Explanatory Consolidation: From 'Best' to 'Good Enough'

Finnur Dellsén

Forthcoming in *Philosophy and Phenomenological Research*
(please cite official version when available)

Abstract

In science and everyday life, we often infer that something is true because it would explain some set of facts better than any other hypothesis we can think of. But what if we have reason to believe that there is a better way to explain these facts that we just haven't thought of? Wouldn't that undermine our warrant for believing the best available explanation? Many philosophers have assumed that we can solve such *underconsideration problems* by stipulating that a hypothesis should not only be 'the best' explanation available; rather, it should also be 'good enough'. Unfortunately, however, the only current suggestion for what it might mean to say that an explanation is 'good enough' is, well, not good enough. This paper aims to provide a better account of what is required for an explanatory hypothesis to be considered 'good enough'. In brief, the account holds that a 'good enough' hypothesis is one that has gone through a process that I call *explanatory consolidation*, in which accumulating evidence and failed attempts to formulate better alternatives gradually make it more plausible that the explanation we currently have is better than any other that could be formulated.

1 Introduction

According to the germ theory of disease, many illnesses are caused by unobservably small organisms, such as bacteria and viruses. Early advocates of the germ theory faced a conundrum familiar to many contemporary scientists. Their theory posited a type of entity, 'germs', whose existence could not be directly established by any experimental methods available at the time. So germ theorists took to arguing for their theory indirectly, by showing how the existence of these unobservable organisms would explain a vast array of observable phenomena, including various natural processes such as fermentation and the spread of human diseases such as cholera. As it became increasingly clear that the germ theory provided better explanations of these phenomena than any competing theory that had or would be proposed at the time, scientists eventually became convinced that the posited 'germs' existed and were responsible for many of the most deadly diseases of the time.

This is an example of *explanatory reasoning*, a form of reasoning in which one infers from some evidence to a hypothesis that would, if true, help to explain that evidence. The best known and most popular account of explanatory reasoning is *Inference to the Best Explanation* (IBE).[1] As it is standardly formulated, IBE is a two-stage process in which one first (i) generates a set of hypotheses that would, if true, explain the evidence, and then (ii) infers the hypothesis in this set that provides the 'best' explanation of the evidence. The features that make one hypothesis a 'better' explanation than another are the so-called *explanatory virtues*, such as how many different kinds of things are assumed to exist by the hypothesis (*simplicity/parsimony*), how many different phenomena the hypothesis would explain (*explanatory scope*), to what extent the hypothesis unifies disparate theories or domains (*unificatory power*), and how well the hypothesis accords with other information in one's possession (*fit with background data*).[2] The hypothesis generated in step (i) that has the greatest balance of such virtues is then inferred in step (ii).

A moment's reflection reveals that this cannot be quite right. What if we have reason to think that none of the hypotheses generated so far is correct – that the most explanatory hypothesis is merely 'the best of a bad lot'? This question points to an asymmetry in IBE as it is standardly formulated: While a hypothesis is evaluated by a *comparison* to other available hypotheses, we are meant to conclude with the *absolute* verdict that the hypothesis is (probably, approximately) true. In response to this asymmetry, proponents of IBE have suggested that the evaluative step in IBE not only requires that the inferred hypothesis be the best explanation that has been considered so far, but also that it be 'satisfactory' (Musgrave, 1988) or 'good enough' (Lipton, 1993, 2004; McCain & Poston, 2019). However, no rigorous

[1] I am using 'explanatory reasoning' as a term for the epistemic phenomenon of inferring from an explanandum to an explanans, and 'Inference to the Best Explanation' as the name of a particular account of this epistemic phenomenon. I am making this distinction here to highlight that one could reject that Inference to the Best Explanation is an adequate account without thereby denying that there is such a phenomenon as explanatory reasoning. For example, one might want to give a fully Bayesian account of explanatory reasoning in which IBE plays no role whatsoever.

[2] This is by no means a complete list of the proposed explanatory virtues, but hopefully a somewhat representative one. See Beebe (2009) for a more comprehensive list.

account has so far been provided of what it might be for an explanatory hypothesis to be 'satisfactory' or 'good enough' in the required sense (cf. Douven, 2017a, §2). And as we shall see, the seemingly straightforward idea of setting some threshold of overall explanatory quality for an inferred hypothesis turns out to be implausible, unworkable, and ill-suited to solve the problem. At least in this respect, IBE thus seems in no better shape than when Lipton himself declared that "[i]t is more a slogan than an articulated philosophical theory" (Lipton, 2004, 2).[3]

This paper aims to do better by finally providing an adequate account of what it would be for an explanatory hypothesis to be 'good enough' in the relevant sense. I will suggest that a 'good enough' hypothesis is one that has been through a process that I call *explanatory consolidation*. Since this process necessarily occurs after the inferred hypothesis has been generated and compared to available alternatives, explanatory consolidation can be thought of as the third and final step in an IBE. In science, explanatory consolidation is normally a gradual process in which we accumulate information which make it increasingly plausible that it's impossible to formulate any explanatory hypothesis that is superior to the best one currently available. I will provide an account of the epistemology of explanatory consolidation as a process in which a second-order explanatory hypothesis is evaluated by comparing it to alternatives in terms of their explanatory virtues. So, on the proposed view, explanatory consolidation gradually secures that the best explanation for the fact that some hypothesis is better than all *available* alternatives is that it is better than all *possible* alternatives.

The plan for this paper is as follows. Section 2 introduces IBE and the epistemic and historical problems that a 'good enough' clause on IBE is meant to address. Section 3 argues that the view that a hypothesis is 'good enough' just in case its overall balance of explanatory virtues exceeds some threshold is inadequate for a number of different reasons. Section 4 briefly examines the historical case of how the transverse wave theory of light was received among physicists in the early 18th century, showing how the acceptance of a theory can be significantly delayed even after there is consensus that it provides a better explanation than available alternatives. Section 5 goes on to develop an account of how this delay in accepting a hypothesis can gradually provide reasons to be optimistic that the best available explanation is also the best possible explanation. Section 6 ties up some loose ends

[3]Similar sentiments are often expressed by those who are skeptical of IBE. For example, Norton (2018a, 2) remarks that the elaboration of IBE "is noticeably thin in the literature."

by addressing two related problems concerning the way in which IBE involves an essentially comparative type of evaluation. Section 7 is the conclusion.

2 IBE and Underconsideration Problems

Although the inference form now know as Inference to the Best Explanation (IBE) traces back to Charles S. Peirce's notion of 'Abduction' (1958, 5.189),[4] Gilbert Harman (1965) is generally considered to have provided the first canonical presentation of IBE. In this inference form, says Harman, "one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true" (Harman, 1965, 89). Since one isn't normally in a position to know whether a hypothesis would provide a better explanation than *any* other hypothesis, Harman should presumably be read as suggesting that we compare the relevant hypothesis with any other explanatory hypothesis *that has been formulated* (i.e., that has been considered, generated, or made available). Another well-known proponent of IBE, Peter Lipton (2004), emphasizes this aspect of IBE by suggesting that the process of generating theories is itself a part of IBE. Hence, on Lipton's view, IBE is a "two-stage mechanism involving the generation of candidate hypotheses and then a selection from among them" (Lipton, 2004, 208; see also 59, 148-151). Lipton's two-stage model of IBE is now standard in the philosophical literature on IBE.

Other aspects of IBE are more controversial. Harman (1965, 88) argued that IBE should be viewed as a basic or fundamental form of non-deductive inference – i.e. as not reducible to any other form of inference.[5] Lipton (2001, 2004, ch.7), by contrast, was among the first to suggest that IBE could be thought of as a heuristic for probabilitistic or 'Bayesian' reasoning. The idea is that IBE is a rule of inference that approximates correct probabilistic reasoning in the sense that employing it in one's non-deductive reasoning delivers sufficiently similar results as if one reasoned

[4]With that said, the connection between Peirce's ideas about Abduction and the contemporary notion of IBE is not straightforward (see, e.g., Kapitan, 1992; Hintikka, 1998; Minnameier, 2004; Campos, 2011; Plutynski, 2011).

