SCIENTIFIC REALISM AND THREE PROBLEMS FOR INFERENCE TO THE BEST EXPLANATION

Abstract

Scientific Realism stands or falls with Inference to the Best Explanation. Realism cannot be accepted if one has reason to think that Inference to the Best Explanation cannot lead to the truth, or is unlikely to. Peter Lipton raises three important problems for his model of Inference to the Best Explanation: Voltaire's objection, Hungerford's objection, and the problem of Underconsideration. In this paper I show that Lipton's own solutions do not fully answer those problems. I argue that what is required to solve these problems is for our conception of explanatory goodness to be truth-conducive because it is sensitive to the way the world actually is. I suggest that the cognitive psychology of exemplars, as described by Kuhn, may provide an answer.

1 Introduction

Science makes frequent use of Inference to the Best Explanation (IBE). And IBE is central to the case made by Stathis Psillos (1999) and others for scientific realism: if we are to uphold scientific realism, it had better be the case that IBE is truth-conducive. Peter Lipton (2004) provides a model of how IBE operates. He then considers three objections to the claim that IBE is truth-conducive. If these objections are good then it is highly unlikely that IBE is truth-conducive. That would drive us towards scientific anti-realism.

Lipton provides his own solutions to the problems he raises. I argue that they are only partially satisfactory and that the sceptical worries the three objections raise remain. I indicate what I think is required for a satisfactory solution.

2 Inference to the Best Explanation and its problems

Inference to the Best Explanation (IBE) is about choosing among explanations. It is a matter of choosing among *potential* explanations of some phenomenon the one that is the best by certain criteria. If there is a suitable best potential explanation, IBE says that we may infer that it is the *actual* explanation, i.e. that the explanatory hypothesis is true.

According to Peter Lipton, IBE is a two-stage process, where both stages are filters of potential explanations (Lipton 2004: 56–64):

> Stage 1: The first stage filters out the implausible explanations. The imaginative capacity of scientists generates all the plausible potential explanations and just leaves the remainder unconsidered.
> Stage 2: At the second stage, scientists investigate the live potential explanations that have passed through the first filter, and ultimately rank

them according to their explanatory goodness, in order to select the top ranking explanation as *the* explanation.

Two qualifications need to be made concerning the second stage:

- For the best explanation to be inferred it should normally, considered on its own, be a *sufficiently good* explanation of *enough* evidence. If our best explanation is a weak explanation even of a large quantity of data (Lipton 2004: 63, 154), or explains only a limited amount of evidence well, then that is some reason to doubt that it is the actual explanation.

- For the best explanation to be inferred it must be significantly better than its nearest rival. If two competing explanations are both good enough, and one is slightly better than the other, our faith in that slightly better one must be slim. While Lipton does not mention this, it is a clear corollary of his account.[1]

Both stages in IBE raise important philosophical questions. A crucial question concerns the first stage. Since it filters out so many logically possible explanations, what confidence can we have that the actual explanation is allowed through? Why should the imagination of scientists have the capacity to pick on the true explanation among those it creates? This problem Lipton (2004: 152) calls 'Underconsideration'. The stage 2 ranking is no good at all if the actual explanation hasn't made it through stage 1 on account of the scientists' failure to think of it.

Assuming that the actual explanation is among those investigated at stage 2, two problems emerge, which Lipton calls 'Hungerford's objection' and 'Voltaire's objection'. Hungerford's objection raises the worry that what we consider to be the goodness of explanations (which Lipton calls 'loveliness') may be too subjective to have any relationship to the truth. However, even if explanatory goodness is objective, there will be many worlds where it does not correlate with truth. Voltaire's objection says that it implausible that our world is the best possible world by those standards, which it would have to be for the best explanation to be true.

In this essay I will first articulate the thee problems—Underconsideration, Hungerford's objection, and Voltaire's objection—in more detail and explain why I find Lipton's own solutions to these problems to be only partially satisfactory. Thereafter I will show how we may find a solution to these problems in the cognitive psychology of scientific research, first articulated in Thomas Kuhn's notion of an exemplar. The latter, suitably developed, will allow us to see why it is plausible that our standards of explanatory goodness are objective and likely to correlate with the truth.

