
Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.
Lando, T. (2016). "Coincidence and Common Cause," *Nous*, 51(1): 132-51.
Lange, M. (2016). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford University Press.
Lange, M. (2017). "The Evidential Relevance of Explanatoriness: A Reply to Roche and Sober," Analysis, 77 (2): 303-312.
Laudan, L. (1977). *Progress and its Problems: Toward a Theory of Scientific Growth.* Berkeley, CA: University of California Press.
Leibniz, G.W. (1714). *The Monadology.*
Leiter, B. (2014) "Moral Skepticism and Moral Disagreement in Nietzsche". In R. Shafer-Landau, (ed.), *Oxford Studies in Metaethics. Vol. 9*, 126–151. Oxford: Oxford University Press.
Lewis, D. (1973). *Counterfactuals.* Cambridge, MA: Harvard University Press.
Lewis, D. (1986). "Causal Explanation", in *Philosophical Papers, Vol. II.* Oxford: Oxford University Press, 214-40.
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*, *2nd ed.* New York: Routledge.
Longino, H. (1996). "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy," in L. Hankinson Nelson & J. Nelson (eds.), *Feminism, Science, and the Philosophy of Science. Dordrecht*: Kluwer Academic Publishers. 39-58.
Lycan, W.G. (1988). *Judgement and justification*. Cambridge: Cambridge University Press.
Lycan, W.G. (2002). "Explanation and Epistemology", in Paul Moser (ed.), *The Oxford Handbook of Epistemology.* Oxford: Oxford University Press.
Magnani, L. (2000). *Abduction, Reason, and Science: Processes of Discovery and Explanation*. New York: Springer.
Magnani L. (2009). *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning.* Berlin: Springer.
Matthews, M. (2015). *Science teaching: The Contribution of History and Philosophy of Science* (2nd ed.). New York: Routledge.
McAuliffe, W. (2015). "How did Abduction Get Confused with Inference to the Best Explanation?", *Transactions of the Charles S. Peirce Society*, 51(3): 300-319.
McCain, K. and Poston, T. (2017a). "Best Explanations: An Introduction," In K. McCain and T. Poston (eds.), *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press, (pp. 1-6).
McCain, K. and Poston, T. (eds.) (2017b). *Best Explanations: New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press,
McCain, K., & Poston, T. (2017c). "The Evidential Impact of Explanatory Considerations", in McCain and Poston (eds.), *Best Explanations: New Essays on Inference to the Best Explanation.* Oxford: Oxford University Press, pp. 121-29.
McGrew, T. (2003). "Confirmation, heuristics, and explanatory reasoning," *Brit. J. Phil. Sci.*, 54 (4): 553-567
McKaughan, D. (2008). "From Ugly Duckling to Swan: C. S. Peirce, Abduction, and the Pursuit of Scientific Theories," *Transactions of Charles S. Peirce Society*, 4(3): 446–68.
Michelson, A. (1896). "XIX. The Department of Physics", in *Annual Register.* Chicago: University of Chicago Press, 159-162.

Millikan, R. (1916). "A Direct Photoelectric Determination of Planck's 'h' ", *Physical Review* 7(3): 355–388.
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.
Mohammadian, M. (2021). "Abduction: The Context of Discovery + Underdetermination = Inference to the Best Explanation," *Synthese*, 198: 4205–4228.
Myrvold, W. (2017). "On the Evidential Import of Unification", *Philosophy of Science*, 84(1): 92–114.
Norton, J. D. (2021). *The Material Theory of Induction*. University of Calgary Press.
Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.
Okasha, S. (2000). "Van Fraassen's Critique of Inference to the Best Explanation," *Studies in the History and Philosophy of Science,* 31: 691-710.
Park, W. (2015). "On Classifying Abduction", *Journal of Applied Logic*, 13(3): 215-238.
Paul, L.A. (2012). "Metaphysics as Modeling: The Handmaiden's Tale", *Philosophical Studies*, 160: 1-29.
Peirce, C. S. (1934). *The Collected Papers of Charles Sanders Peirce*: Cambridge, MA: Harvard University Press.
Popper, K. (1959). *The Logic of Scientific Discovery*. London: Routledge.
Poston, T. (2014). *Reason & Explanation: A Defense of Explanatory Coherentism*. New York: Palgrave-MacMillan.
Pruss, A. (2006). *The Principle of Sufficient Reason*: *A Reassessment*. Cambridge University Press.
Psillos, S. (1996). "On van Fraassen's critique of abductive reasoning," *Philosophical Quarterly*, 46(182): 31–47.
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 F. Stadler (ed.), *Induction and Deduction in the Sciences. Dordrecht: Kluwer Academic Press*, 83-91.
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. & Ullian, J. (1978). *The Web of Belief*. New York: Random House.
