DOI: 10.1111/nous.12349

ARTICLE

NOÛS

Putting explanation back into "inference to the best explanation"

Marc Lange

The University of North Carolina at Chapel Hill

Correspondence
Marc Lange, The University of North Carolina at Chapel Hill.
Email: mlange@email.unc.edu

Abstract

Many philosophers argue that explanatoriness plays no special role in confirmation – that "inference to the best explanation" (IBE) incorrectly demands giving hypotheses extra credit for their potential explanatory qualities beyond the credit they already deserve for their predictive successes. This paper argues against one common strategy for responding to this thought – that is, for trying to fit IBE within a Bayesian framework. That strategy argues that a hypothesis' explanatory quality (its "loveliness") contributes either to its prior probability or to its likelihood. This paper argues that this strategy fails because it must give different treatments to two hypotheses that are unlovely for the very same reason. The strategy therefore loses the insight into scientific reasoning that its reconstruction in terms of IBE is supposed to provide. The paper then provides a Bayesian account of the confirmatory role of explanatoriness that represents explanatory quality as having an impact in the same place for two hypotheses that are unlovely for the very same reason. This approach works by "putting explanation back into IBE" – that is, by invoking hypotheses that refer explicitly to scientific explanation and by invoking the agent's background opinions regarding the kind of explanation that the evidence is liable to have. On this approach, there is no list of "explanatory virtues" the possession of which always helps to make an explanation better. Rather, for different facts, there are different characteristics that our background knowledge of other explanations gives us some reason to expect the given fact's explanation to possess.

1 | INTRODUCTION

Sometimes we argue that one hypothesis derives some plausibility over its rivals from the fact that the explanations it would give (if it were true) are better than those its rivals would give (if they were true). In arguing that evolution by natural selection derives support from the quality of the explanations it would give, Darwin notes that this kind of argument has often been used:

> It can hardly be supposed that a false theory would explain, *in so satisfactory a manner as does the theory of natural selection*, the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing; but it is a method used in judging of the common events of life, and has often been used by the greatest natural philosophers. The undulatory theory of light has thus been arrived at; and the belief in the revolution of the earth on its own axis was until lately supported by hardly any direct evidence. (Darwin, 1873:421, my emphasis)

Following Harman (1965), philosophers have generally termed this "method of arguing" *inference to the best explanation* (IBE).[1]

Views about IBE range widely. At one extreme, Harman (1965) holds that all non-demonstrative inference is fundamentally IBE, a view that Lycan (2002:417) says is not widely accepted (calling it "ferocious explanationism"). At the opposite extreme, some philosophers follow van Fraassen (1980,1989) in denying that the explanatory quality of a hypothesis is ever an epistemic reason for according it greater credence:

> Judgements of simplicity and explanatory power **…** are specifically human concerns, a function of our interests and pleasures, which make some theories more valuable and appealing to us than others. Values of this sort, however, … cannot rationally guide our epistemic attitudes and decisions. For example, if it matters more to us to have one sort of question answered rather than another, that is no reason to think that a theory which answers more of the first sort of questions is more likely to be true (not even with the proviso "everything else being equal"). (van Fraassen, 1980:87)

Van Fraassen argues that for IBE to be rational, it would have to be rational for us to adopt the policy of awarding bonus credibility points to hypotheses for their explanatoriness, augmenting their degrees of credence beyond what Bayesian Conditionalization (BC) assigns them. Van Fraassen (1989:160–161) argues that any such policy is irrational, making one vulnerable to a dynamic Dutch Book. He argues that a hypothesis's making us feel good explanationwise is not any evidence of its truth. On his view, a scientist may use a hypothesis's explanatory quality as a reason (but not an *epistemic* reason) to accept the hypothesis. Many philosophers – including Achinstein (2013), Salmon (2001a, 2001b), and Sober (1990, 2015a) – have agreed with van Fraassen that a hypothesis's explanatory quality is not an epistemic reason for believing in its truth.

My aim in this paper is to show that contrary to van Fraassen, IBE can be reconciled with BC. I defend IBE by arguing that under BC, the explanatoriness of some hypothesis can be a distinct epistemic reason for believing in its truth.[2] I begin (in section 2) by clarifying what it would take for IBE to operate. For example, IBE permits explanatory quality to be overridden in theory choice by other considerations, and IBE can operate in confirmation that falls short of the full acceptance of any theory being confirmed. In section 3, I argue that a prominent strategy for reconciling

IBE with BC – pursued by defenders of IBE such as Lipton (2001b), McGrew (2003), and Okasha (2000) – pays an excessive price for fitting IBE into BC: it makes explanatory quality disappear as a distinct epistemic reason for placing greater confidence in a hypothesis. According to this strategy, the quality of the explanation that hypothesis H would give of evidence E has an impact on H's confirmation either by affecting H's "prior probability" pr(H) or by influencing H's "likelihood" pr(E|H), each of which figures in BC. I apply this strategy to the rivalry between the Copernican and Ptolemaic models of the heavens, where (as Darwin mentioned) explanatory considerations favored the Copernican model. I show that this strategy must treat *differently* various cases where explanatory considerations make *exactly the same* important contribution to confirmation. As a result, an important insight that IBE should bring to our understanding of scientific reasoning is lost under this strategy for combining IBE with BC. Ultimately, the problem with this strategy is that it assigns no confirmatory significance to explanatory quality *per se*, only to its surrogates pr(H) and pr(E|H).

This problem can be avoided by putting explanation back into inference to the best explanation. In section 4, I describe one way in which H's explanatory quality *per se* can make a difference to H's confirmation and I show that this difference arises under BC. Roughly speaking, I argue that H may derive some credibility by virtue of its capacity (if true) to supply E with an explanation of the same kind as we have discovered certain other facts to have – where we judge it plausible that E would have whatever kind of explanation those other facts have. In such a case, the confirming evidence and background opinions essentially concern explanations. On this proposal, there is no list of "explanatory virtues" the possession of which always helps to make an explanation better. Rather (as is actually the case in science), for different facts E, there are different characteristics that our background knowledge of other explanations gives us some reason to expect E's explanation to possess. Therefore, to defend IBE, we do not need to find some way to argue that hypotheses displaying certain explanatory features tend more often to be true over the entire history of science. Rather, H's displaying certain explanatory features justifies our placing greater confidence in H because of our background beliefs regarding explanations of certain other facts like E.

Finally (in section 5), I use this account to argue against the claim made by Earman and Roberts (2005) that if the laws of nature fail to supervene on the Humean base, then we could never be epistemically justified on empirical grounds in believing certain facts to be laws (or in believing them not to be laws). IBE enables hypotheses positing laws to be confirmed by the discovery of certain other explanations. We have discovered that the reason why certain things *do not* happen is that they *cannot* happen – where the prohibitions in these explanations arise from mathematical necessities rather than natural laws. These explanations confirm some other hypotheses that (if true) would give similar explanations, but where these hypotheses posit natural laws (rather than mathematical necessities) as making certain things impossible. In this way, the explanatory qualities of certain hypotheses positing laws can count as epistemic reasons for believing them true. We thus have an account of how we discovered natural necessity: by IBE using background knowledge of explanations appealing to mathematical necessity.

2 | WHAT IS IBE?

In this section, I will identify what I take IBE to be. As I will explain, I construe IBE not as a mechanical rule of inference, but rather as a view about what counts as an epistemic reason for

placing some or another degree of confidence in a given hypothesis. IBE is the view that sometimes at least part of our epistemic justification for our degree of confidence in a given hypothesis is how well that hypothesis would explain a given piece of evidence. Accordingly, IBE entails that the fact that one hypothesis would (if it were true) give a better explanation of some evidence than another, incompatible hypothesis would (if it were true) can sometimes count as an epistemic reason for believing more strongly in the truth of the former hypothesis than in the truth of the latter hypothesis.

I see this view as precisely what van Fraassen (1989:149) identifies as the most "sophisticated" form of IBE against which he argues. As he describes the view to which an advocate of IBE should "retrench":

> "Inference to the Best Explanation" was a misnomer, and the rule properly understood leads to a revision of judgement much more modest than inference to the truth of the favoured hypothesis … Despite its name, it is not the rule to infer the truth of the best available explanation. That is only a code for the real rule, which is to allocate our personal probabilities with due respect to explanation. Explanatory power is a mark of truth, not infallible, but a characteristic symptom. … IBE would be a recipe for adjusting our personal probabilities while respecting the *explanatory* (as well as predictive) success of hypotheses. (van Fraassen, 1989:145–6,149)

Van Fraassen then argues that since IBE demands giving hypotheses credit for their explanatory success over and above the credit they receive for their predictive success, IBE endorses the policy of giving hypotheses bonus points for their explanatory quality over and above the points awarded them by BC for their predictive success. As I mentioned earlier, this policy is the target of van Fraassen's central critique of IBE. (Ultimately, I argue below that contrary to van Fraassen, IBE does not require such bonus credibility points – points beyond what BC would accord hypotheses. But I agree with van Fraassen that IBE does require that hypotheses be rewarded for their explanatory quality *per se*.)

One important aspect of IBE (as so construed) is that explanatory quality is "*a* mark of truth" (my emphasis) – just one mark among many. IBE permits explanatory considerations to be overridden by other considerations so that the "best explanation" of one fact need not be the most plausible hypothesis all things considered. Even while recognizing the support bestowed on a given hypothesis by its capacity (if true) to give a good explanation of some fact, we may justly regard a competing hypothesis that would not "best explain" that fact as nevertheless better supported by the entire body of available evidence.

For instance, although the Sun's diameter is about 400 times greater than the Moon's, it is on average about 400 times farther away, so – remarkably – they have nearly equal angular diameters in Earth's sky. (Hence the Moon is just the right apparent size to cover the Sun's disk without covering the Sun's corona, producing spectacular solar eclipses.) The equality of their angular diameters would be better explained by some theory ascribing their diameters and distances to some common cause than by the theory that this equality is coincidental. Yet all things considered, the coincidence theory is better confirmed (Naeye, 2000:94). This judgment is compatible with IBE in the form that van Fraassen entertains and Lipton (in his most careful formulations) defends: "The distinctive claim of Inference to the Best Explanation is that we use judgements of the quality of potential explanations as a guide to likeliness or probability" (Lipton, 2001b:97).

A guide – not the sole guide or "the whole story about the assessment of scientific hypotheses" (Lipton, 2001b:93).[3]

It is incorrect, then, to interpret van Fraassen as Okasha (2000:698) does: as saying that "the retrencher does not need to hold that the best explanation deserves a high epistemic probability, only that it deserves a higher epistemic probability than alternative, less good explanations." Even this latter contention goes beyond the retrencher's conception of IBE; the retrencher recognizes that other considerations may override explanatory quality, making the best potential explainer less plausible overall than "alternative, less good explanations." Nevertheless, Okasha is correct that on the most "sophisticated" version of IBE entertained by van Fraassen, the hypothesis that would give the best explanation (if that hypothesis were true) need not deserve a high epistemic probability and, in particular, may have insufficient support to justify being accepted. In this respect, "*inference* to the best explanation" is a misnomer.