3 The problem of Underconsideration

No matter how accurate the ranking is at stage 2, the top ranked theory will not be true if the true theory is not among those theories selected at stage 1. The worry is that the theories to which we have actually given conscious consideration are a small subset of the range of all possible theories, and so the chance of our even thinking

[1] Some might complain that even this is not enough. If one hypothesis is still 'live', being consistent with the evidence, can we *know* that any rival hypothesis is true? My own view is that we cannot (Bird 2005a, 2007b, 2010). But I shall not insist on that in this paper. This second qualification and my eliminativist doubts may be set aside if the purpose of the ranking is to infer that the top ranked hypothesis is true, but in order to assess the plausibility of each hypothesis

of the true theory is correspondingly negligible. As Lipton (2004: 152) puts it, if that is correct, then one's thinking that the theory ranked as best at stage 2 is true, is like thinking that Jones will win at the Olympics when all one knows is that he is the fastest miler in Britain. We may give the worry some bite by drawing on a version of the pessimistic meta-induction. We think that scientists of previous eras got things wrong (geocentricism, miasma theory of disease, Newtonian space, phlogiston theory of combustion etc.). When such theories were first conceived and then adopted the currently accepted theories were not even considered as possible alternatives.[2] So it seems at least plausible to doubt that when scientists think of the possible explanations of some phenomenon they are in general able to generate the true theory among them.

Lipton (2004: 152), who develops the problem from van Fraassen (1989: 143), presents it as an argument with two premises:

> (R) (the *ranking* premise): "The testing of theories yields only a comparative warrant." That is, the process of ranking at stage 2 reliably ranks more likely theories higher than less likely theories. But it does not tell us how likely any of these theories is.
> (N) (the *no-privilege* premise): "Scientists have no reason to suppose that the process by which they generate theories for testing makes it likely that a true theory will be among those generated." The true theory may simply not be generated at all, and there is no way of knowing how likely that is.

From which the following conclusion is drawn:

> Conclusion: "While the best of the generated theories may be true, scientists can never have good reason to believe this."

Lipton's strategy in responding to Underconsideration is to argue that the ranking premise undermines the no-privilege premise. He emphasizes that the process of ranking will avail itself of background theories. Here is an example (my own). Why was Luis and Walter Alvarez's impact theory of the K-T extinction held to be a good explanation? One reason was that it explained the iridium anomaly—an unexpectedly high concentration of iridium at the geological K-T boundary; and it could explain that thanks to a background theory which tells us that comets and asteroids have high abundance of iridium, which is rare on Earth. Without such a background theory, the impact theory would have been a less good explanation of the data.

If the background theories employed in ranking were false, then the ranking process would not be reliable. So the reliability of the ranking process implies that we often have true background theories. Lipton then notes that our background theories are themselves products of earlier processes of IBE; in which case those processes did produces true theories. From which is follows that we can reject the no-privilege premise, (N), because it is frequently the case that we have considered and selected the true theory.

Lipton's response shows that if our ranking is reliable, then IBE as a whole is reliable. But I do not believe that he shows that Underconsideration is a self-undermining argument. For the role of the ranking premise, (R), in generating the conclusion is minimal. (N) entails the conclusion on its own, which can be seen straightforwardly by reflecting that if the conclusion were false—scientists did have

[2] Kyle Stanford (2006) develops this line of argument against scientific realism in detail.

reason for thinking that the best generated theory is true—then (N) would be false—those scientists would thereby have reason for thinking that the theory generating process does frequently generate the true theory.

Let us look at this in more detail. (R) may be divided into two components:

(R1) Ranking does not give an absolute measure of likeliness.

and

(R2) Ranking does give a reliable measure of comparative likeliness.

(R1) is a corollary of (N). It tells us that we do not get an absolute measure of theory likeliness from ranking. If we did get an absolute measure, then we would know how likely it is that we have considered the true theory. So it is not the case that (R1) plays a role in generating the conclusion of the Underconsideration argument. Rather both (R1) and the conclusion are consequences of (N) on its own.