[5]This view is also endorsed by, e.g., Armstrong (1983); Foster (1982); Lycan (1988); Weintraub (2013); Poston (2014).

as an ideally rational Bayesian agent.[6] Although this paper touches briefly on this difference between 'fundamentalist' and 'heuristic' conceptions of IBE (see section 3), the central thesis and arguments of this paper should be congenial to those on either side of this debate. Another controversial aspect of IBE concerns whether, or the extent to which, the explanatory virtues are truth-conducive, i.e. such that a more virtuous theory is likelier to be true.[7] According to a prominent contrary view, explanatory virtues are merely pragmatic, i.e. such that a more virtuous theory is more convenient to use in prediction, experimentation, model-building, etc. Again, the argument and thesis of the paper should be congenial to both sides in this debate, but to simplify the discussion in what follows I shall formulate IBE as if the goal were truth rather than convenience.[8]

Some of the more influential criticisms of IBE concern the asymmetry between the *comparative evaluation* of a limited number of available explanatory hypotheses and the *absolute verdict* that IBE is meant to warrant, i.e. that one of these hypothesis is true (or at least probably and/or approximately true).[9] A few subtly different versions of this problem have emerged. Van Fraassen's (1989, 142-143) *bad lot objection* famously argued that since the best available explanatory hypotheses might be 'the best of a bad lot', there is no reason to think that a hypothesis inferred by IBE is in fact true. Although not explicitly concerned with IBE, Sklar (1981) had similarly wondered whether the possibility that there are 'unborn hypotheses' that would be taken seriously if brought to the light of day should lead us to be skeptical about the best scientific theories we have at a given time. Stanford (2001, 2006) developed this thought into the *problem of unconceived alternatives*, which states

[6]This heuristic conception of IBE has been fleshed out in a number of different ways (Niiniluoto, 1999; Okasha, 2000; McGrew, 2003; Cabrera, 2017; Bird, 2017; Dellsén, 2018). Another way to conceive of the relationship between IBE and probabilistic reasoning is van Fraassen's idea that IBE would require assigning higher probabilities to more explanatory hypotheses as one gathers new evidence than the Bayesian conditionalization would have one do. Few, if indeed any, proponents of IBE have found van Fraassen's way of locating IBE within a Bayesian framework to be the most plausible way of doing so (see, e.g., Kvanvig, 1994; Harman, 1997; Douven, 1999, 2017b; Schupbach, 2017). Yet other authors see IBE not so much as an inference rule but as a constraint on subjective probability assignments (Huemer, 2009a; Weisberg, 2009; Poston, 2014). Finally, Henderson (2014, 2017) suggests that a preference for more explanatory hypotheses emerges naturally out of the Bayesian framework, given independently plausible constraints on subjective probability assignments.

[7]See, e.g., van Fraassen (1985); Barnes (1995); Bartelborth (2005); Ivanova (2014); Cabrera (2017); Schindler (2018).

[8]So those who consider explanatory virtues to be pragmatic rather than truth-conducive can then simply mentally replace all instances of 'true' with 'convenient' or 'useful' in the discussion below.

[9]To simplify the discussion, I will often omit the parenthetical qualification in what follows.

that there have historically been many cases in which the best hypothesis was in fact unconceived at the time of inference, leading scientists to infer to hypotheses which we would now consider radically false. Since all these problems concern ways in which scientists' failure to *consider* serious alternatives to their theories, I shall refer to them as *underconsideration problems* for IBE.[10]

The subtle differences between the slightly different underconsideration problems posed by these authors will not be important in what follows. What is important is to understand that underconsideration – i.e., the failure to consider serious alternatives – is a fact of epistemic life, especially in science. It really does happen, for instance, that the best hypothesis scientists have thought of at a given time is not in fact as good an explanation as another hypothesis that hasn't yet been formulated at that time. Indeed, Stanford's (2006) motivation for his version of the problem is in large part based on a number of historical cases in which exactly this type of situation arose in the not-too-distant past, such as the case of Darwin's acceptance of his 'pangenesis' theory of heredity (Stanford, 2006, ch.3). Furthermore, scientists are often aware that inferring the best available hypothesis would be problematic for precisely this reason (Novick & Scholl, 2019). Hence these problems are not of the sort that could or should be addressed by some conceptual maneuver on which they simply disappear or dissolve, leaving everything as it was.[11] Rather, an account of explanatory reasoning should allow us to identify how underconsideration problems arise in science and everyday life, which in turn would tell us what scientists and other epistemic agents can do to counteract them in practice.

If IBE is to provide such an account, it therefore needs to be supplemented with an account of the conditions under which underconsideration is not likely to be an issue. In other words, we would need to specify under what conditions an explanatory hypothesis that provides a better explanation than any other available hypothesis is 'good enough' to be inferred. These additional conditions should preferably be such that the epistemic agents who are making the inferences, e.g. scientists, are able to tell with at least some reliability when the conditions are in fact satisfied. In what

[10]See also Lipton (1993) and Wray (2008, 2011). There are related issues with IBE that concern the asymmetry between the comparative evaluation involved in IBE and the absolute conclusion it is meant to warrant (Dellsén, 2017a; McCain & Poston, 2019), but as I explain in section 6 these differ from the type of asymmetry problems that I want to focus on in that they are not concerned with the possibility of unconceived or unconsidered alternatives.

[11]Lipton (1993) and Schupbach (2014) respond to underconsideration problems along these lines, effectively arguing that these problems are based on a philosophical or logical confusion. These responses are criticized by Wray (2008), Khalifa (2010), and myself (Dellsén, 2017b).

follows, I will sketch an account of such conditions in terms of what I call *explanatory consolidation* – a process in which it becomes increasingly plausible that a hypothesis is not just better than all available alternatives, but also better than all alternatives that haven't been made available. A hypothesis is thus 'good enough' to be inferred via IBE if it has been through such a process of explanatory consolidation. Before I flesh out this suggestion, however, I will consider a different proposal for how to specify when the best available explanatory hypothesis is 'good enough'.

3 'Good Enough' as a Loveliness-Threshold

Recall that an explanatory hypothesis is considered 'better' than another for the purposes of IBE just in case it has a greater balance of various explanatory virtues, such as simplicity and explanatory scope. Following Lipton (2004), I will refer to this balance of explanatory virtues as 'loveliness', and thus say (with Lipton) that in the context of IBE, 'better' can be defined as *lovelier*. This immediately suggests that 'good enough' can similarly be defined as *lovely enough*, which in turns suggests that a hypothesis should be considered 'good enough' to be inferred via IBE just in case its loveliness – i.e. its balance of explanatory virtues – exceeds some designated threshold.[12] On this proposal, although there may be no specific threshold for each explanatory virtue, there is a threshold for the overall loveliness that is in turn determined by these explanatory virtues collectively. Thus, for example, a somewhat complex hypothesis may still be 'good enough' to be inferred via IBE if it compensates for its lack of simplicity by possessing other explanatory virtues to a sufficiently high degree, e.g. by having a much greater explanatory scope.

This is a straightforward and attractive proposal for what is required for an explanatory hypothesis to be 'good enough'. It promises to reduce the problem of specifying when a hypothesis is 'good enough' to that of spelling out how explanatory virtues should be defined and weighed against each other in an evaluation of the hypothesis' overall loveliness. Since the latter is something that proponents of IBE will have to do anyway, it would be unfair to object that compelling definitions of explanatory virtues and their respective weightings in overall loveliness are not

[12]Lipton seems to have had this type of account in mind in a brief remark on how one could respond to van Fraassen's Bad Lot Objection: "[A]n explanationist can endorse a two-stage inferential process without maintaining that the explanatory virtues that operate in the selection phase are comparative rather than absolute" (Lipton, 2004, 154). Unfortunately, however, Lipton did not elaborate on this suggestion.

yet forthcoming. Moreover, the underconsideration problems for IBE reviewed in the previous section would seem to disappear on the resulting model of explanatory reasoning, since there would now be a non-comparative form of evaluation built into IBE, viz. the evaluation of an explanatory hypothesis as sufficiently lovely, that would match the non-comparative conclusion that the relevant hypothesis is probably, approximately true.

Unfortunately, however, this loveliness-threshold view of 'good enough' is subject to at least three serious problems. The first problem concerns how a hypothesis' absolute level of loveliness could be measured in a way that would make it possible for someone to tell whether this loveliness-level exceeds the loveliness-threshold. Since loveliness is an agglomeration of various explanatory virtues, one would first have to measure the absolute level of virtuousness a given hypothesis enjoys for each virtue before agglomerating each such virtue into an overall level of loveliness. The problem is not just that it would have to be possible in principle to measure absolute degrees of virtuousness to cover all theories at all times, but that the resulting measures of virtuousness would have to be accessible to those making explanatory inferences as they determine the absolute degrees of virtuousness possessed by the theory. Only if this latter condition is satisfied could these absolute degrees of virtuousness subsequently be agglomerated in some way by the relevant agents, thus producing an overall absolute degree of loveliness which could be checked against the relevant loveliness-threshold. However, in many of the paradigmatic cases of explanatory reasoning, it is hard to see how this accessibility condition could possibly be satisfied.