Let us continue refining what IBE consists in. Denying that IBE requires accepting the best of a "bad lot," Lipton (2001b:104; cf. 2004:154) says: "'Inference to the Best Explanation' must thus be glossed by the more accurate but less memorable phrase, 'inference to the best of the available competing explanations, when the best one is sufficiently good'." But this is still not the right way to express IBE. That the best explanation would be a "sufficiently good" *explainer* (if it were true) does not ensure that it is plausible enough, *all things considered*, to justify being accepted. Furthermore, we should not understand IBE as "inference to the best of the available competing explanations, when the best one is sufficiently good" even if we interpret "sufficiently good" as meaning "sufficiently plausible, all things considered" rather than as "sufficiently good at explaining, if true." IBE does not require that explanatory quality be the sole criterion for judging whether a hypothesis that is plausible (all things considered) should be accepted.

Lipton's formulation is not only too strong, but also too weak. As I mentioned, IBE says that the quality of the explanation that some hypothesis would give (if it were true) can count as an epistemic reason for our having a given degree of confidence in the hypothesis, even when we lack sufficient reason to justify accepting (i.e., inferring to) the hypothesis. As McGrew (2003:555) puts it, Lipton's view in the passage I just quoted "is tantamount to saying that when the data are sparse, the [explanatory] virtues are worthless: they can do no epistemic work until there are enough of them in place to justify a definite conclusion." (Lipton is not the only one who officially restricts IBE to cases of acceptance; Thagard (1978:77) likewise characterizes IBE as "accepting a hypothesis on the grounds that it provides a better explanation…".)

Lipton (2004:59) defines the "most lovely" hypothesis (among various rivals) as the hypothesis that "would, if correct, be the most explanatory or provide the most understanding." Thus, "loveliness" is explanatory quality – "a measure of how good a potential explanation is, of how much understanding it would provide if it is or if it were an actual explanation" (Lipton, 2001b:119). Expressed in terms of loveliness, then, I construe IBE as the view that loveliness is a distinct epistemic virtue – that at least part of our justification for our degrees of confidence in certain hypotheses is their loveliness.[4]

Lipton (2001b:93–4, 105) emphasizes that loveliness must be distinguished from likeliness, since IBE would shed no light on confirmation if it were the view that (at least part of) our justification for our degree of confidence in a given hypothesis is its likeliness. I will now argue that on one popular strategy for reconciling IBE with BC, we would again lose the light that IBE is supposed to shed on the way that explanatory quality functions in confirmation.

3 | HOW EXPLANATION THREATENS TO DISAPPEAR FROM "INFERENCE TO THE BEST EXPLANATION"

Lipton (2001b:105) remarks that those who embrace IBE owe us an account of the factors that make a hypothesis more or less lovely (i.e., that contribute to a putative explanation's goodness). Harman (1965:89), for instance, mentions simplicity, plausibility, scope of explanation, and lack of ad hocness. However, Salmon (2001b:125) rightly notes that plausibility and lack of ad hocness "pertain directly to prior probabilities; indeed, I would take plausibility to be the very same thing as prior probability." If this is loveliness, then loveliness threatens to become nothing but likeliness. In that case, as we saw at the end of the previous section, IBE would shed no light on confirmation. Lipton (2001b:106) offers a somewhat different list of explanatory virtues: "Better explanations explain more types of phenomena, explain them with greater precision, provide more information about underlying mechanisms, unify apparently disparate phenomena, or simplify our overall picture of the world."

Of course, after identifying the explanatory virtues, an account of IBE should reveal why we should regard them as guides to the truth.[5] Salmon maintains that loveliness does not make a hypothesis more plausible: "*explanatory* beauty does not enter into our choice of the hypothesis as most likely" (Salmon, 2001a:74). Likewise, Achinstein (2013:114), after quoting van Fraassen approvingly, writes that loveliness is not an epistemic virtue: "'Loveliness' makes for a better theory, not for a believable or more believable one." Salmon maintains that BC leaves no work for IBE to do in helping us to understand confirmation: "Bayesian confirmation can account for the probabilistic evaluation of the statements that are employed in constructing explanations. … The probabilities that enter into Bayes's theorem refer to epistemic, not explanatory, virtues" (Salmon, 2001a:88).

Some of IBE's supporters have attempted to defend it from this charge by arguing that IBE has a place within the Bayesian approach. The obvious suggestion for them to make is that H's loveliness in how it would explain E makes itself felt in BC by contributing either to H's prior probability pr(H) or to H's likelihood pr(E|H) – always to one or to the other factor (or to both), but to different factors in different cases. In particular, Lipton (2004), McGrew (2003), and Okasha (2000) propose that T_1's being more lovely than T_2 in its explanation of E entails that either $pr(T_1) > pr(T_2)$ or $pr(E|T_1) > pr(E|T_2)$. For instance, Okasha (2000:73) writes:

> The correct way of representing IBE, I suggest, views the goodness of explanation of a hypothesis *vis-à-vis* a piece of data as reflected in the prior probability of the hypothesis P(H), and the probability of the data given the hypothesis P(e/H). The better the explanation, the higher is one or both of these probabilities. Relative to this account, favouring a hypothesis on the grounds that it provides a better explanation of one's data than other hypotheses, and indeed making it a rule to do so, is perfectly consistent with Bayesian principles.

Lipton and McGrew also suggest that (in Lipton's (2001b:111) words) loveliness "is in fact the scientists' guide to" prior probability or likelihood (as the case may be) in that H's loveliness is often more epistemically accessible to an ordinary epistemic agent than her own personal prior probability and likelihood. For instance, Lipton (2001b:111–112) writes:

> On this proposal, the resulting transition of probabilities in the face of new evidence might well be just as the Bayesian says, but the mechanism that actually brings about the change is explanationist. … What would be required, I think, is that lovelier explanations tend to make what they explain likelier (even if high likelihood is no guarantee of good explanation), and that we sometimes exploit this connection by using judgements of loveliness as a barometer of likelihood. For example, when we consider the loveliness of a potential causal explanation, we may consider how the mechanism linking cause and effect might run, and in so doing we are helped in forming a judgement of how likely the cause would make the effect and how unlikely the effect would be without the cause. … Explanatory loveliness is used as a symptom of likelihood, and likelihoods help to determine … posterior probability. This is one way Inference to the Best Explanation and Bayesianism may be brought together.

Although this example involves scientists using H's loveliness as a guide to pr(E|H), scientists in other cases (according to Lipton) use H's loveliness as a guide to pr(H) because "the priors assigned to competing explanations … were themselves generated in part with the help of explanatory considerations" (Lipton, 2001b:113).

I will devote this section to examining this view more closely and, ultimately, to arguing that this view fails to capture the role in confirmation that is played by considerations of explanatory quality ("loveliness"). To begin with, is H_1's being lovelier than H_2 in its potential explanation of E extensionally equivalent to $pr(H_1) > pr(H_2)$ or $pr(E|H_1) > pr(E|H_2)$? No, since it can be that H has a high prior probability and E is likely given H although H's truth would not explain E. For instance, if E is that it will rain in Seattle on a given future date and H is that the streets will be wet there on that date, then H (even if true) would not explain E and yet both pr(H) and pr(E|H) are high. But advocates of this view recognize this possibility. They argue that their view does not require that H_1's having higher loveliness than H_2 be *necessary* for $pr(H_1) > pr(H_2)$ or $pr(E|H_1) > pr(E|H_2)$. Their view requires only that H_1's having higher loveliness than H_2 be *sufficient* for $pr(H_1) > pr(H_2)$ or $pr(E|H_1) > pr(E|H_2)$ – and, perhaps, that loveliness tend to be epistemically more accessible than these factors – since sufficiency is enough for higher loveliness always to be manifested in one or another factor in BC.[6]

In making this point, Okasha (2000:705) is nicely explicit in the course of rejecting an objection to his view:

> Objection…: but there are many cases where both P(T) and P(e/T) are high, and yet T does not explain e at all, less still provide the best explanation of e. Reply: true but irrelevant. Certainly, T can be a well-established theory which entails but fails to explain e, as many counterexamples to the D-N model of explanation show; in such a case, P(T) will be high and P(e/T) equal to one. But this does not undermine my proposed way of modelling IBE in Bayesian terms. It only highlights the obvious fact that not all cases of updating by Bayesian conditionalization involve explanatory considerations. My claim is that when scientists do attach confirmatory weight to a theory because the theory yields a better explanation of the evidence than rival theories, this piece of reasoning can be given a plausible reconstruction in Bayesian terms. That is compatible with allowing that not all cases of conditionalization are cases of IBE. So the fact that high values for P(T) and P(e/T) do not suffice for T to explain e is not to the point. … What my position does require is this: if one regards T1 as a better explanation of e than T2, then one must either set P(e/T1) > P(e/T2), or P(T1) > P(T2), or

both. This is crucial to my proposed reconciliation of IBE with Bayesianism, and it seems perfectly reasonable. Indeed, it is hard to see what it could mean to believe that T1 explains e better than T2 if one's personal probability function satisfied neither of the above inequalities.

However, the claim that Okasha says his position requires is implausible. Suppose T_1 is that Jones comes in contact with some infected matter, E is that Jones is afflicted with the infection, and T_2 is that Jones's blood contains antibodies to the infection. We can have $pr(E|T_1) < pr(E|T_2)$ and $pr(T_1) < pr(T_2)$ even though T_1 explains E better than T_2 does – since T_2 cannot explain E; on the contrary, E causes and hence can explain T_2. For instance, we can have $pr(E|T_1) = .2 < .4 = pr(E|T_2)$ and $pr(T_1) = .01 < .1 = pr(T_2)$ while T_1 causes E (we can have $pr(E|T_1) = .2 > .1 = pr(E)$, for those interested in the probabilities, by having $pr(E|{\sim}T_1) = .09898\ldots$) and E causes T_2 (we can have $pr(T_2|E) = .4 > .1 = p(T_2)$ by having $pr(T_2|{\sim}E) = .06666\ldots$).

Perhaps Okasha intends to claim that if one regards T_1 (if true) as a better explanation of E than T_2 (if true), *but one believes that T_2 (if true) would explain E*, then one must either set $pr(E|T_1) > pr(E|T_2)$ or $pr(T_1) > pr(T_2)$.[7] This restricted claim would not be undermined by the example I just gave, since in that example, one does not believe that T_2 (if true) would explain E. However, the italicized restriction may seem inappropriate in a view that is supposed to make explanatory quality count as an epistemic reason in favor of believing (or according higher credence to) a hypothesis. If T_1's capacity (if true) to give a better explanation of E than T_2 (if true) can sometimes count as a reason favoring belief in T_1 over T_2 even when T_2 (if true) would explain E, then why wouldn't this explanatory consideration sometimes count as a reason favoring T_1 over T_2 when T_2 (even if true) would not explain E and T_1 (if true) would explain E – and so perforce T_1 (if true) gives a better explanation of E than T_2 (if true)? Sober (2015a) thinks that it is a problem for IBE that it incorrectly entails that a hypothesis that would (if true) explain E is always better confirmed by E than a rival hypothesis that (even if true) would not explain E. Sober compares "T is true" to "T is empirically adequate" and to "T is false in what it says about unobservables but true in what it says about observables." He points out that although neither of the latter two hypotheses can explain why T has been empirically successful so far, we have

$$\begin{aligned}&pr(\text{T has been empirically successful so far} \mid \text{T is true}) =\\ &pr(\text{T has been empirically successful so far} \mid \text{T is empirically adequate}) =\\ &pr(\text{T has been empirically successful so far} \mid \text{T is false in what it says about unobservables but true in what it says about observables})\end{aligned}$$

and so, Sober (2015a:913) concludes, the "likelihood principle" (that E favors T_1 over T_2 to the extent that $pr(E|T_1) > pr(E|T_2)$) undercuts IBE: "the law of likelihood does not care about 'explanation'." We could add that there is nothing to prevent pr(T is true) from being less than pr(T is empirically adequate) and pr(T is false in what it says about unobservables but true in what it says about observables). So by the same reasoning (extended to encompass the priors as well as the likelihoods), BC joins the law of likelihood in not caring about explanation – undermining Okasha's strategy for incorporating IBE within BC, unless his claim is restricted to the case where any of the rival hypotheses would (if true) explain E. But Sober apparently sees no motivation from IBE for imposing this restriction: IBE should favor a good explainer over a non-explainer for the same reason as it favors a good explainer over a less good one.