Lipton's intended argument against Underconsideration aims to undermine (N) by pointing out that (N) is inconsistent with (R2), the claim that ranking does reliably assess comparative likeliness. Hence Lipton's argument will work only if (R2) is indeed part of the Underconsideration argument. But, as emphasized above, (R2) plays no role at all in generating the conclusion of the Underconsideration argument. So the fact that (R2) undermines the ranking premise does not show that the argument for Underconsideration as presented is self-defeating.

Lipton (2004: 158) does note that the proponent of the Underconsideration argument can avoid his response by weakening the ranking premise, but replies, "Of course, if ranking were completely unreliable, the skeptic would have his conclusion, but this just takes us back to Hume. The point of the argument from Underconsideration was rather to show that skeptical conclusion follows even if we grant scientists considerable inductive powers." However, I do not see how dropping (R2) would 'just takes us back to Hume'. Yes, the conclusion, 'we have no reason to believe that our inductive practices yield the truth', is much the same as Hume's conclusion. But what is important is the *argument* for the conclusion, and the sceptical concern with IBE raised by the no-privilege premise is quite different from anything in Hume's problem, which may be regarded a stage 2 problem rather than a stage 1 problem. Underconsideration worries that we might not have thought of the correct hypothesis. Hume's problem worries that even if we have thought of the correct hypothesis, any attempt to prove it will be question-beggingly circular. Lipton shows that a solution to the stage 2 problems implies a solution to the stage 1 problem. But that does not entail that the stage 1 problem reduces to the stage 2 problem; Lipton's argument can be contraposed: if there is no solution to the stage 1 problem, there is no solution to the stage 2 problem. The problem of Underconsideration *adds* to the problems surrounding stage 2. Although Lipton explicitly presents the ranking premise, and its component (R2), as a premise in the Underconsideration argument, the quotation given in the first sentence of this paragraph suggests instead that (R2) is supposed to be a concession made by the sceptic, a concession that emphasizes the power of the Underconsideration argument. In effect, the sceptic is taken to be saying, "Let us agree that stage 2 is justified, insofar as it is correct that if theory A is a better explanation than theory B, then A is more likely to be true than B. Still, that is no reason to believe that the best ranked theory is true, since we have no reason to believe that we have considered the true theory when we carried out the ranking." We should conclude from Lipton's argument that the sceptic should *not*

assert *that*. If ranking is reliable, then IBE is safe from the problem of Underconsideration. But as we have seen the sceptical proponent of the Underconsideration argument doesn't *need* to claim that ranking is reliable. A perfectly plausible position for the IBE sceptic is to claim that we have no idea how reliable ranking is, just as we have no idea how likely it is that our generating process generates the true theory among those is puts forward for consideration. Lipton is right that Underconsideration fails as an intermediate scepticism—conceding to us the capacity of (relative) inductive ranking while denying absolute ranking and so knowledge. Nonetheless it remains as a full-blooded sceptical problem, and a distinct one from Hume's problem. For note that the Underconsideration sceptic can still concede that *if* the scientist were in general able to think of the true hypothesis among those she considers *then* she would be able to use IBE to discern which one that is—a concession that the Humean sceptic would deny. So Underconsideration is a distinctive sceptical problem that still needs to be answered.

4 Hungerford's objection

The final lines of Keats's 'Ode on a Grecian Urn' tell us:

> 'Beauty is truth, truth beauty,—that is all
> Ye know on earth, and all ye need to know.'

These lines capture the central insight of Inference to the Best Explanation. Keats exaggerates poetically in saying that truth and beauty are identical. According to IBE they are *correlated* (when we take 'beauty' to refer to the good-making features of a explanation). Hungerford's objection is so-called since it is encapsulated in Margaret Hungerford's famous line in *Molly Bawn*, that 'beauty is in the eye of the beholder'. If we put Keats and Hungerford together we get the conclusion that truth is correlated with the subjective perception of beauty. Unless relativism about truth of a radical sort is correct, that cannot be right. For the realist about truth, Hungerford's objection is simply the worry that our standards of explanatory goodness (beauty) are too subjective to have any correlation with something as objective as truth.[3] Helge Kragh (1990: 287) articulates it thus:

> The principle of mathematical beauty, like related aesthetic principles, is problematical. The main problem is that beauty is essentially subjective and hence cannot serve as a commonly defined tool for guiding or evaluating science.