Consider, for instance, the classic example of how ontological simplicity, i.e. parsimony, was used to adjudicate between the heliocentric and geocentric models of our solar system. It is relatively easy to see that the heliocentric model of the solar system is simpler than the geocentric model in a comparative sense, viz. because the former posits fewer ellipses than the latter. By contrast, it is rather less clear how one could measure the absolute levels of parsimony that are supposedly possessed by each model in a way that would enable us to tell whether the loveliness of each exceeds the relevant threshold.[13] The problem is exacerbated by the fact that the

[13]It is telling that those who the use of parsimony in theory choice are invariably concerned only with a comparative notion of parsimony (e.g., Quine, 1963; Swinburne, 1997; Baker, 2003; Huemer, 2009b; Biggs & Wilson, 2017; Jansson & Tallant, 2017). Indeed, Ockham's Razor is also standardly understood as appealing only to comparative parsimony: "Other things being equal, if T_1 is more ontologically parsimonious than T_2 then it is rational to prefer T_1 to T_2" (Baker, 2016). As far as I know, not a single author has so far even attempted to defend or define an absolute notion of parsimony.

absolute measure of parsimony that would be required would of course not be relative to ellipses, or any other specific entities, since that would make the resulting view of 'good enough' inapplicable to explanatory inferences in which those specific entities are not posited. Rather, the measure of parsimony with which we are operating would have to be one according to which any hypothesis whatsoever, regardless of its domain, is assigned an absolute degree of parsimony.[14]

The issue here is not merely that the loveliness-threshold view of 'good enough' requires that there *exist* such absolute measures of explanatory virtuousness, but that those who are making explanatory inferences would have to be able to *tell* what level of virtuousness each measure assigns to each explanatory hypothesis. If IBE is made to include a condition that theories must exceed some threshold of absolute loveliness, then those who are making explanatory inferences, e.g. working scientists, would have to have at least a rough idea of how to tell when this absolute degree of loveliness has been reached. Given that we are very far from having any sort of measure of absolute degrees of explanatory virtuousness,[15] let alone any firm conceptions of how to then agglomerate them into a measure of overall loveliness, it simply is not plausible that working scientists – or other agents who are in fact engaged in explanatory reasoning – could have any reliable means of telling whether a given theory possesses the requisite absolute degree of loveliness. Thus, the resulting version of IBE turns out to be unworkable, due to the practical impossibility of determining whether the required loveliness threshold has been reached in a given case.

I turn now to a second problem with the loveliness-threshold view of 'good enough'. It happens frequently in science that a firmly accepted theory is challenged by the emergence of a radically different alternative that is deemed superior to the previous theory. In such situations, scientists often hesitate to accept the new theory, despite its superiority to a theory that had previously been firmly accepted. Of course, scientists will be even more skeptical of the previous theory, which will be

---

[14]Notice that if hypotheses could be said to possess absolute values of individual explanatory virtues, e.g. parsimony, then it would have to be possible to compare any hypothesis to any other in terms of each virtue and overall loveliness. So in our example, we should be able to compare not just the heliocentric and geocentric models (or other possible models of the solar system) in terms of their relative parsimony, but also compare each model's parsimony to, for instance, the parsimony of Newtonian mechanics, the parsimony of Lamarck's theory of inherited characteristics, and the parsimony of the law of supply and demand. Each theory would have an absolute degree of parsimony that would be higher than, equal to, or lower than the absolute degree of parsimony possessed by any other theory.

[15]See footnote 13.

either discarded or at best viewed as a useful approximation or fiction. The point here is that even the new theory is frequently viewed with suspicion, at least for the time being – especially if the emergence of the new theory is unexpected and thus indicates that yet other theoretical alternatives may still be overlooked. I discuss an historical case of this sort in in section 4 below, viz. the reception of Fresnel's transverse wave theory of light in the decades after it was first formulated in 1815-1818. To foreshadow that discussion, Fresnel's theory was already in 1819 considered to be explanatorily superior to its available alternatives, including Newton's corpuscular theory, which had been accepted up to that time. So Fresnel's new theory was considered lovelier than Newton's theory, which must have been considered sufficiently lovely to be accepted. And yet Fresnel's theory was viewed with considerable suspicion by many prominent optical physicists for most of the 1820s.[16]

Situations of this sort create the following problem for the loveliness-threshold view of 'good enough'. If IBE is applicable at all to these cases, then the emergence of the new theory must have undermined the acceptance of the previous theory in virtue of its greater loveliness. And since the previous theory was firmly accepted, it must have been taken to exceed the required loveliness threshold. It follows that the new theory should have been taken to exceed the loveliness threshold as well. But then how come this new theory is not accepted? Given the loveliness threshold view of 'good enough', the new theory would not just be considered better than any available alternative – including the previously accepted theory – but also *better than good enough*. What we actually find in at least some of these cases is that scientists hesitate to accept the new theory, instead preferring to hold off on reaching any decisive conclusions until much later. In sum, then, the loveliness-threshold view of 'good enough' implies an overly-simplistic picture of explanatory reasoning, in which newly-proposed, lovelier alternatives would automatically be deemed inferable via IBE if their predecessors had been deemed to be inferable.

A third problem with the loveliness-threshold view of 'good enough' concerns where the loveliness-threshold is set, and how that in turn relates to loveliness-comparisons between available explanatory hypotheses. Is the threshold set so high as to guarantee that any given hypothesis that exceeds the threshold is lovelier than all possible alternatives, or is it possible for a hypothesis to exceed the loveliness-threshold and yet be equally or less lovely than some (perhaps currently unconceived) alternative? These options are exhaustive, but each one is problematic. If we choose

[16] See section 4 for a fuller discussion of this case.

the former, then a hypothesis' exceeding the loveliness-threshold would already guarantee that it is lovelier than all possible alternatives, so there would be no point in comparing such a hypotheses to available alternatives. In that case, one might as well skip the superfluous step of actually comparing hypotheses in IBE and go straight to an absolute evaluation of its explanatory loveliness. Thus this option would make a mockery of the idea that IBE involves a comparative step in which different hypothesis are juxtaposed and contrasted with one another. Furthermore, since scientists do in fact reason by comparing and contrasting different available explanations, an account of 'good enough' that makes such comparisons utterly superfluous cannot hope to adequately account for how explanatory reasoning actually works in science.

This leaves us with the latter option, i.e. of allowing that an explanatory hypothesis that exceeds the required loveliness-threshold is not necessarily lovelier than all possible alternatives. On this suggestion, the loveliness-threshold is sufficiently low so as to allow that a number of possible explanatory hypotheses are 'good enough', even if only one of these is loveliest and thus the one that ought to be inferred via IBE. The problem with this suggestion, however, is that it fails to solve the very problem that the 'good enough'-condition on IBE is introduced in order to solve. After all, a hypothesis could then be lovelier than all extant alternatives, and moreover exceed the loveliness-threshold, and yet still be less lovely than some currently-unconceived alternative hypothesis. Indeed, the agent in question could even have compelling *reasons to believe* that such a currently-unconceived alternative could be formulated, e.g. because they are aware that such unconceived alternatives have often emerged in the recent past. And yet on the current suggestion the hypothesis should nevertheless be inferred, due to its being both lovelier than available alternatives and also lovelier than the relevant loveliness-threshold. Since this is exactly the type of underconsideration problem that a 'good enough'-clause is meant to save IBE from, the view currently under consideration is just not fit for purpose.

Taking stock, I have identified three problems for the loveliness-threshold view of 'good enough'. First, there is the problem of defining and measuring absolute degrees of explanatory virtues and overall loveliness in a way that could possibly be used by those who are meant to be making explanatory inferences. Second, the view wrongly implies a simplistic picture of scientific theory change where new theories should automatically be accepted if they are deemed lovelier than previously-accepted alternatives. Third and finally, the loveliness-threshold view of 'good enough' is subject to a dilemma, the upshot of which is that the loveliness threshold would either (a) make explanatory comparisons between available hypotheses utterly superfluous, or

(b) be subject to a version of the very same underconsideration problem that a 'good enough'-clause is meant to save us from. Any one of these three problems seems to me to be sufficiently serious to look elsewhere for a plausible account of what makes a hypothesis 'good enough' to be inferred via IBE. But where to look? I suggest it is high time that we examine how working scientists themselves approach underconsideration problems in science.