More importantly, even as restricted to cases where T_2 (if true) would explain E, it is implausible to claim (with Okasha, Lipton, and McGrew) that if one regards T_1 (if true) as a better explanation

of E than T_2 (if true), then one must either set $pr(E|T_1) > pr(E|T_2)$ or $pr(T_1) > pr(T_2)$. Higher loveliness (between two rival hypotheses that each would explain, if true) is not always accompanied by either higher likelihood or higher prior probability. As we saw in the previous section, IBE permits explanatory considerations to be overridden by other considerations so that the "best explanation" need not be the most plausible hypothesis all things considered. Therefore, even if T_1's loveliness contributes toward $pr(T_1)$ or toward $pr(E|T_1)$, other factors could override its contribution so that neither $pr(E|T_1) > pr(E|T_2)$ nor $pr(T_1) > pr(T_2)$. It is easy to imagine this happening in the case we discussed in the previous section: where E is that the Sun's and Moon's angular diameters in the Earth's sky are about equal, T_1 is a proposed explanation that depicts the two diameters as having an important common cause, and T_2 is a rival proposed explanation that depicts the two diameters' equality as coincidental. As a second example, suppose P is the Ptolemaic geocentric model of the heavens, which puts a certain period (about 365 days) in the separate models of the positions of the Sun and various planets, but gives no common reason why this period appears in all of them. Suppose C is the Copernican heliocentric model of the heavens, which proposes a common reason why that period appears in all of these models (namely, because it is really the Earth's orbital period, not the periods of each of various unconnected crystalline spheres). We can well imagine a scientist (after Copernicus and before Galileo's telescopic discoveries) taking the Copernican model as better at explaining the appearances than the Ptolemaic model (just as in 1596 Kepler (1981:75-6, 81; cf. Jardine, 1984:141,145) did). Nevertheless, the scientist's opinions might well be that $pr(P) > pr(C)$, by virtue of P's fitting and C's failing to fit with accepted (Aristotelian) physics (as in Ptolemy, 1984: 45)[8] and also that $pr(E|C)$ is no higher than $pr(E|P)$ since C and P "save the phenomena" about equally well (Gingerich, 1975:86).[9]

Because the priors and likelihoods lump together explanatory considerations with all of the other factors relevant to confirmation, I deny that if T_1 gives a lovelier explanation than T_2, then its prior or its likelihood must be higher. But I think that Okasha is mistaken in regarding this conditional as required by his general strategy for integrating IBE within BC. That strategy, as I understand it, is to regard loveliness as *contributing* to the prior or the likelihood. Its contribution, as we have seen, can be outweighed by others; loveliness is not the only relevant consideration. But all that is required to vindicate IBE, on this strategy, is for loveliness to make *some* contribution (even if outweighed) in every case; a hypothesis must get some credit for its loveliness, and the only kinds of credit available under BC are in terms of the prior or the likelihood. This view would entail that in a case where T_1's greater loveliness is the *principal* relevant factor, then either $pr(T_1) > pr(T_2)$ or $pr(E|T_1) > pr(E|T_2)$. But otherwise, this disjunction need not hold even though IBE correctly identifies loveliness as among the factors responsible for the priors and likelihoods.

Having now arrived at this refined version of the approach pursued by Lipton, McGrew, and Okasha, I will devote the remainder of this section to arguing that even this refined approach is inadequate to vindicate IBE. One way of expressing the problem is that this account must treat differently various cases in which explanatory considerations make exactly the same important contribution to confirmation. In particular, this account must regard loveliness as in some cases contributing to priors and in other cases contributing to likelihoods even when the explanatory considerations in the cases are exactly the same and so should be understood as playing the same role in confirmation. IBE properly treats as alike various cases that the account I have just described must treat differently. Therefore, an important insight that IBE should supply to our understanding of scientific reasoning is lost under the above strategy for assimilating IBE to BC.

For example, let's return to the Copernican model's explanatory superiority over the Ptolemaic model that depicts the planets as moving on small circles ("epicycles") whose centers orbit the Earth (see Figure 1). As the evidence, take the long-known fact that a "superior planet" (that is, a

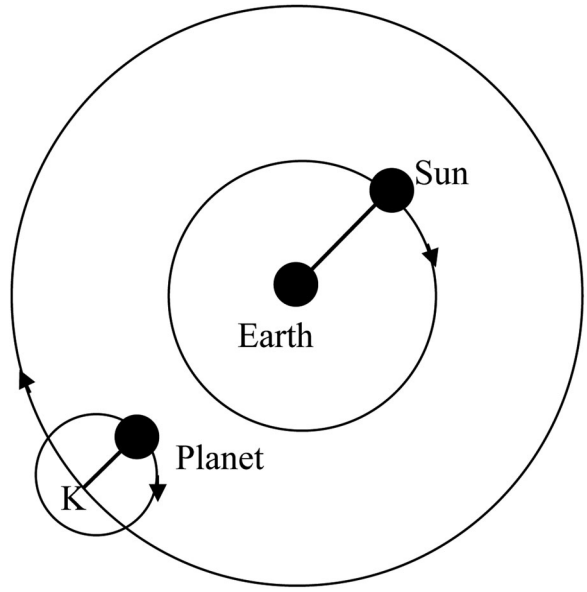

FIGURE 1 A superior planet at opposition to the Sun on the Ptolemaic model. The Sun orbits Earth while the planet orbits on an epicycle, the center (K) of which orbits Earth on a larger circle. As long as the orbital velocities are fine-tuned so that the line from K to the planet remains parallel to the line from the Earth to the Sun, the planet will reach perigee (the point on its epicycle that is closest to Earth) exactly at opposition. The two lines must turn synchronously

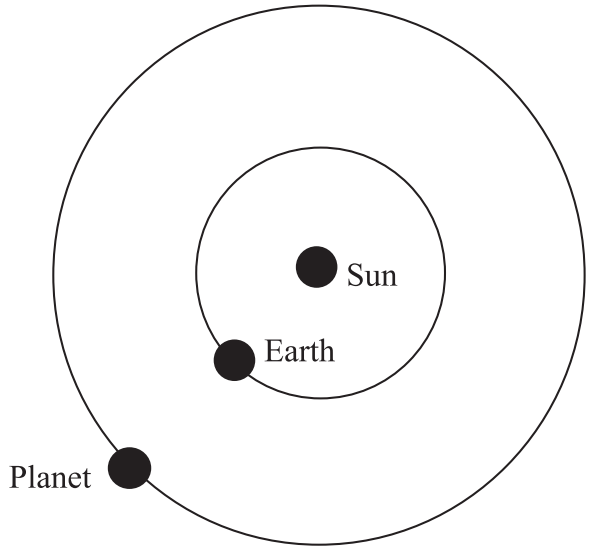

FIGURE 2 Perigee on the Copernican model occurs automatically at opposition. If the planet were anywhere else on its orbit (with the Earth's position unchanged), then the planet would be farther from the Earth and no longer at opposition

planet that takes more than one earth-year to complete one circuit through the zodiac in Earth's night sky – for instance, Mars, Jupiter, and Saturn) is at "opposition" (opposite the Sun, as seen from Earth – rising in Earth's sky at sunset and setting at sunrise) exactly when the planet becomes brightest as seen from Earth and so when it is at its "perigee" (its closest approach to Earth). The Ptolemaic model can explain this correlation, but only by making each superior planet's epicycle turn at the same rate (angular velocity) as the Sun orbits the Earth.

In contrast, the correlation between opposition and perigee falls immediately out of the geometry of the Copernican model (see Figure 2); the correlation would hold no matter what the orbital velocities were (as long as the Earth sometimes passes between the superior planet and the Sun). The Copernican model (if true) explains the opposition/perigee correlation without having to posit any special fine-tuning of the orbital parameters, any suspicious correlation among the angular velocities of different planets' epicycles, or any mysterious incorporation of the Sun's

orbital velocity into each planets' epicycle. The Copernican model's proposed explanation is lovelier than the Ptolemaic model's. Its greater loveliness was regarded (e.g., by Kepler) as supporting Copernicus' theory.[10]

Let's see how the refined version of the approach pursued by Lipton, McGrew, and Okasha accounts for the role in confirmation that is played by the Copernican model's greater loveliness. Let C be the Copernican model and P be the Ptolemaic model without any specification of the angular velocities of the various circles. Let E be the opposition/perigee correlation. As we just saw, pr(E|C) = 1 because E follows from the bare geometry of the Copernican model; that model portrays E as independent of the angular velocities. By contrast, pr(E|P) is very low since in the Ptolemaic model, E requires a very special combination of values for the angular velocities. So the Copernican model's greater loveliness contributes to its having greater likelihood, in accord with the idea (that we refined from Lipton, McGrew, and Okasha) that greater loveliness in every case makes some contribution to greater prior or greater likelihood.

Now let F be the Fine-tuned Ptolemaic model with just the right combination of angular velocities to entail E. Since pr(E|F) = 1 = pr(E|C), the Copernican model's greater loveliness does not contribute to its having greater likelihood than F. However, the range of angular velocities that would produce E is so narrow that presumably, even if pr(P) is large, pr(F) is small. Thus, the Copernican model's greater loveliness contributes toward making pr(C) exceed pr(F), in accord with the idea (from Lipton, McGrew, and Okasha) that greater loveliness in every case makes some contribution to greater prior or greater likelihood.[11]

So according to this idea, the Copernican model's having greater loveliness than P is manifested in a contribution to its having greater likelihood, whereas the Copernican model's having greater loveliness than F is manifested in a contribution to its having greater prior probability. *But P and F are unlovely for the very same reason*: to explain E, the orbital parameters in P and F must be fine-tuned in a special, suspicious way. P and F depict as coincidental various facts that are not plausibly coincidental: that the different planets' epicycles have the same angular velocity as the Sun on its orbit (and hence as one another). P and F have the same weakness as potential explanations of the opposition/perigee correlation. What makes P unlovely in its explanation of E is not that E is unlikely given P or that E follows from P only for a narrow range of P's adjustable parameters (features that F does not share with P). Rather, what makes P unlovely is that the explanation it offers depends on values of the orbital velocities that are (for the reasons I just mentioned) suspicious – a feature that F shares with P.[12] But despite F and P having exactly the same unlovely feature, the approach pursued by Lipton, McGrew, and Okasha says that P's unloveliness has an impact on confirmation through likelihoods, whereas F's unloveliness makes itself felt through priors. What IBE correctly regards as the very same explanatory defect is treated differently in P's case than in F's case by the approach pursued by Lipton, McGrew, and Okasha.[13] In treating these cases differently, this approach mischaracterizes how explanatory quality functions as an epistemic reason.[14]

This approach is susceptible to this problem because it does not assign any confirmatory significance to explanatory quality *per se*. Rather, whatever confirmatory significance is possessed by explanatory quality must go through either the priors or the likelihoods. Because explanatoriness on this view has no confirmatory significance in itself, this view has the opportunity to treat very differently two cases that involve the same explanatory considerations. In short, while this view aims to incorporate IBE within BC, explanation itself ends up going missing, and its disappearance has unfortunate consequences.[15]

Admittedly, for both P and F, the approach I am criticizing finds a factor in BC that is low: the likelihood in P's case and the prior in F's case. It might seem, then, that there is no problem;

the approach covers both P and F, correctly taking their low explanatory quality regarding E as counting against them by comparison to C. The approach I am criticizing gets the right answer. So what is the problem with it?