Railton, P. (1978). "A Deductive-Nomological Model of Probabilistic Explanation," *Philosophy of Science*, 45: 206–26.
Reichenbach, H. (1938). *Experience and Prediction.* Chicago: University of Chicago Press.
Rescher, N. (1991). *G. W. Leibniz's Monadology*. Pittsburgh, PA: University of Pittsburgh Press.
Reutlinger, A. (2017). "Explanation Beyond Causation? New Directions in the Philosophy of Scientific Explanation", *Philosophy Compass*, 12(2), doi.org/10.1111/phc3.12395.
Roche, W. & Sober, E. (2013). "Explanatoriness is evidentially irrelevant, or inference to the best explanation meets Bayesian confirmation theory," *Analysis*, 73: 659-668.
Roche, W. (2018). "The Perils of Parsimony," *The Journal of Philosophy*, 115 (9):485-505.
Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton University Press.
Salmon, W. (1989). "Four Decades of Scientific Explanation", in P. Kitcher and W. Salmon (eds.), *Scientific Explanation,* Vol. 13, *Minnesota Studies in the Philosophy of Science.* Minneapolis: University of Minnesota Press, 3-219.
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. (2001). "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. (2005). *Reality and Rationality*. Oxford: Oxford University Press.
Schickore, J. & Steinle, F. (2006). *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*. Dordrecht: Springer.
Schindler, S. (2018). *Theoretical Virtues in Science: Uncovering Reality through Theory*. Cambridge University Press.
Schupbach, J. & Sprenger, J. (2011). 'The Logic of Explanatory Power', *Philosophy of Science*, 78: 105–27.
Schupbach, J. (2014). "Is the Bad Lot Objection Just Misguided?" *Erkenntnis*, 79(1): 55-64.
Schupbach, J. (2017). "Inference to the Best Explanation, Cleaned Up and Made Respectable", in K. McCain and T. Poston (eds.), *Best Explanations: New Essays on Inference to the Best Explanation*. Oxford: Oxford University Press, pp. 39-61.
Schurz, G. (1999). "Explanation as Unification," *Synthese*, 120(1): 95-114.
Sider, T. (2009). "Ontological antirealism", in D. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics: New Essays on the Foundation of Ontology*. New York: Oxford University Press, 384–421.
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. (1988). *Reconstructing the Past, Parsimony, Evolution, and Inference*. Cambridge: MIT Press.
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. (2009). "Parsimony Arguments in Science and Philosophy—A Test Case for Naturalism$_P$", *Proceedings and Addresses of the American Philosophical* Association, 83(2): 117-55.
Sober, E. (2012). "Coincidences and How to Reason about Them", in H. de Regt, S. Hartmann, & S. Okasha (eds.), *European Philosophy of Science Association 2009*. Springer, Amsterdam, 355–74.
Sober, E. (2015). *Ockham's Razors: A User's Manual.* Cambridge: Cambridge University Press.
Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives.* New York: Oxford University Press.
Stump, D. (2007). "Pierre Duhem's virtue epistemology," *Studies in History and Philosophy of Science Part A, 38*(1), 149–159.
Swartz, N. (1985). *The Concept of Physical Law*. New York: Cambridge University Press.
Swinburne, R. (1997). *Simplicity as Evidence for Truth*. Milwaukee: Marquette University Press.
Swoyer, C. (1999). "Explanation and inference in metaphysics", *Midwest Studies in Philosophy*, 23:100–131.
Thagard, P. (1978). The Best Explanation: Criteria for Theory Choice," *The Journal of Philosophy*, 75(2): 76-92.
Tulodziecki, D. (2011). "A case study in explanatory power: John Snow's conclusions about the pathology and transmission of cholera", *Stud. History Philos. Biol. Biomed. Sci.*, 42: 306–316.
van Fraassen, B.C. (1980). *The Scientific Image.* Oxford: Oxford University Press.
van Fraassen, B.C. (1989). *Laws and Symmetry.* Oxford: Oxford University Press.
Vogel, J. (1990). "Cartesian Skepticism and Inference to the Best Explanation", *Journal of Philosophy*, 87: 658-666.
Walker, D. (2012). "A Kuhnian Defence of Inference to the Best Explanation," *Studies in History and Philosophy of Science*, 43: 64–73.
Weisberg, J. (2009). "Locating IBE in the Bayesian Framework." *Synthese*, 167: 125–43.

Whewell, W. (1840/1968).*William Whewell's Theory of Scientific Method*. R. Butts (ed.), Pittsburgh, PA: University of Pittsburgh Press.
White, R. (2000). "Fine-Tuning and Multiple Universes", Nous, 34(2): 260–76.
White, R. (2005). "Explanation as a Guide to Induction," *Philosophers' Imprint, 5*(2), 1–29.
Williamson, T. (2016). "Abductive Philosophy," *The Philosophical Forum*, 47: 263–80.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.