4 Consolidating the Transverse Wave Theory of Light

In a series of scientific papers published between 1815 and 1818, the young French chemist Augustin Fresnel proposed and argued for a new type of wave theory of light.[17] At the time, the dominant theory of light was Isaac Newton's corpuscular theory, according to which light consists in discrete particles emitted from the light source. By contrast, Fresnel's wave theory holds that light is a *transverse wave*, i.e. a wave that oscillates in a direction perpendicular to its movement. Fresnel argued for his theory by showing how it alone could explain various observed behaviors of light without requiring any additional theoretical posits or adjustments to the core theory. In particular, Fresnel showed in some detail how the wave theory effortlessly explained the diffraction of light, i.e. how light appears to 'bend' around the edges of opaque objects – a phenomenon that becomes particularly apparent when light falls on narrow slits. Fresnel's theory was further confirmed when a particularly surprising prediction made by the theory, viz. that a bright spot would appear in the midpoint of a shadow cast from an opaque disk, was shown to obtain by Fresnel's countryman Francois Arago. No explanation of this surprising phenomenon could be provided by Newton's corpuscular theory, which posited that light travels in perfectly straight lines originating in the light source.[18]

Fresnel's theory was thus unquestionably explanatorily superior to the corpuscular theory that had dominated optics in the decades before, especially after Arago confirmed the wave theory's surprising prediction. And yet, in the decade or so following the presentation of Fresnel's theory – i.e. in the late 1810s and the early to mid 1820s – scientists still disagreed considerably about whether the new theory

[17]The following discussion of Fresnel's theory owes much to Buchwald's (1989) thorough and comprehensive study of the transverse wave theory and its early history.

[18]This episode is often taken as a paradigm example of how IBE is employed in science (see, e.g., Thagard, 1978; Norton, 2018b).

should be accepted as true. In particular, the revered optical physicist Sir David Brewster, and several other British physicists, argued that neither Fresnel's theory, nor the competing corpuscular theory, should be accepted outright (Cantor, 1975). Although even Brewster admitted that the transverse wave theory was superior to the corpuscular theory, he suggested that it should be employed "only as a temporary auxiliary" (Brewster, 1837, 143). Apparently, then, Fresnel's case for the wave theory had not convinced British physicists at the time that his wave theory was true; instead, it served only to secure the transverse wave theory's status as the most plausible theory of light among those that had been proposed thus far.

It is worth noting that this does not sit well with the two-stage model of IBE in which one first generates a set of explanatory hypotheses and then infers the hypothesis that provides the best explanation of the relevant data. According to this two-stage model, British optical physicists ought to have inferred the wave theory already in 1819 or soon thereafter, since it unquestionably did provide the best explanation of the available hypotheses at the time – even by their own lights. Nor does this sit well with a modified two-stage model of IBE on which the second stage includes the clause that the inferred hypothesis' explanatory loveliness must also exceed some loveliness-threshold. On this modified two-stage model, there are three options for what Brewster and his colleagues ought to have done with the new theory. First, if they deemed the theory's explanatory loveliness to be *above* the threshold, they ought to have accepted it outright – which they didn't do. Second, if they deemed its explanatory loveliness to be *below* the designated threshold, they ought to have rejected it outright – which they didn't do either. A third and final option is that Brewster and his colleagues were *unsure* as to whether the theory had reached the loveliness threshold. Would this be a convincing interpretation of the episode?

No, because there is simply no indication that Brewster and his colleagues were genuinely unsure about whether Fresnel's theory ought be accepted (see Cantor, 1975, 118-120). Brewster argued extensively that the wave theory – and indeed any theory available at the time – was unable to explain a number of experimental results, such as why light of certain wavelengths is absorbed in nitric acid gas while other light isn't (Brewster, 1833). When William Whewell (1837) argued that Fresnel's theory should be considered to be an established truth, Brewster objected strenuously to Whewell's positive evaluation of the theory, referring to it as "wholly unfounded" (Brewster, 1837, 143). Hence, Brewster's view was certainly not that Fresnel's theory enjoyed some ambiguous epistemic status such as to make it hard to tell whether the theory ought to be accepted or not. Rather, his view was unequivocally that the theory was not good enough to be accepted. The complication here is that, in Brewster's view, this unambiguously negative verdict about the acceptability of the theory was consistent with his recommendation that the theory should be adopted "as a temporary auxiliary", since in his opinion no other theory was more plausible than Fresnel's.

Of course, as it turned out this tentative adoption of the theory "as a temporary auxiliary" was itself only temporary, to be replaced with a more resolute attitude towards the theory in due course. In the late 1820s and early 1830s, the transverse wave theory gradually became accepted by the most influential scientists of the day, including British physicists George Biddel Airy, William Herschel, and Whewell. A decade later, in the early 1840s, the core tenents of the transverse wave theory had become nearly universally accepted (Buchwald, 1989, 308). Through further empirical and conceptual work on the theory, the transverse wave theory would be developed into James Clerk Maxwell's electromagnetic theory of light in the latter half of the 18th century, which would in turn lead to and be incorporated into Einstein's theories of special and general relativity in the early 20th century. Thus, in a way, Fresnel's transverse wave theory of light is still accepted as true in a modified form – a point that is often emphasized by selective scientific realists (e.g., Worrall, 1989; Psillos, 1999; Chakravartty, 2007).

Crucially, however, the eventual acceptance of the transverse wave theory was not simply based on a generation of the theory and subsequent comparative evaluation of it against available alternatives. Rather, the theory would not become widely endorsed in the scientific community until it had already been employed "as a temporary auxiliary" among British physicists for over a decade. Only then was the transverse wave theory considered sufficiently secure for it to be accepted outright, i.e. as probably, approximately true. What emerges from this episode, then, is a more nuanced picture than the two-stage model of IBE provided by Lipton in which explanatory hypotheses are first generated and then compared with an eye

[19]Norton (2018b) similarly describes paradigmatic cases of IBE as including a step in which "favoring is rendered absolute: we are authorized to infer to the favored hypothesis or theory" Norton (2018b, 1). However, Norton does not include what Lipton (and I) would refer to as the *first* step, i.e. the generation of available theories. Norton also appears to be skeptical that anything general could be said about how the final step works: "there can be no universally applicable schema that fully characterizes inference to the best explanation" Norton (2018b, 1). In the following section, I effectively argue that Norton's pessimism is premature, since a general account of the final step can be provided based on the core insight behind IBE.

towards inferring the one that provides the best explanation. Instead, the episode would be more accurately described by a three-stage model of explanatory inference in which (i) a set of potentially explanatory hypotheses are generated, (ii) one of these is adopted as a working hypothesis ("temporary auxiliary") on the grounds that it provides the loveliest, i.e. the best, explanation of the available evidence, and then (iii) this working hypothesis is gradually considered increasingly plausible and is eventually accepted by the scientific community.[19]

The last of these steps is what I refer to as *explanatory consolidation*. It is in this step, I suggest, that explanatory hypotheses go from merely being considered the 'best' among available alternatives to being considered 'good enough' to be inferred via IBE. Of course, this is not yet much of a solution to the problem of what makes a hypothesis 'good enough', since we will still have to say what it is that changes during explanatory consolidation so as to make a hypothesis that was previously merely considered to be the best among available alternatives into one that is considered to be sufficiently plausible to be accepted. However, there is an important insight to be gleaned from the history of Fresnel's transverse wave theory, viz. that in actual science the logical gap between 'best' and 'good enough' is often bridged, if at all, *over time*. This highlights that explanatory consolidation is not so much a condition that an explanatory hypothesis either meets or fails, as it is a temporally extended process in which the hypothesis is 'good enough' only at its endpoint. The key question, however, is how explanatory consolidation works epistemically: Why and when should a working hypothesis eventually be accepted by the scientific community at the end of this process — not just as the best available explanation, but as (probably, approximately) true?

5 The Epistemology of Explanatory Consolidation

It might be tempting to say that explanatory consolidation should proceed in much the same way as the previous step in IBE, i.e. by accumulating various bits of evidence that the hypothesis explains better than its available alternatives. There is surely something to this suggestion.[20] However, it should be clear by now that this cannot be the full story of how explanatory consolidation works. For recall from the second problem discussed in section 3 that when an explanatorily superior alternative

[20]As we shall see, the account of explanatory consolidation presented below accommodates the kernel of truth in this idea.

to a previous theory emerges, the new theory is frequently met with suspicion even if the previous theory had been firmly accepted. Indeed, the brief case study of the previous section fits this pattern, since the transverse wave theory failed to be accepted even by some of those scientists who acknowledged its superiority to the corpuscular theory (which they themselves had previously accepted as true).[21] Cases of this sort do not fit the view that explanatory consolidation simply requires 'more of the same' – i.e., more evidence that the theory explains better than its available alternatives – since the new theory is, if anything, supported by even more such evidence than the previous theory had been when it was accepted.