The problem, I am arguing, is that an adequate account of how explanatory quality makes a difference to confirmation should not merely get the right answer, but also get it for the right reason. That is, it should correctly identify where explanatory quality enters into confirmation. An account of how explanatory quality contributes to confirmation is better (all other things equal) insofar as it treats like cases alike. The account given by Lipton, McGrew, and Okasha gets the right answer in the Copernican example, but only by assigning the very same explanatory considerations different roles in cases where IBE regards them as playing the very same role (since P and F share the same source of unloveliness). This approach therefore fails to capture IBE and, more importantly, loses the insight into our reasoning that its reconstruction in terms of IBE is supposed to provide. A Bayesian account that represented explanatory quality as having an impact in the same place for both P and F would in this respect have an advantage over the account I am criticizing. (I will propose such a Bayesian account in the next section.)

Let me clarify this reply to the charge that there is nothing wrong with the account I am criticizing because that account gets the right answer for both P and F. My reply is not that there is something automatically problematic about a given confirmatory consideration being represented in BC by different factors in different cases. Indeed a Bayesian approach can adequately account for scientists' appeal to a given epistemic virtue even if the Bayesian approach finds no single factor in BC that embodies that virtue in every case where scientists properly appeal to it. For instance, Sober (1990) argues that if parsimony determines which phylogenetic hypothesis "is best supported by the data, this will be because the most parsimonious hypothesis is the hypothesis of maximum likelihood," i.e., maximum pr(E|H) (Sober, 1990:89), whereas if parsimony determines that individual selection is better supported by the data than group selection, then that is because "the prior probability of a group selection hypothesis is lower than the prior probability of a hypothesis of individual selection" (Sober, 1990:83). There is nothing implausible about parsimony's confirmatory impact receiving a different rational reconstruction in these two cases.

But that is because in these two cases, the background beliefs that make parsimony relevant to confirmation in one of these cases are unrelated to the background beliefs that make parsimony relevant to confirmation in the other case. In the case of group versus individual selection, the background beliefs concern the frequencies with which populations have various structures (involving the amounts of variation and the rates of colonization and extinction). In the case of phylogenetic inference, by contrast, the relevant background belief concerns the rate at which mutations occur. Hence, there is no particular rationale for giving the same account of parsimony's impact in these two cases. Matters are entirely otherwise for P and F: the same background beliefs are responsible for our regarding P and F as giving unlovely explanations of the opposition/perigee correlation. P and F have the same trouble explaining this correlation because their putative explanations exhibit the very same unlovely features: suspicious fine-tuning involving a mysterious incorporation of the Sun's orbital velocity into each planet's epicycle without any common cause of its appearance in each planet's model. The same explanatory deficiency must be treated as different epistemic reasons by the approach we refined from Lipton, McGrew, and Okasha.[16]

My point is not that their approach fails because there is a *sui generis* virtue called "explanatoriness" that affects plausibility; for that to be my argument would be too question-begging.[17] Rather, my argument is that the approach from Lipton, McGrew, and Okasha fails because it does

not accord similar treatment to similar cases where one hypothesis is supported over another by virtue of giving a better potential explanation than the other does. Perhaps sometimes one hypothesis T_1 gives a lovelier potential explanation than another hypothesis T_2 for a *different* reason than T_1 gives a lovelier potential explanation than T_3. But whenever T_1 gives a lovelier explanation than T_2 for the *same* reason as T_1 gives a lovelier explanation than T_3 (as is the case for C compared with P and for C compared with F), this similarity should be captured rather than obscured by our account of loveliness's role in confirmation.[18]

Let me summarize the problem with the approach that we refined from Lipton, McGrew, and Okasha. On that approach, the quality of the explanations that a given hypothesis would supply does not count as an epistemic reason for or against believing the hypothesis. Rather, some other consideration (either the likelihood or the prior) is the reason. In the case of F and P, this other consideration results from the suspicious fine-tuning required for the Ptolemaic model to explain the opposition/perigee correlation. That suspicious fine-tuning is also what makes P (or F) unlovely. But its unloveliness is not (on this approach) what counts as an epistemic reason against the hypothesis. The suspicious fine-tuning makes the hypothesis less well confirmed not by virtue of making it unlovely, but rather by virtue of influencing either its likelihood or its prior. The suspicious fine-tuning's making the hypothesis less good as an explainer is (on this approach) an incidental byproduct of its making the hypothesis less well confirmed. Low confirmation and low loveliness, on this view, are joint "effects" of a common "cause." But IBE is the idea that the degree of loveliness *itself* sometimes counts as an epistemic reason. So the approach from Lipton *et al*., to my mind, does not count as "one way Inference to the Best Explanation and Bayesianism may be brought together" (Lipton, 2001b:112) since it effectively removes E from IBE. However, this approach holds that we can still account for scientists' practice of appealing to loveliness. As I quoted Lipton as saying, scientists use loveliness "as a symptom" of likelihood or prior probability (just as one effect of a common cause can be a "barometer" of another). But consequently (I have argued) this approach ends up giving different accounts of two cases involving the very same epistemic reasons for belief (or for degrees of credence). It therefore cannot do justice to the role that appeals to loveliness play in confirmation.

We should portray P's unloveliness and F's unloveliness as making the same difference to confirmation, since P and F have identical explanatory deficiencies as compared to C. We can portray P's and F's unloveliness as making the same difference to confirmation by portraying explanatory quality as *itself* making a difference to confirmation – that is, by putting E back into IBE. I will now propose one way of assimilating IBE to BC while ascribing confirmatory significance to explanatory quality *per se*. This approach works by invoking hypotheses that refer explicitly to scientific explanation and by invoking the agent's background opinions regarding the kind of explanation that the evidence is liable to have.

4 | HOW EXPLANATORY QUALITY MAKES A DIFFERENCE TO CONFIRMATION

I will now identify one way for some evidence to confirm H by virtue of H's capacity, if true, to supply certain kinds of explanations – and I will identify one way for BC to capture this confirmatory role played by explanatory considerations. IBE is thereby incorporated within BC since this confirmation proceeds by virtue of H's explanatory quality *per se*.

Whether a scientist believes some E or merely regards E as a live epistemic possibility, she may have some degree of confidence that if E turns out to obtain, then E has the same kind of

explanation as some known fact G. By discovering G's explanation, then, the scientist may confirm a given hypothesis H by virtue of H's capacity (if true) to supply E (if E obtains) with the same kind of explanation as G has. H's confirmation then proceeds by virtue of H's explanatory quality, not some other feature of which its explanatory quality is a symptom. H is confirmed by virtue of the background opinion that E and G have the same kind of explanation – a background opinion about explanation *per se*.[19]

For instance, suppose that several people suddenly begin to display the same uncommon symptom (such as an upset stomach), and we discover the explanation to be that the symptoms in all of the victims have a common cause: the victims all ate the same tainted food. Suppose the discovery of this particular explanation is among our earliest discoveries of common-cause explanations of this kind. Then if we have on hand another group of similar, as yet unexplained occurrences spread among different individuals, our discovery of the common-tainted-food explanation plausibly confirms (to some degree, in the absence of other considerations) a hypothesis attributing a similar common cause to this other group of occurrences. (Of course, once we have discovered many such common-cause explanations, the discovery of another one will make little confirmatory difference.) Further afield, if the question of whether to adopt C (the Copernican model) or P (and F) is still open, the discovery of the common-tainted-food explanation plausibly confirms C and disconfirms P and F. That is because C purports to identify a common cause (Earth's motion around the Sun) for the different superior planets' displaying the same "symptom" (from a geocentric perspective: an epicycle with a period of about 365 days). In contrast, P and F attribute the planets' all having this symptom to no common cause, and so P and F are disconfirmed by the discovery of the common-tainted-food explanation. Of course, not every known instance of common symptoms arising together turns out to result from common causes. But the existence (and frequency) of cases that do plausibly has a bearing on hypotheses positing that other cases do.[20]

It is C's *explanatory* potential that enables it to be confirmed by the common-causal explanations we discover. To be confirmed in this way, C need not be much like the hypothesis that we discovered to explain the upset stomach. What they have in common, enabling one to confirm the other, is the kind of explanation they would supply (as well as various similarities among the phenomena they would explain). Among the respects in which two explanations may be alike is in positing spatiotemporally local causes or common causes or mechanical causes – or in identifying the facts to be explained as the products of regression to the mean (rather than a common cause) or as independent of certain petty causal details (i.e., as explained "at a certain higher level"). Insofar as we are inclined to expect that the two facts will have similar explanations, discoveries regarding one fact's explanation will tend to bear confirmationwise on hypotheses that would explain the other fact.

As we have seen, IBE's fans have usually tried to specify the characteristics that make a potential explanation better or worse. IBE's defenders then owe us an account of why these characteristics are truth-conducive. On my proposal, by contrast, there is no canonical list of characteristics that automatically help to make one potential explanation lovelier than another. Rather, for different facts E, there are different characteristics where our background knowledge of other explanations gives us some reason to expect that E's explanation will turn out to have these characteristics – and these other explanations thereby tend to confirm any hypotheses that have those characteristics and that would explain E. In this way, the simplicity of (say) a proposed phylogenetic explanation might contribute to its "loveliness" even though regarding a different explanandum, the simplicity of a proposed explanation might detract from its loveliness. This seems to be the case in scientific practice. As Salmon (2001a:81) emphasizes, "in anthropology and sociology. . . an

extremely simple hypothesis may be deemed *implausible* because it is likely to be an oversimplification"; multifactorial explanations tend to be more common in the social sciences. (Likewise, Crick (1988:6,59,138–139) emphasizes that whereas simplicity may be a useful guide in physics, in biology it is often "very misleading" because a "contrived or even ugly" mechanism may well be what natural selection arrived at, "because nature could only build on what was already there."[21]) Whether a given feature makes lovelier a given hypothesis that would explain E depends on our background knowledge of the features that E's explanation is likely to have, based on our knowledge of explanations of facts belonging to kinds that we believe tend to have explanations like E has. There is then no need for IBE's defenders to offer some general account of why loveliness tends to make a hypothesis more likely.[22]

Nevertheless, this view does not turn IBE into "inference to the likeliest explanation." Again, H_1 can be such that it would (if true) supply a better explanation of E than H_2 would (in that H_1 has more than H_2 of the features that we expect E's explanation to have, based on other explanations we know) and yet all things considered, we justly regard H_2 as more credible.[23] But it remains worth distinguishing the support that a given hypothesis receives by virtue of the way it would explain E from the support that it receives in other ways. The discovery of the common-tainted-food explanation confirms C only because C (like the tainted-food explanation) aims to explain a similarity among apparently unrelated, spatially separated phenomena. If we understood astronomical theories as aiming merely to save the phenomena, then the discovery of the illness's explanation would have no bearing on astronomy.[24]

Of course, a hypothesis' characteristics as a potential explainer enable it to be confirmed by the explanations that we discover of other phenomena only when we have some confidence that those other phenomena have the same kind of explanations as phenomena that the hypothesis would explain. This means that a given quality that the hypothesis would exhibit as an explainer counts in its favor not automatically, but only when there is appropriate evidence (from explanations of other phenomena) and appropriate background opinions (involving some confidence that those other phenomena have explanations of the same kind as the phenomena that the hypothesis would explain). That the confirmatory significance of the hypothesis' qualities as an explainer is beholden to this evidence and background seems to me a point in favor of my proposal: it means that a hypothesis' characteristics as an explainer count in its favor only when there is some empirical ground for them to do so (rather than by some mysterious *a priori* reason for expecting phenomena to have certain kinds of explanations), and explanatory characteristics can have different confirmatory significance in different cases. I believe that when scientists have confirmed hypotheses by reasoning in ways that we reconstruct in terms of IBE, scientists have been guided by their empirical knowledge of the kinds of explanations that hold in other cases they deem similar. For instance, Copernicus famously writes (in Book 1, Chapter 10 of *De Revolutionibus*) that "we thus follow Nature, who producing nothing vain or superfluous, often prefers to endow one cause with many effects." Copernicus's understanding of the kind of explanation that nature "often prefers" clearly must be acquired empirically – from cases as humble as the common-tainted-food case.[25] Only given our experience with other explanations does the Ptolemaic model's incorporation of the Sun's orbital velocity into each superior planet's epicycle qualify as suspicious, ad hoc, mysterious, or "crying out" for explanation – that is, as detracting from the model's loveliness and as suggesting that the opposition/perigee correlation has a different sort of explanation.