These situations arise, I suggest, because the arrival of the new theory changes the scientific community's estimation of whether yet other theoretical alternatives may have been overlooked. Specifically, the emergence of an explanatorily superior new theory to replace a previous theory suggests that the new theory may itself be vulnerable in the very same way as the previous theory was thereby revealed to be. In such situations, the emergence of a new theory undermines its own acceptance – paradoxical as that may seem. In the case of Fresnel's transverse wave theory, for example, Fresnel's presentation of the theory in 1818 acted as a catalyst for speculation about whether yet other theories, currently unconceived, could explain the relevant phenomena as well or better. The epistemic problem that arises in these situations is how explanatory consolidation could provide the scientific community with the means to rationally regain its confidence that there are no such alternative theories in logical space, i.e. that the most plausible alternatives that *could be* formulated *have been* formulated. And of course, the same type of problem arises also in the simpler case where the new theory isn't replacing an extant theory, but is rather the first serious theory to explain the relevant phenomena better than all available alternatives. In general, the epistemic problem is to describe what sort of reasons explanatory consolidation could provide for thinking that no explanatorily superior theory could be formulated with respect to the relevant phenomena.

If possible, our solution to this problem should appeal to resources already con-

---

[21] One might think that cases of this sort can be explained by invoking Kuhnian paradigm-shifts [acknowledgements omitted]. On this view, the new theory is met with suspicion by proponents of the old paradigm because it somehow fails to fit into other aspects of the existing paradigm. However, in order for this this suggestion to work in the case of the transverse wave theory, there would have had to be a Kuhnian paradigm-shift between 1818 and the 1830s, which would seem to exaggerate the differences between optical scientists before and after this period. So unless 'paradigm-shift' is used in a very loose sense – in which case it would arguably lose all explanatory force – I don't think these types of episodes can be usefully described in Kuhnian terms.

tained within a broadly-speaking explanationist framework, rather than appealing to other models of non-deductive reasoning, such as Bayesianism. Otherwise, IBE could not possibly be an autonomous or fundamental form of non-deductive reasoning, as many of its proponents maintain (e.g., Lycan, 1988; Weintraub, 2013; Poston, 2014). Indeed, incorporating elements from other frameworks of non-deductive reasoning would also undermine the project of viewing IBE as an accessible heuristic for some other more ideal form of inference, such as Bayesianism (see, e.g., Cabrera, 2017; Bird, 2017; Dellsén, 2018), since the resulting heuristic would thus require the relevant agents to master an entirely different form of reasoning in addition to standard IBE. In particular, if explicating explanatory consolidation required appealing to Bayesian principles or calculations, then IBE would clearly not work as a an accessible heuristic for Bayesian reasoning, since the type of reasoning IBE is meant to help us approximate would then be involved in IBE itself.

What we need, then, is an *explanationist* account of explanatory consolidation. I suggest the following approach, parts of which are anticipated in another context by Richard Dawid's 'No Alternatives Argument' (Dawid, 2013; Dawid et al., 2015; Dawid, 2018).[22] In any process of explanatory consolidation, we know from the get-go that some hypothesis, H, provides a better explanation than any of the competing explanatory hypotheses that have so far been considered. (After all, that is the upshot of the previous step in our new three-step model of IBE.) Now, what might explain this fact about H? One explanation is that H in fact provides the best explanation of all explanatory hypotheses in logical space, i.e. that H provides the best explanation that could, logically speaking, be formulated. Call this *the*

[22]With that said, there are several important differences between Dawid's argument and the the account of explanatory consolidation given here. I will just mention three key differences: First, Dawid's argument is primarily put forward in the context of the debate about scientific realism rather than IBE (although see Dawid, 2013, 64-68). Accordingly, Dawid does not address the main concern of this paper, viz. what it takes for an explanatory hypothesis to be 'good enough' to be inferred in the context of IBE. Second, whereas 'explanatory consolidation' is a *diachronic* process in which theories gradually gain plausibility over some period of time, Dawid's No Alternatives Argument works as a *synchronic* consideration in favor of a theory at a given time. Dawid's synchronic approach makes perfect sense in the context of the realism debate, where the question is what epistemic status to assign to scientific theories *now*, whereas my diachronic approach is more appropriate in the context of understanding explanatory reasoning, which is a temporally extended process. Finally, Dawid's No Alternatives Argument is meant to support the 'predictive viability' of a given theory – i.e. that the theory will turn out to make correct prediction in the next round of empirical testing – whereas I view explanatory consolidation as incrementally supporting the notion that the relevant theory is *true*. Dawid's restriction to 'predictive viability' is natural in the context of contemporary theorizing in fundamental physics, but since my aim is to give an account of IBE that applies in more mundane scientific contexts, as well as in everyday contexts, it is natural to view truth as the conclusion of the relevant IBEs (although see footnote 8).

optimistic explanation. Another explanation is that although there is at least one better explanatory hypothesis than H in logical space, these explanations have not (yet) been formulated. Call that *the pessimistic explanation*. Thus we have the following competing explanations for the fact that H is lovelier than its available alternatives:[23]

The optimistic explanation: The hypothesis H is the loveliest explanation that could be formulated, i.e. the loveliest explanation in logical space.

The pessimistic explanation: Although a lovelier explanation than H has not been formulated, there is a lovelier explanation than H in logical space.

By construction, these two explanations are mutually exclusive and jointly exhaustive. So, necessarily, one and only one is true.[24]

The question is just which one. Fortunately, that is the type of problem that we already know how to solve – using the resources of IBE. We simply ask which of these two explanations is 'better', i.e. lovelier, in light of background evidence. What makes this case somewhat special is only that the phenomenon to be explained is not some ordinary set of empirical data, but rather the fact that a hypothesis H currently provides the best available explanation of some other phenomena. Since different hypotheses can give better and worse explanations of this fact, they too can be evaluated in an explanatory comparison of the very same sort as in the second step in an IBE. In particular, note that the explanatory loveliness of each such second-order explanation will of course be relative to the background information one possesses at a given time, just as in any other IBE.[25] Thus, as such background information accumulates, the comparison in overall explanatory loveliness may tilt in favor of the explanation that had previously been deemed inferior.

[23] I refer to these as 'explanations' rather than 'explanatory hypotheses' simply to distinguish them from claims like H and its alternatives.

[24] This point is important because it means that there is no question about whether the better explanation among the optimistic and pessimistic explanations is also the best of all possible explanations; rather, that simply follows immediately from the fact that the two explanations are jointly exhaustive. Thus there is no need for a second-order explanatory consolidation once either of these two explanations has been shown to be superior to the other.

[25] Indeed, the necessity of evaluating explanations with reference to background information is so crucial that it is often listed as one of the explanatory virtues (see, e.g., Lycan, 1985; Lipton, 2004; Beebe, 2009; Cabrera, 2017). That said, even those who do not list fit with background evidence as one of the explanatory virtues will have to acknowledge that the plausibility of any non-deductive is clearly relative to the state of this background evidence, as dictated by 'the principle of total evidence' (Carnap, 1947).

This is what I suggest happens in explanatory consolidation. At the beginning of the process, the explanatory comparison between the two explanations is tilted towards the pessimistic explanation. (Necessarily so, since otherwise the problem to which explanatory consolidation is a solution would never arise.)[26] As time goes by and different types of background information accumulates – more on which below – this gradually favors the optimistic explanation. That is, it becomes increasingly plausible that the fact that H is the loveliest explanatory hypothesis currently available is explained not by our failure to conceive of even lovelier alternatives to H, but by there being no lovelier alternatives to H in logical space. Of course, this gradual favoring of the optimistic explanation over the pessimistic explanation will take time, which is why explanatory consolidation is a temporally extended process rather than an instantaneous event. At some point, however, the overall comparison in loveliness finally tilts decisively in favor of the optimistic explanation, at which point the process is complete. H has been explanatorily consolidated.

The 'background information' that accumulates in this process is of two crucially different kinds. First of all, there is the familiar type of information that we might label 'direct empirical evidence', e.g. experimental results and observations, that H explains better than do available alternatives. Each piece of evidence of this kind does not just support the hypothesis H (although it certainly does that as well); it also adds to the stock of evidence that any alternative to H would have to explain in order to be as explanatorily lovely as H. So, as this kind of evidence for H accumulates, it becomes gradually less plausible that it would be possible to formulate an alternative to H that is also able to explain this growing amount

[26]There are of course many cases in which explanatory consolidation is not needed, because it will be clear from the get-go what are the most plausible explanations for the relevant phenomena. This happens frequently in everyday circumstances, where the most plausible explanations will come to mind more or less immediately. Consider, for example, the case that Weisberg uses as an illustration of an everyday IBE: "Suppose you come home one day to find the front door open and the lock broken. Furniture is overturned, the contents of the shelves are on the floor, and valuables are missing" (Weisberg, 2009, 129). In this case, there is no need for explanatory consolidation because it is obvious from the get-go that the hypothesis that your home has been burglarized better explains these facts than any competing hypothesis that could be offered. Thus the relevant optimistic explanation – viz. that the burglary hypothesis is lovelier than any alternative hypothesis in logical space – is already vastly superior to the pessimistic explanation. However, I am interested in the more difficult set of cases in which it is not obvious how to explain some set of facts, as is arguably the case for most phenomena of scientific interest [acknowledgements omitted].

of evidence.[27] For example, it was more plausible that a lovelier alternative to Einstein's theory of general relativity could be formulated before the solar eclipse expeditions of 1919, because at that point such an alternative would not have had to explain the bending of light in strong gravitational fields such as that around the sun. More generally, every time some new evidence emerges that is explained by general relativity, this makes it less plausible that it would be possible to formulate a lovelier alternative to general relativity, because such an alternative would then have to explain this new evidence in addition to all the previous evidence in order to match general relativity's explanatory scope. In this way, direct empirical evidence for a hypothesis incrementally supports the optimistic explanation (that there is no lovelier alternative to the hypothesis) over the pessimistic explanation (that there are lovelier alternatives that haven't yet been considered). The accumulation of direct empirical evidence for a hypothesis thus consolidates the hypothesis.[28]

Importantly, though, direct empirical evidence is not the only type of background information that accumulates in explanatory consolidation and helps to favor the optimistic explanation over the pessimistic explanation. Another kind of information that accumulates in this process concerns the repeated non-occurrence of a specific type of event, viz. the emergence of an alternative hypothesis that is indeed lovelier than H.[29] The pessimistic explanation implies that this type of event should with some probability occur sooner or later, assuming only that it is humanly possible to formulate such a hypothesis. By contrast, the optimistic explanation implies that no such event could ever take place, since it holds that the loveliest possible explanation is already among those that have been formulated. So each period of time in which a lovelier alternative to the currently loveliest explanatory hypothesis fails to be formulated – especially if scientists are actively searching for such alternatives, as

[27]Compare Dawid (2018, 503): "The strength of limitations to scientific underdetermination crucially depends on the strength of of the confirming data: the more striking the agreement between the collected data and a theory's predictions, the fewer alternative theories may be expected to match that predictive success."

[28]This explains the kernel of truth in the thought that explanatory consolidation simply consists in collecting more evidence in favor of the hypothesis. As I am suggesting here, this is one way in which consolidation occurs; but as I am about to suggest, it is not the only way.

[29]This is the type of event that fuels Dawid's No Alternatives Argument for the predictive viability of string theory (see esp. Dawid, 2013, 31-33). According to Dawid's argument, the fact that no serious alternatives to string theory have been discovered despite an intense search, in conjunction with certain theoretical constraints that seem to imply string theory, lend credence to the viability of the theory. Elsewhere, Dawid (2017) argues that similar reasons supported the existence of the Higgs boson before it could be detected empirically. See also Dawid et al. (2015) for a Bayesian reconstruction of this type of non-empirical argument.

they often are in explanatory consolidation – we gain some small additional reason to think the loveliest hypothesis is not just lovelier than any extant alternative, but also lovelier than all possible alternatives. In other words, as we collectively and repeatedly fail to formulate lovelier alternatives to our currently-loveliest explanatory hypotheses, we gradually gain reason to favor the optimistic explanation over the pessimistic one. Of course, each failure to formulate a lovelier hypothesis isn't by itself a strong reason to think no such hypothesis exists in logical space. But as such attempts accumulate and begin to jointly constitute a long history of systematic attempts to formulate a lovelier alternative, all of which have failed, that will eventually provide a quite significant reason to think this failure is not due to a lack of imagination on the part of the theorists, but rather due to the hypothesis being lovelier than any hypothesis that could be formulated.

In most cases of explanatory consolidation, these two types of information accumulate in tandem to consolidate the relevant hypothesis. That is, we both gain more direct empirical evidence in favor of the hypothesis H, and also repeatedly fail to come up with lovelier alternatives to H. Interestingly, there appear to be cases in which only the second type of information accumulates, due to some difficulty or other involved in gathering direct empirical evidence. For example, it is notoriously difficult – indeed, practically impossible, as far as we know – to gather direct empirical evidence in favor of the novel posits of string theory. However, as Dawid (2009, 2013) points out, string theory is also remarkable for the fact that there is a notable lack of plausible rivals to the theory, despite numerous attempts to formulate such alternative 'theories of everything'. Dawid argues that this history of failed searches for superior alternatives constitutes a strong reason to accept the theory. Whether we agree with Dawid's argument in this particular case, it should at least be clear that a history of repeated searches for an alternative theory can lend support to the optimistic explanation that no such alternative theory exists in logical space.

The converse might also be possible, at least in the sense that failure to find plausible alternatives plays little or no role in the consolidation of the hypothesis. This occurs when the direct empirical evidence for a hypothesis H is so strong that we can tell from that evidence alone that H is not merely the best one currently available but also the best one that could be made available. For example, several philosophers of science have suggested that Jean-Baptiste Perrin's experimental tests of the atomic

[30] See, for example, Salmon (1984, 2005), Achinstein (2002), Maddy (1997, 2007), Roush (2005); although see also Stanford (2009) and van Fraassen (2009).

theory of matter, building on Einstein's predictions of Brownian motion, provide nearly conclusive evidence of its truth.[30] If so, this can be viewed as a case in which the explanatory consolidation works by accumulating enough direct empirical evidence to rule out beforehand the possibility that any equally-or-more plausible alternative to the atomic theory could be developed. In that case, there would be no need to attempt – and fail – to develop alternatives to the atomic theory, for the direct empirical evidence has effectively already narrowed down the space of theories that are reasonably consistent with the evidence so as to leave only the atomic theory in the running among those that should be taken seriously.

Let us take stock. Our goal has been to construct an account of the process by which a hypothesis H goes from merely being lovelier than available alternatives, to being 'good enough' to be inferred via IBE. My suggestion is that H is 'good enough' just in case the optimistic explanation – that the hypothesis H is the loveliest explanation that could possibly be formulated – has itself become a lovelier explanation than the pessimistic explanation – that although a lovelier explanation than H could be formulated, this explanation has not in fact been formulated. 'Explanatory consolidation' is my term for the gradual process by which this happens, and I have suggested that it consists in the accumulation of two quite different types of information that incrementally favor the optimistic explanation over the pessimistic explanation. Specifically, explanatory consolidation occurs as (a) direct empirical evidence for H accumulates, thereby decreasing the plausibility that any alternative to H could explain all the evidence as well as H, and (b) repeated attempts are made to formulate lovelier explanations than H, but every such attempt fails. If all goes well, enough information of this type will eventually have accumulated that a comparison of the overall explanatory loveliness tilts decisively in favor of the optimistic explanation as against the pessimistic explanation, at which point H has been consolidated and is thus 'good enough' to be inferred via IBE.

6 Consolidated IBE: Rivalry and Risk

Having spelled out what explanatory consolidation is and how it works, I will now briefly examine the nature and limits of the three-step model of IBE in which explanatory consolidation serves as the third and final step. I call this model *consolidated IBE* in order to distinguish it from Lipton's (2004) standard two-stage model. In particular, I will respond to two potential worries about consolidated IBE in the hope that doing so clarifies the role of explanatory consolidation within this new three-step model. In different ways, both of these worries concern the fact that consolidated IBE still involves reaching an *absolute* verdict – that a given hypothesis is probably and/or approximately true – on the basis of a *comparative* evaluation – that the hypothesis is lovelier than some collection of alternatives.