However, some philosophers may regard IBE as the view that certain explanatory characteristics weigh automatically and in every case. For example, Weisberg (2009) worries that if any assignment of priors is permitted, then agents are not obliged to attend to explanatory

considerations, contrary to IBE. In particular, Weisberg is concerned that in the absence of any objective constraints on prior probabilities, "[i]t is possible to be very confident that H is a better explanation of E than H' is, and yet have p(H|E) < p(H'|E); this does not contradict the probability axioms or any other standard Bayesian principle" (Weisberg, 2009:131). However, I deny that this possibility conflicts with IBE, since as I have emphasized, I do not think that IBE requires explanatory considerations to override all others.[26]

Nevertheless, Weisberg may press the worry that an agent might recognize that H (if true) would be a better explanation of E than H' (if true) would be and yet regard this fact as contributing nothing toward making H more credible than H'. In response, let's see how under very modest constraints on personal probabilities, BC *requires* that explanatory considerations *per se* have a confirmatory impact (by the mechanism I have been describing). Suppose the epistemic agent already believes that H, if true, would at least help to give E (if true) a certain sort of explanation. (H need not be enough by itself to explain E; it suffices that H would be part of such an explanation and is the only part that the agent has yet to ascertain.) So H entails C (for "Conditional"): that if E is the case, then E has a certain sort of explanation (namely, the kind that H would help to give it). Therefore – under BC, presuming the agent's credence function (pr) to be such that $0 < pr(H), pr(C) < 1$ – the agent's discovery that C would confirm H by some increment.[27] Let X (for "eXplanation") be that a certain other fact G has a certain explanation – the same kind of explanation as H would help to give E. Suppose that the epistemic agent has some confidence that G and E, if they are the case, have the same kind of explanation – enough confidence that the discovery that X justly raises the agent's confidence that C. (Of course, rationality does not require that the agent regard X as confirming C; rationality alone does not suffice to require that an agent regard a hypothesis's capacity to give E the same kind of explanation as G possesses as contributing to the hypothesis's explanatory quality. But the point is to show that *if* an agent is prepared to regard a hypothesis's possession of certain feature as contributing to its explanatory quality, then rationality requires that the agent regard a given hypothesis's possession of this feature as thereby making it more credible.) Suppose further (for simplicity) that besides by confirming C, there is no other possible way for X (or ∼X) to confirm H. In other words (under BC), suppose that the agent's $pr(H|X\&C) = pr(H|C)$ and $pr(H|X\&\sim C) = pr(H|\sim C)$; that is, once C (or ∼C) has been ascertained, X makes no confirmatory difference to H (and similarly for ∼X). Under these conditions, BC says that the agent must regard X as confirming H. (I relegate the proof to an endnote.[28])

This confirmation depends upon H's potential explanatory power regarding E. Its dependence is demonstrated by the fact that the agent cannot regard X as confirming H*, where H* is like H in all of the above respects except that it is not confirmed by C; in particular, H* does not entail C because H* would not give E the given sort of explanation – perhaps because it would not help to give E any sort of explanation.[29]

On this account, then, a rational agent cannot recognize that H would give E a better explanation than H' would and yet fail to regard this fact as favoring H over H'. A rational agent could still fail to regard a given feature as contributing to the explanatory quality of H – but once having recognized that feature as doing so, she would have to regard that feature as having confirmatory significance. On this account, a hypothesis can be supported or disconfirmed by virtue of its character *as an explainer*. Its explanatory quality can be a distinct epistemic reason for or against having some degree of belief in it. In this way, IBE is integrated with BC. This approach shares with van Fraassen's "bonus points" approach the idea that for IBE to be vindicated, explanatory quality must make a difference to confirmation over and above whatever confirmation a hypothesis receives by virtue of its predictive success. But unlike van Fraassen's construal of IBE, my account accomplishes this without resorting to "bonus points" that are not awarded through BC.

Van Fraassen is correct that to vindicate IBE, an account must accord some epistemic significance to explanatory quality *per se*. But van Fraassen is mistaken in thinking that in order for explanatory quality *per se* to make a difference to confirmation, BC must be violated. By the same token, Salmon is mistaken in thinking that for IBE to add something to BC, rival hypotheses must be compared with respect to their explanatory virtues alone, and the most explanatorily virtuous rival must then be accorded the highest posterior probability:

> I do not believe that archaeologists set up three different explanations, comparing them with respect to their explanatory virtues without regard for the truth of the premises, and then say that the premises of the best are most likely to be true. … From a Bayesian standpoint, this procedure (IBE) simply attacks the problem from the wrong end. … Rather, I think, they make a rough estimate of the prior probabilities of the hypotheses, and then prefer (but not necessarily adopt), for the purposes of explaining the existence of the artifact, the one with the highest posterior probability. (Salmon, 2001a:82–83)

The procedure Salmon identifies with IBE mistakenly treats explanatory considerations as incapable of being overridden. IBE requires only that rival hypotheses' explanatory qualities be among the epistemic reasons for our degrees of confidence in those hypotheses. I have tried to identify one way (perhaps not the only way) for this to be.[30]

Of course, to make explicit (in Bayesian terms) one mechanism by which H's potential explanatory quality *per se* makes a difference to its confirmation, I have invoked hypotheses that refer explicitly to scientific explanation, such as C: that E has a certain sort of explanation (if E obtains). Although work in Bayesian confirmation theory does not typically appeal to hypotheses that make explicit references to explanation, such hypotheses should not be regarded suspiciously. After all, science presumably confirms such hypotheses in exactly the same way as it confirms hypotheses of other kinds. For instance, a physician might justly take certain features of a newly discovered disease, combined with her knowledge that all of the many other diseases known to possess those features are explained by infectious microbes, as confirming that the newly discovered disease is explained in the same way; this confirmation would motivate a search for the microbe responsible for the disease. Bayesian confirmation theory ought to give greater attention to hypotheses of the form "H is explained by…".[31] At the very least, no harm should result from plugging into the Bayesian apparatus various degrees of confidence in hypotheses of the form "H is explained by…" – just as the apparatus can treat agents who assign credence not simply to H, but also to the hypothesis "It is a law that H" (or "H is true but not a law"). In the next section, I will pursue this parallel.

On my account, that a hypothesis is "lovely" – that it possesses the features that we believe E's explanation likely to have – cannot be a matter of indifference confirmationwise. There is no room for an "open question" here; an agent cannot rationally recognize H as explanatorily lovely regarding E and yet not be committed to regarding its loveliness as favoring its truth. A hypothesis's meeting our expectations in various ways always favors its truth.

5 | CONTACTING THE NOMIC

My account of IBE can help us understand the confirmation of hypotheses positing laws of nature. Earman and Roberts (2005) argue roughly that if the laws of nature fail to supervene on the

Humean base, then we could never be epistemically justified on empirical grounds in believing certain facts to be laws (or in believing them not to be laws). Their argument is that the evidence (the facts believed noninferentially in response to observation) must belong to the Humean base, and if the Humean base does not determine which facts are laws, then that evidence could never support theory T (which posits L's lawhood) over theory T* (which holds that L is true but not a law, and is otherwise the same as T) or vice versa. T and T* agree regarding the Humean base but disagree about the laws.

However, T can differ from T* in explanatory quality. For instance, in holding that it is a law that the universe's net electric charge is conserved, T supplies an explanation of the fact that we have never managed to change the universe's net charge: it *has* never changed because it *cannot* change. That charge *must* be conserved explains why it *has* been conserved.[32] Because T* denies that charge conservation is naturally necessary, it cannot give T's explanation of our failure to alter the net charge. Even if T* gives some other explanation of our failure (e.g., by citing initial conditions along with other putative laws (Earman & Roberts, 2005:257)), T* cannot offer the same kind of explanation as T. According to T, the failures of all of our various, diverse attempts to create or to destroy charge have a single explainer that makes the explanandum (naturally) necessary. My account shows how T's capacity to offer a better explanation than T* may support T over T*; T may be better confirmed by our discoveries of other cases in which we conclude that our repeated failure to bring something about is explained by the task's impossibility.[33] In this way, T's explanatory superiority over T* could be an epistemic reason favoring T over T* even if Humean Supervenience is false.

Of course, these explanatory considerations favoring T over T* could be outweighed by other considerations. Moreover, although in the charge-conservation example, T arguably would give a better explanation than T*, it obviously is not the case that for *every* m, "It is a law that m" gives a better explanation of m than some theory that would explain m by citing initial conditions as well as laws (or in some other way). For it to be plausible that there is some law of nature prohibiting any change to the net charge, our failures to change the net charge have to be sufficiently many and diverse that we have some confidence that these failures have the same kind of explanation as other cases where we have concluded that we failed because success was impossible.

This suggests an objection that Earman and Roberts could make to my argument: any other case where we discovered that success had been impossible must be a case where we discovered a natural law precluding success. To posit such a discovery begs the question against Earman and Roberts, since they are arguing that (if Humean Supervenience is false) we are never epistemically justified on empirical grounds in believing certain facts to be laws.