The first worry concerns situations in which there are several known alternative hypotheses each of which is at least somewhat lovely as compared to the loveliest hypothesis H. Consider, in particular, an example I have previously used to illustrate how these situations undermine IBE (Dellsén, 2017a, 23-25). The loveliest explanation for the origin of life on Earth is arguably the RNA world hypothesis, according to which life began with the formation of self-replicating RNA molecules. However, there are also several plausible alternative explanations, which appear to prevent biologists from inferring that the RNA world hypothesis is correct. In these types of cases, there is an embarrassment of riches in the number and plausibility of possible explanations which undermines inferring that even the best of these is probably, approximately true. Accordingly, I have chosen to call this *the problem of multiple plausible rivals*.[31]

Although the problem of multiple plausible rivals is a real problem for consolidated IBE, it is important to see that it is distinct from the kind of underconsideration problems that I set out to discuss in this paper. Underconsideration problems concern the possibility or plausibility that the best explanatory hypothesis has *not been considered*, whereas the problem of multiple rivals concerns the possibility or plausibility that one of the less-than-loveliest hypotheses *that has been considered* is nonetheless true. Put differently, the problem of multiple rivals concerns explanations that have already been formulated, whereas underconsideration problems concern explanations that haven't been formulated. Of course, some solution will need to be offered at some stage to the problem of multiple rivals, but since this problem is clearly distinct from underconsideration problems, we can simply choose

[31]A related objection is developed by McCain & Poston (2019), based on a brief remark by Fumerton (1995, 209); see also van Fraassen (1989, 149). McCain and Poston refer to their objection as *the disjunction objection*, since the idea is that the disjunction of alternative hypothesis can sometimes be shown to be more plausible than the hypothesis that offers the best explanation. McCain and Poston offer a response to their disjunction objection, which is that in cases of this sort the loveliest explanatory hypothesis isn't 'good enough' to be inferred via IBE because it doesn't exhibit a sufficiently good fit with background evidence. This response appears to invoke a loveliness-threshold view of 'good enough' – which would be problematic for reasons given in section 3 – or else leave 'good enough' undefined – in which case it is incomplete. For this reason, I prefer my own approach to the problem (Dellsén, 2017a), which is further discussed below.

whatever solution we prefer to the latter independently of our solution to the former.

So consider, in particular, my own previous own approach to the problem of multiple rivals (Dellsén, 2017a, 25-28). In brief, I argued that IBE is a special case of a more general inference pattern in which one infers to what is entailed by the k best explanations available, for some number k. In so far as one wants to avoid inferring the best explanation from a set of multiple plausible ones, one simply employs the instantiation of this pattern where k is sufficiently large, thus inferring only what all the k best explanations have in common. For example, one may note that the three best alternative explanations for the origin of life all posit that life began with the formation of a nucleic acid – although they disagree about *which type* of nucleic acid (PNA, TNA or GNA). Since this is also true of the RNA world hypothesis, we can infer this 'robust' result using the instantiation of the pattern in which $k = 4$. I have referred to this pattern as 'abductively robust inference', but I also emphasized that it is really a generalization of IBE in which IBE emerges as the limiting case where $k = 1$.

This approach to the problem of multiple rivals turns out to be entirely compatible with, and indeed nicely complementary to, the approach to underconsideration problems I have been advocating in the current paper, i.e. explanatory consolidation. Thus far I have presented explanatory consolidation as a process in which it becomes increasingly plausible that the loveliest explanatory hypothesis that is available is also the loveliest such hypothesis in logical space. But the same process can be used, *mutatis mutandis*, for becoming rationally convinced that the k loveliest explanations in logical space have all been made available, for any k. Specifically, as direct empirical evidence for these k hypotheses accumulates, and as attempts to formulate even lovelier hypotheses than these k alternatives repeatedly fails, the fact that no available hypothesis is lovelier than these k hypotheses increasingly becomes better explained by the optimistic explanation that there is no such hypothesis in logical space than by the pessimistic explanation that such a hypothesis exists but just hasn't (yet) been formulated. Thus the process of explanatory consolidation is exactly the same as described earlier, *modulo* only the fact that its aim is not to consolidate a *single* hypothesis but a set of k such hypotheses, for some $k > 1$. Once that process is complete, one simply infers to what is entailed by all of these k hypothesis. In this way, this solution to the problem of multiple rivals would be incorporated without remainder into consolidated IBE.

So much for the first worry about consolidated IBE; let us turn to a second, albeit related, worry. Even with this latest modification to avoid the problem of multiple rivals, one might worry about a remaining asymmetry in consolidated IBE between the evaluation and the verdict of consolidated IBE. After all, the output of explanatory consolidation is still the comparative judgment that a hypothesis H provides (or that a set of hypotheses $H_1, ..., H_k$ contains) the loveliest explanation(s) of all hypotheses in logical space. On this basis consolidated IBE tells us to infer that the hypothesis H (or whatever is entailed by every $H_1, ..., H_k$) is probably and/or approximately true – an absolute verdict. Thus one might still think that IBE, even after adding explanatory consolidation as its final step and avoiding the problem of multiple rivals in the way described above, provides a kind support that is structurally of the wrong type to warrant any kind of absolute verdict. The worry, in brief, is that you cannot derive an absolute verdict from a comparative evaluation.

I see two plausible responses to this worry. One response is to weaken the verdict, conceding that IBE only really licenses the comparative verdict that the inferred hypothesis H is *more probably* and/or *more approximately* true than all possible alternatives. This would restore the symmetry between consolidated IBE's evaluation and its verdict by virtue of making the latter comparative in the same way as the former.[32] Although the verdict reached in IBE would thus be comparative, it is still a very strong type of verdict since it entails that H is epistemically superior to – i.e., more probably or more approximately true than – any other explanatory hypothesis that could possibly be formulated. Perhaps our canons of inductive inference can take us no further than to that point; or perhaps we should rely on some other model of inductive inference to take us further.

Another option is to dig in one's heels and insist that we should not let this remaining asymmetry motivate any further modifications to our account of explanatory inferences. After all, one might ask, what exactly is epistemically problematic about this remaining asymmetry? Consolidated IBE would involve the risk that the truth might not be contained within the k loveliest possible explanatory hypotheses, i.e. that a relatively *unlovely* hypothesis contains the truth. That is surely possible, but so what? We knew beforehand that explanatory inferences are *ampliative*, i.e. such that the 'inputs' from which one is inferring would not guarantee the truth of that which is inferred. So the mere fact that consolidated IBE still involves an epistemic risk is not by itself a serious objection to it as an account of explanatory

[32] A couple of different suggestions of this kind are provided by Kuipers (2000) and myself (Dellsén, 2018).

inferences. Things would be different if there was another account of explanatory inferences available in which this epistemic risk could be eliminated even while retaining the features of IBE that made it so attractive in the first place. No such account is forthcoming. Moreover, it is doubtful that such an account could be developed, since part of what makes IBE attractive is precisely that it seems descriptively accurate to characterize most actual explanatory inferences in science as involving a comparative evaluation of hypotheses as against their available alternatives, as opposed to some sort of absolute evaluation in which hypotheses are evaluated in isolation from one another.

Thus I tentatively suggest that we should learn to live with the remaining asymmetry between the comparative evaluation involved in IBE and the absolute verdict it is meant to license, at least in the absence of an alternative account in which the asymmetry has been eliminated. Alternatively, one could choose to weaken the verdict of consolidated IBE to mimic the comparative nature of the evaluation involved therein. In either case, we would have a modified version of IBE – consolidated IBE – which is designed to address the type of underconsideration problems that arise for explanatory inferences. Specifically, the final step in IBE would be a process in which, when all goes well, it becomes increasingly clear that the best available explanatory hypothesis is better than all alternative explanatory hypotheses that could be formulated.

7 Conclusion

A plausible version of Inference to the Best Explanation cannot simply claim one should infer the best explanatory hypothesis that is currently available. Rather, the inferred hypothesis must also be 'good enough'. But what makes a hypothesis 'good enough' in this sense? The answer, I have argued, is not that a 'good enough' hypothesis possesses some minimum balance of explanatory virtues, for that account turns out to be implausible, unworkable, and ill-suited to solve the problem. Instead, I have suggested that a hypothesis becomes 'good enough' when it has been through *explanatory consolidation*, a process in which direct empirical evidence for the hypothesis, and repeated failures to come up with plausible alternatives, gradually accumulate so as to support the optimistic explanation that the reason we haven't found a better explanatory hypothesis is that there is no such hypothesis to be found. Although this is a highly fallible process, to be sure, I have argued that it is both more epistemically secure and more descriptively accurate than the standard model of IBE in which one simply infers to the best explanation one has thought of so far.[33]

[33]This paper was greatly improved by helpful comments from Kevin McCain, Elmar Geir Unnsteinsson, Kareem Khalifa, and an anonymous reviewer for *Philosophy and Phenomenological Research*.

References

Achinstein, P. (2002). Is There a Valid Experimental Argument for Scientific Realism? *Journal of Philosophy*, 99, 470–495.

Armstrong, D. M. (1983). *What Is A Law Of Nature?* Cambridge: Cambridge University Press.

Baker, A. (2003). Quantitative Parsimony and Explanatory Power. *British Journal for the Philosophy of Science*, (pp. 245–259).

Baker, A. (2016). Simplicity. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*. Metaphysics Research Lab, Stanford University.

Barnes, E. (1995). Inference to the loveliest explanation. *Synthese*, 103, 252–277.

Bartelborth, T. (2005). Is the best explaining theory the most probable one? *Grazer Philosophische Studien*, 70, 1–23.

Beebe, J. R. (2009). The Abductivist Reply to Skepticism. *Philosophy and Phenomenological Research*, 79, 605–636.

Biggs, S. & Wilson, J. (2017). The A Priority of Abduction. *Philosophical Studies*, 174, 735–758.

Bird, A. (2017). Inference to the Best Explanation, Bayesianism, and Knowledge. In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 97–120). Oxford: Oxford University Press.

Brewster, D. (1833). Observations on the Absorption of Specific Rays, in Reference to the Undulatory Theory of Light. *Philosophical Magazine*, 2, 360–363.

Brewster, D. (1837). Review of Whewell's History of the Inductive Sciences. *Edinburgh Review*, 66, 110–151.

Buchwald, J. Z. (1989). *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Eearly Nineteenth Century*. Chicago, IL: University of Chicago Press.

Cabrera, F. (2017). Can there be a Bayesian explanationism? On the prospects of a productive partnership. *Synthese*, 194, 1245–1272.

Campos, D. G. (2011). On the distinction between Peirce's abduction and Lipton's Inference to the Best Explanation. *Synthese*, 180, 419–442.

Cantor, G. (1975). The Reception of the Wave Theory of Light in Britain: A Case Study Illustrating the Role of Methodology in Scientific Debate. *Historical Studies in the Physical Sciences*, 6, 109–132.

Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research*, 8, 133–148.

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.

Dawid, R. (2009). On the Conflicting Assessments of the Current Status of String Theory. *Philosophy of Science*, 76, 984–996.

Dawid, R. (2013). *String Theory and the Scientific Method*. Cambridge: Cambridge University Press.

Dawid, R. (2017). Bayesian perspectives on the discovery of the Higgs particle. *Synthese*, 194, 377–394.

Dawid, R. (2018). Delimiting the unconceived. *Foundations of Physics*, 48, 492–506.

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

Dellsén, F. (2017a). Abductively Robust Inference. *Analysis*, 77, 20–29.

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

Dellsén, F. (2018). The Heuristic Conception of Inference to the Best Explanation. *Philosophical Studies*, 175, 1745–1766.

Douven, I. (1999). Inference to the Best Explanation Made Coherent. *Philosophy of Science (Proceedings Supplement)*, 66, S424–S435.

Douven, I. (2017a). Abduction. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Summer 2017 Edition)*. Metaphysics Research Lab, Stanford University.

Douven, I. (2017b). Inference to the Best Explanation: What Is It? And Why Should We Care? In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 7–24). Oxford: Oxford University Press.

Foster, J. (1982). Induction, Explanation, and Natural Necessity. *Proceedings of the Aristotelian Society*, 101, 145–161.

Fumerton, R. (1995). *Metaepistemology and Skepticism*. Lanham, MD: Rowman & Littlefield.

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

Harman, G. (1997). Pragmatism and reasons for belief. In C. B. Kulp (Ed.), *Realism/Antirealism and Epistemology*. Lanham, MD: Rowman and Littlefield. Reprinted in **?**.

Henderson, L. (2014). Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science*, 65, 687–715.

Henderson, L. (2017). Bayesianism and IBE: The Case of Individual vs. Group Selection. In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 248–261). Oxford: Oxford University Press.

Hintikka, J. (1998). What is Abduction? The Fundamental Problem of Contemporary Epistemology. *Transactions of the Charles S. Peirce Society*, 34, 503–533.

Huemer, M. (2009a). Explanationist Aid for the Theory of Inductive Logic. *British Journal for the Philosophy of Science*, 60, 345–375.

Huemer, M. (2009b). When Is Parsimony a Virtue. *Philosophical Quarterly*, 59(216-236).

Ivanova, M. (2014). Is There a Place for Epistemic Virtues in Theory Choice? In A. Fairweather (Ed.), *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science* (pp. 207–266). Dordrecht: Springer.

Jansson, L. & Tallant, J. (2017). Quantative Parsimony: Probably for the Better. *British Journal for the Philosophy of Science*, 68, 781–803.

Kapitan, T. (1992). Peirce and the Autonomy of Abductive Reasoning. *Erkenntnis*, 37, 1–26.

Khalifa, K. (2010). Default Privilege and Bad Lots: Underconsideration and Explanatory Inference. *International Studies in the Philosophy of Science*, 24, 91–105.

Kuipers, T. (2000). *From Instrumentalism to Constructive Empiricism*. Dordrecht: Springer.

Kvanvig, J. (1994). A Critique of van Fraassen's Voluntaristic Epistemology. *Syn-*

these, 98, 325–348.

Lipton, P. (1993). Is the Best Good Enough? *Proceedings of the Aristotelian Society*, 93, 89–104.

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

Lipton, P. (2004). *Inference to the Best Explanation*. London: Routledge, 2nd edition.

Lycan, W. G. (1985). Epistemic Value. *Synthese*, 64, 137–164.

Lycan, W. G. (1988). *Judgment and Justification*. Cambridge: Cambridge University Press.

Maddy, P. (1997). *Naturalism in mathematics*. Oxford: Clarendon Press.

Maddy, P. (2007). *Second Philosophy: A Naturalistic Method*. Oxford: Oxford University Press.

McCain, K. & Poston, T. (2019). Dispelling the Disjunction Objection to Explanatory Inference. *Philosopher's Imprint*, 19(36), 1–8.

McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science*, 54, 553–567.

Minnameier, G. (2004). Peirce-Suit of Truth - Why Inference to the Best Explanation and Abduction Ought Not to Be Confused. *Erkenntnis*, 60, 75–105.

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

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

Norton, J. D. (2018a). *The Material Theory of Induction*, chapter 8: Inference the Best Explanation: The General Account. Unpublished manuscript available online at https://www.pitt.edu/~jdnorton/papers/material_theory/material.html.

Norton, J. D. (2018b). *The Material Theory of Induction*, chapter 9: Inference the Best Explanation: Examples. Unpublished manuscript available online at https://www.pitt.edu/~jdnorton/papers/material_theory/material.html.

Novick, A. & Scholl, R. (2019). Presume It Not: True Causes in the Search for the Basis of Heredity. *British Journal for the Philosophy of Science*, DOI: 10.1093/bjps/axy001.

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

Peirce, C. S. (1958). *The Collected Papers of Charles Sanders Peirce*. Harvard University Press.

Plutynski, A. (2011). Four Problems of Abduction: A Brief History. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 1, 227–248.

Poston, T. (2014). *Reason and Explanation: A Defense of Explanatory Coherentism*. New York, NY: Palgrave Macmillan.

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

Quine, W. V. O. (1963). On Simple Theories of a Complex World. *Synthese*, 15, 103–106.

Roush, S. (2005). *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Clarendon Press.

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

Salmon, W. C. (2005). *Reality and Rationality*. Oxford: Oxford University Press.

Schindler, S. (2018). *Theoretical Virtues in Science*. Cambridge: Cambridge University Press.

Schupbach, J. N. (2014). Is the Bad Lot Objection Just Misguided? *Erkenntnis*, 79, 55–64.

Schupbach, J. N. (2017). Inference to the Best Explanation, Cleaned Up and Made Respectible. In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 39–61). Oxford: Oxford University Press.

Sklar, L. (1981). Do Unborn Hypotheses Have Rights? *Pacific Philosophical Quarterly*, 62, 17–29.

Stanford, P. K. (2001). Refusing the Devil's Bargain: What Kind of Underdetermination Should We Take Seriously? *Philosophy of Science (Proceedings Supplement)*, 68, S1–S12.

Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Stanford, P. K. (2009). Scientific Realism, the Atomic Theory, and the Catch-All Hypothesis: Can We Test Fundamental Theories Against All Serious Alternatives. *British Journal for the Philosophy of Science*, 60, 253–269.

Swinburne, R. (1997). *Simplicity as Evidence for Truth*. Milwaukee: Marquette University Press.

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

van Fraassen, B. C. (1985). Empiricism in the Philosophy of Science. In P. M. Churchland & C. A. Hooker (Eds.), *Images of Science*. Chicago, IL: Chicago University Press.

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

van Fraassen, B. C. (2009). The Perils of Perrin, in the Hands of Philosophers. *Philosophical Studies*, 143, 5–24.

Weintraub, R. (2013). Induction and Inference to the Best Explanation. *Philosophical Studies*, 166, 203–216.

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

Whewell, W. (1837). *History of the inductive sciences, from the earliest to the present times*. London: J. W. Parker.

Worrall, J. (1989). Structural Realism: The Best of Both Worlds. *Dialectica*, 43, 99–124.

Wray, K. B. (2008). The Argument from Underconsideration as Grounds for Anti-Realism. *International Studies in the Philosophy of Science*, 22, 317–326.

Wray, K. B. (2011). Epistemic Privilege and the Success of Science. *Nous*, 46, 375–385.