My response to this objection is that my proposal does not presuppose that in every other case where we have discovered the explanation of our failures to be that success had been impossible, our explanation posited a law of nature making success impossible. There are many cases where the impossibility is stronger than natural impossibility. For instance, suppose that despite several attempts, Mother fails to divide her 22 strawberries evenly (without cutting any) among her 3 children. The explanation is that her success is mathematically impossible. The same applies to repeated failures to untie trefoil knots or to cross every bridge exactly once in a certain bridge arrangement in Königsberg.[34] With our discovery of these explanations (all involving mathematical necessities), we may have acquired an epistemic reason for believing that our similarly repeated failure to change the universe's net charge is also explained by the impossibility of success. But the creation of charge is not *mathematically* impossible. In this way, our evidence could give us an empirical epistemic reason to posit a variety of necessity unrepresented in our previous explanations: natural necessity.[35]

I have just proposed a genealogy of natural necessity – or a rational reconstruction of our discovery of natural necessity. By IBE, we discovered that the world contains a variety of necessity that is weaker than mathematical necessity.[36]

6 | CONCLUSION

Lipton (2001a:56) correctly says, "According to Inference to the Best Explanation…[i]t is not just that the observations support the hypothesis that explains them; it is precisely because that hypothesis would explain the observations" (and, I would add, how well it would explain them) "that they support it." We saw in section 3 that if we sacrifice this idea in trying to reconcile IBE with BC, then we lose some of the insight into scientific reasoning that IBE should provide. On the account in section 4, the confirming evidence and the background opinions powering that confirmation can essentially concern explanations, thereby vindicating IBE. This account permits explanatoriness *per se* to count in favor of a hypothesis, putting E back into IBE without awarding good explainers any of the BC-violating "bonus points" that van Fraassen believes IBE requires.[37]

ENDNOTES

[1] I join many fans of IBE in finding it impossible to resist quoting Darwin's remark; Darwin (1873) appeals repeatedly to his theory's explanatory advantages. Lycan (2002) surveys philosophical work on IBE; Dellsen (2018:1745–1746) nicely botanizes more recent approaches. Perhaps a slightly more accurate name would be "inference to the best *potential* explanation" since many of the rival candidates turn out not to be genuine explanations. IBE involves (very roughly speaking) inference to the hypothesis that *would, if true*, do the best job at explaining the evidence – better, that is, than any other hypothesis that would explain the evidence if that hypothesis were true. What makes one potential explanation "better" than another is not its likeliness of being true, but rather the quality of the explanation it would supply, if it were true. This will come up again (and again).

[2] Some defenders of IBE, such as Douven (2013), question whether BC and its generalizations (such as Jeffrey conditionalization) are the only rational rules for updating opinions. Even if they are, I will argue, IBE can be vindicated.

[3] In the next section, I will say more about how an assessment of explanatory quality functions as a "guide", according to Lipton.

[4] Dellsen (2018) argues that IBE typically plays a more modest role with respect to BC than many IBE fans have sought, namely, as guiding us to judgments of hypotheses' *comparative* plausibility (in particular, as guiding us to the most probable of the hypotheses under consideration, which should then be further investigated), not as any guide to hypotheses' *absolute* probability values. Although I cannot here assess Dellsen's various arguments for this view, it seems to me both too strong and too weak: too strong in that other considerations besides explanatory ones should influence even our comparative plausibility judgments among hypotheses, and too weak in that explanatory considerations often rightly have large influences on even absolute (not merely comparative) credences, sometimes supplying a key rationale for accepting hypotheses. This last is emphasized by Darwin in the passage quoted above, and I will elaborate it in section 4. Dellsen might reply (as he suggests on p. 1759) that even when explanatory considerations cannot guide us to absolute credences, other means may exist for arriving at those credences. He says (p. 1760) that the Darwinian case is atypical in allowing explanatory considerations to access absolute credences in that all alternatives to Darwin's explanation of the evidence are poor. But Darwin (in the quoted passage) maintains that cases where explanatory considerations provided a key rationale for accepting hypotheses are common in science and everyday life rather than atypical (as borne out by the example of Kepler on Copernicanism – a case that I will discuss below).

[5] I accept this point only provisionally. Lycan (2002:421–422) argues that an unrepentant proponent of IBE should reject the demand for an account of why various explanatory virtues should be regarded as guides to the truth. In section 4, I will (for a different reason from Lycan's) deny that IBE's defenders owe us a general account of why possession of various explanatory virtues tends to make a hypothesis more likely.

[6] Perhaps for this strategy to account within BC for scientists using loveliness as a guide to truth, we do not even need H's high loveliness to be *sufficient* to ensure high pr(H) or high pr(E|H). Perhaps it would be enough that *in the vast majority of cases*, H's high loveliness is accompanied by either high pr(H) or high pr(E|H), as long as loveliness is epistemically more accessible than these factors. I believe that this is McGrew's view. As McGrew says in acknowledging that high pr(H) or high pr(E|H) can occur outside of explanatory contexts: "It is sufficient that the virtues are frequently, perhaps overwhelmingly frequently, exhibited in inferences of an overtly explanatory sort and that our attention is drawn to them precisely because of the role they play in explanatory reasoning" (McGrew, 2003:565). The arguments I give below that H's high loveliness is not always accompanied by either high pr(H) or high pr(E|H) suggest that it is also not true that in the vast majority of cases, H's high loveliness is accompanied by either high pr(H) or high pr(E|H).

[7] That this is Okasha's intention is suggested a bit by his remark (2000:705n.16) that a certain case is no counterexample to his view because the hypothesis there is not an unlovely explainer but rather no explainer at all. For defense of a similar restriction on IBE, see Dellsen (2018:1758).

[8] Thanks to Wayne Myrvold for this reference.

[9] Throughout this paper, I set aside the problem of old evidence for BC, just as do Okasha (2000:705) and the other philosophers I am discussing. Accordingly, the fact that E is old evidence does not make pr(E|P) equal to 1. I have characterized IBE in terms of explanatory quality serving as an epistemic reason for greater confidence. I have argued elsewhere (Lange, 1999) that talk of such reasons should be understood *not* in terms of a diachronic updating of opinions (allowing the problem of old evidence to arise), but rather in terms of the role that these reasons play in *justifying* one's current opinions (where justifying these opinions need not involve retracing their etiology). On my view, we justify these opinions by beginning with a probability distribution that is sufficiently neutral in the given context to require no justification itself, and by then bringing the individual reasons (e.g., pieces of evidence, including "old evidence") to bear individually, in each case "updating" the probability assignments in accordance with BC (or a generalization thereof). Pieces of old evidence can thus be used as reasons for our current opinions.

[10] Likewise, in order to explain why every inferior planet (Mercury and Venus) remains near to the Sun in the Earth's sky, the Ptolemaic model requires that the line from the Earth to the Sun be parallel to the line from the Earth to the center of an inferior planet's epicycle. This, then, is a further instance of fine-tuning, of suspicious correlation among different planets, and of coincidental incorporation of the Sun's orbital velocity into each planet's orbital behavior. This unloveliness is avoided on the Copernican account: each inferior planet remains near the Sun in Earth's sky simply because the inferior planets' orbits lie entirely within Earth's orbit.

[11] There have been many Bayesian discussions (e.g., Huemer, 2009:224–225) comparing rival hypotheses with different numbers of adjustable parameters – and there it has been widely noted that the likelihoods of these rival hypotheses become equal (to 1) if the hypotheses are augmented with the parameters' values set precisely to enable the hypotheses to entail the evidence. For a recent discussion (with some earlier references) see Sober, 2015*b*:124–125. (On pp. 17–21, Sober gives the Copernican example, citing in turn one of my earlier discussions of it.)

[12] What makes P unlovely is *not* that the explanation it offers depends on values of the orbital velocities that are *unlikely* given P (i.e., is not that only a narrow range of the orbital velocities will allow P to yield E). Rather, what makes P unlovely is that the requisite values of the orbital velocities are (for the reasons I just mentioned) *suspicious*. To explain E, P requires not merely certain particular orbital velocities, but very special orbital velocities. (To see the difference between unlikely values and suspicious values, suppose that I toss a pencil casually onto the floor. Anyplace it lands was a unlikely place for it to have landed. But if I had marked an "X" on the floor before tossing the pencil, then its landing on that spot is not merely unlikely; it is suspicious.) *Any* required fine-tuning of orbital parameters would make pr(F) small – even if the requisite fine-tuning did not involve *suspicious* values of the orbital parameters. But only fine-tuned values that are not plausibly coincidental make F (or P) unlovely.

[13] It might be objected that Lipton *et al*. do not treat these cases differently: both the posterior probability of P updated by E and the posterior probability of F updated by E are low (compared with the corresponding probability of C). I reply that the philosophical goal is to *explain why* these posterior probabilities are both low, and the explanations given by Lipton *et al*. are importantly different in the two cases. The agent's justification for low pr(P|E), according to Lipton *et al*., is importantly different from the justification for low pr(F|E).

[14] It might be objected that P, even if true, could not explain E. Rather, P could help to explain E only after P is supplemented by auxiliary hypotheses concerning the orbital velocities. But then the explainer is effectively F, and so we no longer have two cases, P and F. To this objection, I reply, first of all, that P alone is at least a partial explanation of E; in certain contexts, it would even constitute a complete explanation. (After all, even F in certain contexts is merely a partial explanation – namely, in those contexts where a complete explanation would have to specify why the orbital velocities have the particular values in F or why the planets circle the Earth.) Secondly, I reply that even if P would not by itself potentially explain E, P could be turned into a potential explainer by being supplemented with auxiliary hypotheses that do *not* turn P into F – that is, by being supplemented with auxiliary hypothesis that do not allow the explainer to entail E. Suppose we supplement P with the auxiliary hypothesis that the orbital velocities were fixed randomly by some physical process that could have produced any values within a suitably wide range ("most" values in which do not yield E), where the physical process is associated with a natural, flat probability density over the possible values of the orbital velocities. This strengthened hypothesis P*, if true, would give a statistical explanation of E (at least, according to those philosophers who do not require that the explainers in a statistical explanation make the fact being explained highly likely). Nevertheless, pr(E|P*) is very low. The Copernican model's having greater loveliness than P* is again (on the view I am criticizing) reflected in C's greater likelihood, not in its greater prior. Yet P* and F have the same unloveliness since in order to yield E, they require the same suspicious values of the angular velocities. They do not give the same explanation of E, but their explanations share the same source of unloveliness – and any Bayesian reconstruction of IBE should capture this similarity.

[15] Of the approach to reconciling IBE to Bayesianism that I am criticizing, Douven and Schupbach (2015:299) say that it "provides a gloss on, rather than alternative to, Bayesianism." I agree with this characterization (though their paper offers little by way of argument for it). Henderson (2014) also investigates the Copernican model's explanatory advantages over the Ptolemaic model. But her aim is to argue for precisely the supposed reconciliation between IBE and Bayesianism that I am criticizing: that these explanatory advantages *per se* do not count as epistemic reasons favoring the Copernican model over the Ptolemaic, but "a Bayesian who adopts constraints on her probabilities, which are reasonable on her own terms, would end up favouring more explanatory theories" (698–699) as a result of their automatically having higher likelihoods or priors.

[16] Here is an argument that it is perfectly appropriate for a Bayesian reconstruction of IBE to treat F and P differently. Bird (2017:100–104,113) distinguishes between the "internal" and the "external explanatory virtues" of a hypothesis. Internal virtues (such as simplicity and unification) reflect the hypothesis alone and so are captured by pr(H). External virtues reflect H's relationship to E and so are captured by pr(E|H)/pr(E). Armed with this distinction, we might expect that since F and P are distinct hypotheses, an explanatory consideration that is internal to one could appropriately be external to the other. Though the internal/external distinction may sound natural, the argument I have been making shows that it is highly artificial in the Copernican example. F and P have exactly the same explanatory deficiency regarding E (the opposition/perigee correlation): each relies on (the same) suspicious values of the orbital velocities. That these suspicious values are built into F, whereas P's adjustable parameters must be tuned to those values, makes no difference to the explanatory deficiency, but makes it "internal" to F while "external" to P. An account of this explanatory consideration's bearing on confirmation would be better if it accorded the same treatment to this consideration for both F and P. The internal/external distinction obscures the fact that F and P suffer exactly the same deficiency as explainers of E. (For additional argument against Bird's internal/external distinction, see note 30.)

[17] Sober (1990:76) writes: "Notice that the Bayesian biconditional [pr(H1/E) > pr(H2/E) iff pr(E/H1)pr(H1) > pr(E/H2)pr(H2)] does not use the word 'explanation'. Explanations have likelihoods; and sometimes they even have priors. This means that they can be evaluated for their overall plausibility. But there is no *sui generis* virtue called 'explanatoriness' that affects plausibility."

[18] The Copernican example is not unusual; the same argument applies to any number of other examples. For instance, Salmon (1984) considers two students who turn in word-for-word identical papers. The analogue to C is that they collaborated. The analogue to P is that they worked independently. The analogue to F is that they worked independently and here…are the specific initial conditions (including the initial states of the students' brains), where these conditions lead by law to each student's writing the same given sequence of words. Whereas P's analogue has a healthy prior, it has a low likelihood; F's analogue has a high likelihood but low prior. They have the same source of unloveliness.

[19] That a potential explainer can be confirmed in this way has been noted before. For instance, Thagard (1978:91) says that "analogy between phenomena suggest[s] the existence of analogy between explanatory hypotheses", but he regards this as just one of the explanatory virtues (and presents it as a rule of inference). Thagard (1978:90), in turn, cites Achinstein (1971:132–133) on "analogical-explanatory inference". Sober (2015b:280) briefly interprets one step of Reichenbach's cubical-world argument for the existence of the "external world" as appealing to an established scientific explanation as confirming a hypothesis that would supply a similar explanation of a similar phenomenon.

[20] Achinstein (2013) argues that for any allegedly canonical explanatory virtue, the history of science has not shown that theories with it have turned out to be true more often than theories without it: "As Larry Laudan has emphasized…, if you look at the success of theories historically, using any criterion of goodness, including "loveliness," the induction will be pessimistic" (Achinstein, 2013:106). But I am not proposing that over the entire history of science, theories positing (e.g.) common causes have been more successful than other theories. Rather, I am proposing that scientists sometimes use a different, narrower line of reasoning: that for some E, most phenomena that have been explained and that are known to share certain characteristics with E have been found to have common-causal explanations, suggesting that E does, too. (The data could even be more modest: that phenomena that have been explained and are known to share certain characteristics with E have been found to have common-causal explanations more often than phenomena do generally.) For instance (recalling the example from note 18), in my experience, most cases where two students turned in word-for-word identical papers are cases where the two papers have important common causes.

[21] Likewise, the pioneering meteorologist Lewis Fry Richardson wrote: "Einstein has somewhere remarked that he was guided towards his discoveries by the notion that the important laws of physics were really simple. R.H. Fowler has been heard to remark that, of two formulae, the more elegant is the more likely to be true. Dirac sought an explanation alternative to that of spin in the electron because he felt that Nature could not have arranged it in so complicated a way. These mathematicians have been brilliantly successful in dealing with mass-points and point-charges. If they would condescend to attend to meteorology the subject might be greatly enriched. But I suspect that they would have to abandon the idea that truth is really simple." (Ashford, 1985:124)

[22] If Salmon is correct about explanations in anthropology and sociology, then it is a consequence of my view that a proposed explanation in those fields is often "lovelier" insofar as it is *less* simple. It might seem a bizarre usage for "loveliness" to be associated with multifactorial complexity. But bear in mind that "loveliness" is not a term from scientific practice; rather, it is Lipton's term for explanatory quality. "Good explanation" is a term from scientific practice, and it *is* realistic to expect an anthropologist or sociologist to say of a proposal that it would make a "good explanation" of some phenomenon partly because it does not seem like an oversimplification, but rather is properly multifactorial. Likewise, if early-modern astronomers had thought that the superior planets' all displaying the same "symptom" probably has the same sort of explanation as certain other phenomena, and if those other phenomena had already been explained as coincidences, then early-modern astronomers should have judged the Ptolemaic model as giving a good potential explanation and as confirmed (to some degree) by those explanations that had already been discovered. But I presume that enough "suspicious" correlations had already turned out to be non-coincidental that astronomers should have regarded the Copernican model as giving the far better explanation.

[23] Because explanatory considerations can be outweighed by other considerations, a hypothesis that would give explanations of a novel kind (and so receiving no support from other, previously discovered explanations) can nevertheless be best supported all things considered. It can then, in turn, support further hypotheses of the new kind. So my account faces no regress problem; it does not entail that the adoption of one lovely explanation requires that other, similar lovely explanations have already been adopted.

[24] One hypothesis can be more credible than a rival partly by virtue of supplying a better explanation *and* partly by virtue of other evidence favoring it. For instance, Lipton (2004:136) discusses the seventeenth-century hypothesis that a sword wound can be made to heal by "sympathetic powder" sprinkled on the sword after it caused the wound. A rival hypothesis to explain why wounds left untreated but having the powder sprinkled on the sword healed more quickly than treated wounds is that seventeenth-century unsterile "treatments" infected the wounds; wounds left untreated were less prone to infection. The latter hypothesis is favored over the former by IBE (based on other explanations, we expect healing to be explained locally, not by action at a distance) and, as Lipton notes, it would also be favored by other evidence (such as that we would see no difference in healing between "sympathetic powder" treatment and no treatment at all).

[25] An IBE defender who (following Lipton) sees explanatory considerations as playing merely a heuristic role (in that H's loveliness is often more epistemically accessible to an ordinary epistemic agent than her own pr(H) and pr(E|H)) need not regard certain characteristics as known *a priori* to enhance explanatory quality in every scientific theory exhibiting them. She could instead take those characteristics as I do: as known empirically to enhance or to detract from explanatory quality differently for different theories. But as long as the IBE defender treats explanatory considerations as mere heuristics that an agent can use for accessing personal priors and likelihoods, the IBE defender cannot use BC to understand the way that explanatory considerations operate as evidence that confirms or disconfirms a potentially explanatory hypothesis. That is, only by ceasing to view explanatory considerations as mere heuristics can an IBE defender use BC (as I am about to do) to understand the way that the empirical evidence of other scientific explanations confirms that certain of H's characteristics enhance the loveliness of H's potential explanations and thereby enhance H's plausibility. Only by departing from the heuristic view can IBE's defender treat the evidence of other scientific explanations (in supporting a given potential explanation's loveliness) as playing the same role as other empirical evidence does in confirming explanatory theories.

[26] The possibility that Weisberg identifies is not worrisome. Suppose that H is that *m* is a law whereas H' is that *m* is true. These are obviously not rivals but some explanations appealing to H may not work if H is replaced by H'. Nevertheless, since H entails H' but not vice versa, it may well be that pr(H|E) < pr(H'|E).

[27] That C's discovery would confirm H follows from Bayesian Conditionalization in a familiar way: pr(H|C) = pr(H) cr(C|H)/pr(C), but pr(C|H) = 1 since H entails C whereas 0 < pr(C) < 1, so pr(H|C) > pr(H).

[28] That X's discovery would confirm H follows from Bayesian Conditionalization as follows:

By the theorem of total probability,

pr(H&X) = pr(H&X&C) + pr(H&X&~C).

Hence, by the definition of conditional probability,

pr(H&X) = pr(H|X&C) pr(X&C) + pr(H|X&~C) pr(X&~C).

We have supposed that other than by confirming C, there is no other possible way for X to confirm H – in other words, that pr(H|X&C) = pr(H|C) and likewise pr(H|X&~C) = pr(H|~C). Therefore,

pr(H&X) = pr(H|C) pr(X&C) + pr(H|~C) pr(X&~C).

By the definition of conditional probability,

pr(H&X) = pr(H|C) pr(C|X) pr(X) + pr(H|~C) pr(~C|X) cr(X).

So

pr(H&X)/pr(X) = pr(H|C) pr(C|X) + pr(H|~C) pr(~C|X).

By the definition of conditional probability,

pr(H|X) = pr(H|C) pr(C|X) + pr(H|~C) pr(~C|X).

By analogous reasoning,

pr(H|~X) = pr(H|C) pr(C|~X) + pr(H|~C) pr(~C|~X).

Subtracting the previous two equations, we find

pr(H|X) – pr(H|~X) = pr(H|C) [pr(C|X) – pr(C|~X)] + pr(H|~C) [pr(~C|X) – pr(~C|~X)].

Since pr(~C|X) = 1 – pr(C|X) and pr(~C|~X) = 1 – pr(C|~X), it follows that [pr(~C|X) – pr(~C|~X)] = [pr(C|~X) – pr(C|X)]. Hence,

pr(H|X) – pr(H|~X) = pr(H|C) [pr(C|X) – pr(C|~X)] + pr(H|~C) [pr(C|~X) – pr(C|X)]
= [pr(H|C) – pr(H|~C)] [pr(C|X) – pr(C|~X)].

Therefore, since C would incrementally confirm H ([pr(H|C) – pr(H|~C)] > 0) and X would incrementally confirm C ([pr(C|X) – pr(C|~X)] > 0), it follows that X would incrementally confirm H ([pr(H|X) – pr(H|~X)] > 0). After writing this paper, I noticed that Sober (2015b: 249) appeals to the same "transitivity" principle in evaluating one version of the argument from evil and the same result is proved by Shogenji (2003:614-615). It may appear elsewhere as well.

[29] If "H" is replaced by "H*" in the argument in the previous note, then since C would not incrementally confirm H* ([pr(H*|C) – pr(H*|~C)] ≤ 0), it follows from the conclusion of that argument that X would not incrementally confirm H* ([pr(H*|X) – pr(H*|~X)] ≤ 0).

[30] As on my proposal, Bird (2017:112) sees a hypothesis's loveliness as evaluated in light of evidence rather than *a priori*. Accordingly, Bird (2017:115) likewise sees the explanatory virtues as able to vary with the scientific field. However, Bird regards some of these evaluations as not being captured by Bayesian conditionalization on evidence (and so sees these explanatory virtues as "internal" (i.e., "evidence-independent") rather than "external" –

see note 16 above). By contrast, I see no reason why these evaluations cannot be captured by the kind of Bayesian updating that I have just described, where the evidence on which H is being conditionalized consists of the explanations we discover in accepting other theories. For example, regarding Coulomb's proposed inverse-square law of electrostatics, Bird (2017:113) says that "even before the torsion balance experiments that he carried out, this hypothesis carried a high degree of (evidence-independent) plausibility because of its analogy to the inverse square law of gravitation. This is evidence-independent plausibility in that it is not the product of any process of conditionalization (e.g., on the law of gravitation)." By contrast, I see this contribution to the plausibility of Coulomb's hypothesis as captured by conditionalization on evidence in the manner I have just described. In particular, Coulomb's hypothesis gained some plausibility precisely from its potential to provide explanations similar to those provided by Newton's gravitational-force law. This confirmation of Coulomb's hypothesis (H), prior to the torsion balance experiments, can be understood in terms of conditionalization on X (that various facts are explained by Newton's law – explanations of the same sort as H would give some possible results E of future torsion balance experiments). It seems advantageous to understand the way that evidence of other explanations underwrites our confidence in a given theory (by virtue of that theory's characteristics as a potential explainer) by using the same means as we use to understand the way that evidence confirms theories generally. Nevertheless, I agree with Bird's (2017:114) general point that "scientists acquire their sense of explanatory virtues from their training with and use of exemplars" (though these exemplars should include newly discovered explanations, not merely explanations that institute a Kuhnian paradigm).

[31] Recently, Hartmann and Fitelson (2015) have appealed to such a hypothesis.

[32] To presuppose that laws are necessary and can explain is not to beg the question against the Humean. On the contrary, Humeans and non-Humeans alike recognize that laws possess these features in scientific practice. For instance, Lewis (1999:232) writes: "If you're prepared to grant that theorems of the best system are rightly called laws, presumably you'll also want to say that they underlie causal explanations; that they support counterfactuals; that they are not mere coincidences; that they and their consequences are in some good sense necessary; and that they may be confirmed by their instances." Of course, many philosophers have argued that Humeans are mistaken in regarding various truths as able to acquire a genuine kind of necessity or power to explain by virtue of their status as theorems of the best system. But this is not the argument that I am concerned with here. I am not arguing for a non-Humean account of law (e.g., as the only way to account for the laws' necessity and explanatory power); rather, I am arguing against one prominent Humean argument against a non-Humean account (i.e., that a non-Humean account renders lawhood epistemically inaccessible), so I am free to adopt the Humean's premises (e.g., that laws are necessary and have explanatory power) for the sake of argument in order to show that even granting these premises, the argument against a non-Humean account does not succeed.

[33] As we have seen, this confirmation requires our having some prior confidence that these other repeated failures, if they occur, have the same kind of explanation as our repeated failures to change the net charge, if they occur. Perhaps Earman and Roberts would say that we are not entitled to such prior opinions (cf. Earman & Roberts, 2005:264). But these opinions are not about laws specifically and so they are not the kind of opinions that Earman and Roberts believe we are never epistemically justified in holding if Humean Supervenience is false. Furthermore, they seem little different from other kinds of prior opinions that underwrite ampliation, and Earman and Roberts are not wholesale inductive skeptics.

[34] Pincock (2007) and Lange (2013, 2017) discuss these and similar explanations. It makes no difference to my argument whether we initially discovered *a priori* (by mathematical proof) that our success was mathematically impossible, or whether we discovered its impossibility empirically (by IBE): we tried so many ways to cross those bridges yet always failed, and we considered whether that failure was from lack of imagination or bad luck, and we eventually concluded that there was a deeper reason.

[35] Of course, there are many differences between mathematical necessity and natural necessity; for instance, mathematical necessities can be discovered *a priori* and mathematical necessity is stronger than natural necessity. But these differences do nothing to keep mathematical and natural necessities from explaining in the same way: by revealing that some fact holds because it had to hold. This similarity between their roles in one kind of explanation is not undermined by the existence of other kinds of explanation in which natural necessities play roles that could not be played by mathematical necessities, or vice versa. Moreover, there are obviously many cases where our failure to bring about some result cannot be explained by the result's impossibility; we justly conclude that we failed despite the fact that success is possible. But insofar as we have tried repeatedly, employing diverse means under seemingly propitious circumstances, and insofar as other kinds of explanation for our

repeated failure are implausible, we may come to be warranted in believing that we failed because success was impossible. (For more on such explanations, see Lange, 2017.)

[36] Again, this account of how we discovered natural necessity does not conflict with a Humean account of natural law. As I mentioned in note 32, the Humean believes that laws of nature possess a kind of necessity (though, of course, the Humean takes their necessity as arising from their possessing some feature involving no irreducibly modal ingredient – on Lewis's account, for instance, from their membership in the best system). Once again, whether the Humean's conception of necessity is substantial enough to do the work required of it is a perennial issue that is not being addressed here. For more on what natural necessity has in common with other varieties of necessity by virtue of which they all qualify as varieties of the same thing, while possessing different strengths, see Lange, 2009.

[37] My thanks to F. Dellsen, B. Kment, M. Kotzen, W.G. Lycan, W. Myrvold, F. Nappo, J. Roberts, and audiences at King's College London; the "Hempel and Beyond" conference at the University of Cologne; the 2016 "Metaphysics at the Ranch" conference; and the inaugural Stephen Humphrey Lecture at the University of California, Santa Barbara.

REFERENCES

Achinstein, P. (1971). *Law and Explanation*. Oxford: Clarendon.

Achinstein, P. (2013). *Evidence and Method*. Oxford: Oxford University Press. Retrieved from http://doi.org/10.1093/acprof:oPo/9780199921850.001.001

Ashford, O. (1985). *Prophet – or Professor? The Life and Work of Lewis Fry Richardson*. Bristol and Boston: Adam Hilger.

Bird, A. (2017). Inference to the best explanation, Bayesianism, and knowledge. In K. McCain & T. Poston (Eds.), *Best Explanations* (pp. 97–120). Oxford: Oxford University Press. Retrieved from http://doi.org/10.1093/oso/9780198746904.003.0007

Crick, F. (1988). *What Mad Pursuit*. New York: Basic Books.

Darwin, C. (1873). *The Origin of Species* (6th ed). London: Murray.

Dellsen, F. (2018). The heuristic conception of inference to the best explanation. *Philosophical Studies*, *175*, 1745–1776. Retrieved from http://doi.org/10.1007/s11098-017-0933-2

Douven, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimisation. *Philosophical Quarterly*, *63*, 428–444. Retrieved from http://doi.org/10.1111/1467-9213.12032

Douven, I., & Schupbach, J. (2015). The role of explanatory considerations in updating. *Cognition*, *142*, 299–311. Retrieved from http://doi.org/10.1016/j.cognition.2015.04.017

Earman, J., & Roberts, J. (2005). Contact with the nomic: A challenge for deniers of Humean supervenience about laws of nature, Part II. *Philosophy and Phenomenological Research*, *71*, 253–286. Retrieved from http://doi.org/10.1111/j.1933-1592.2005.tb00449.x

Gingerich, O. (1975). 'Crisis' vs. aesthetic in the Copernican revolution. In A. Beer & K. Strand (Eds.), *Copernicus: Yesterday and Today* (pp. 85–93). Oxford: Pergamon.

Harman, G. (1965). Inference to the best explanation. *Philosophical Review*, *74*, 88–95.

Hartmann, S., & Fitelson, B. (2015). A new Garber-style solution to the problem of old evidence. *Philosophy of Science*, *82*, 712–717. Retrieved from http://doi.org/10.1086/682916

Henderson, L. (2014). Bayesianism and inference to the best explanation. *British Journal for the Philosophy of Science*, *65*, 687–715. Retrieved from http://doi.org/10.1093/bjps/axt020

Hon, G., & Rakover, S. (2001). *Explanation: Theoretical Approaches and Applications*. Dordrecht: Kluwer. Retrieved from http://doi.org/10.1007/978-94-015-9731-9

Huemer, M. (2009). When is parsimony a virtue? *Philosophical Quarterly*, *59*, 216–236. Retrieved from http://doi.org/10.1111/j.1467-9213.2008.569.x

Jardine, N. (1984). *The Birth of History and Philosophy of Science: Kepler's "A Defense of Tycho Against Ursus"*. Cambridge: Cambridge University Press.

Kepler, J. (1981). *Mysterium Cosmographicum*. New York: Abaris.

Lange, M. (1999). Calibration and the epistemological role of Bayesian conditionalization. *The Journal of Philosophy*, *96*, 294–324. Retrieved from http://doi.org/10.2307/2564680

Lange, M. (2009). *Laws & Lawmakers*. New York: Oxford University Press. Retrieved from http://doi.org/10.1093/acprof:oSo/9780195328134.001.0001

Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *British Journal for the Philosophy of Science*, *64*, 485–511. Retrieved from http://doi.org/10.1093/bjps/axs012

Lange, M. (2017). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. New York: Oxford University Press. Retrieved from http://doi.org/10.1093/acprof:oSo/9780190269487.001.0001

Lewis, D. (1999). Humean supervenience debugged. In D. Lewis (Ed.), *Papers in Metaphysics and Epistemology* (pp. 224–247). Cambridge: Cambridge University Press. Retrieved from http://doi.org/10.1017/cbo9780511625343

Lipton, P. (2001a). *What good is an explanation?* In Hon and Rakover (pp. 43–59). Retrieved from http://doi.org/10.1007/978-94-015-9731-9_2

Lipton, P. (2001b). *Is explanation a guide to inference?* In Hon and Rakover (pp. 93–120). Retrieved from http://doi.org/10.1007/978-94-015-9731-9_4

Lipton, P. (2004). *Inference to the Best Explanation* (2nd ed). Abingdon: Routledge. Retrieved from http://doi.org/10.4324/9780203470855

Lycan, W. (2002). Explanation and epistemology. In Paul Moser (Ed.), *The Oxford Handbook of Epistemology* (pp. 408–433). Oxford: Oxford University Press. Retrieved from http://doi.org/10.1093/0195130057.003.0015

McGrew, T. (2003). Confirmation, heuristics, and explanatory reasoning. *British Journal for the Philosophy of Science*, *54*, 553–567. Retrieved from http://doi.org/10.1093/bjps/54.4.553

Naeye, R. (2000). Ask astro. *Astronomy*, *28*(10). 94.

Okasha, S. (2000). Van Fraassen's critique of inference to the best explanation. *Studies in History and Philosophy of Science*, *31*, 691–710. Retrieved from http://doi.org/10.1093/s0039-3681(00)00016-9

Pincock, C. (2007). A Role for Mathematics in the Physical Sciences. *Noûs*, *41*, 253–275. Retrieved from http://doi.org/10.1111/j.1468-0068.2007.0064.x

Ptolemy, C. (1984). *Ptolemy's Almagest*, G. Toomer (Ed. and trans.). London: Duckworth.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, W. (2001a). *Explanation and confirmation: A Bayesian critique of inference to the best explanation*. In Hon and Rakover (pp. 61–92). Retrieved from http://doi.org/10.1007/978-94-015-9731-9_3

Salmon, W. (2001b). *Reflections of a bashful Bayesian*. In Hon and Rakover (pp. 121–136). Retrieved from http://doi.org/10.1007/978-94-015-9731-9_5

Shogenji, T. (2003). A condition for transitivity in probabilistic support. *British Journal for the Philosophy of Science*, *54*, 613–616. Retrieved from http://doi.org/10.1093/bjps/54.4.613

Sober, E. (1990). Let's razor Ockham's razor. In D. Knowles (Ed.), *Explanation and its Limits* (pp. 75–93). Cambridge: Cambridge University Press. Retrieved from http://doi.org/10.1017/cbo9780511599705.006

Sober, E. (2015a). Two Cornell realisms: Moral and scientific. *Philosophical Studies*, *172*, 905–924. Retrieved from http://doi.org/10.1007/s11098-014-0300-5

Sober, E. (2015b). *Ockham's Razors*. Cambridge: Cambridge University Press. Retrieved from http://doi.org/10.1017/cbo9781107705937

Thagard, P. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy*, *75*, 76–92. Retrieved from http://doi.org/10.2307/2025686

Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Clarendon. Retrieved from http://doi.org/10.1093/0198244274.001.0001

Van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Clarendon. Retrieved from http://doi.org/10.1093/0198248601.001.0001

Weisberg, J. (2009). Locating IBE in the Bayesian framework. *Synthese*, *167*, 125–143. Retrieved from http://doi.org/10.1007/s11229-008-9305-y

How to cite this article: Lange M. Putting explanation back into "inference to the best explanation". *Noûs*. 2022;56:84–109. https://doi.org/10.1111/nous.12349