Arguably some subjectively determined properties might correlate with objective features of the world: if it is correct that a person's experience of colour is subjective, then that would be a subjective quality that does correlate with an objective property. On the other hand, so Hungerford's objection goes, we have reason to believe that judgments of explanatory goodness are rather more like judgments of beauty that the experience of colour. For one thing, the terminology used to articulate notions of explanatory goodness, or aspects thereof, are taken from the aesthetic realm: loveliness, beauty, elegance. So the claim is that explanatory goodness

[3] It should *not* be thought that what I am calling 'explanatory goodness' coincides with the 'beauty' of a scientific theory, although the latter may be a component of the former, perhaps an important one. Walker (2012) renames Hungerford's objection 'the subjectivity objection' in order to avoid this conflation.

is subjective in the respect that aesthetic properties are, which makes them unlikely to correlate with the truth.[4]

Lipton (2004: 143–4) responds to Hungerford's objection by pointing out that while there is interest-relativity in our assessment of the goodness of explanations, that relativity is to be welcomed, since it correlates with the interest-relativity of IBE itself. For example, we expect audience relativity, since different people have different evidence and background beliefs. IBE is also relative to the interests of the audience, since different interests determine different contrasts (Lipton argues that explanation is contrastive). I might be asked to explain why I flew to Vienna; but that request for an explanation could be intended in more than one way: to explain why I came by plane rather than by train, or why I came to Vienna rather than visiting Munich or staying at home, or even why *I* was invited to Vienna rather than some other, more illustrious person. A difference of interests might determine a different intended contrast amongst these possibilities, and hence what is to be explained and so what facts are inferred by IBE.

While I agree with Lipton that IBE does show these forms of relativity, and that they are no threat to the objectivity of IBE, I also maintain that they do not get to the heart of the 'beauty is in the eye of the beholder' objection. One way to see this is to note that Hungerford's objection might still be raised in a case where none of the relativity to which Lipton refers arises. We may find a case were we are focussing on one specific contrast and where the relevant evidence is agreed on by all; if in such a case the competing hypotheses are ranked according to some feature F, where the possession of by a theory of F is a matter of subjective opinion, one would doubt that the ranking produced would correlate with likeliness of truth.

In times past, scientific ideas were often expressed in poetic form—Lucretius's *De Rerum Natura* is a prime instance.[5] Imagine that in some community a theory is often preferred (as regards truth) to another because the poetry of the first is deemed aesthetically superior to that of the latter. Clearly that would be a poor basis for a theory preference. The challenge of Hungerford's objection is that the judgments of our scientists in using IBE have something in common with such a community. Even if our scientists' judgments are not transparently subjective, they are subjective nonetheless. Note that the subjectivity implied by Hungerford's dictum 'beauty is in the eye of the beholder' is typically individual—a building that you find beautiful I may find ugly. However, the subjectivity of Lipton's problem is public and shared. The community of scientists is usually agreed on which explanations they find lovely. The objection is not that explanatory goodness is personal, but rather that it is something more like a fashion or the taste exhibited and extolled in a particular historical era. There is agreement but like a shared aesthetic response it is too shifting and ungrounded a basis for evaluating objective truth. Thus ancient

[4]Consider the claim made by some evolutionary psychologists, that our ideas of human physical beauty correlate with health, fertility and other properties that go towards biological fitness, the propensity to survive and reproduce (e.g. Grammer et al. 2003). One objection made to this proposal is that ideas of beauty are too subjective to permit such a correlation. Different individuals have different views about what they find attractive; what people find beautiful may vary with their individual circumstances, with fashion, and with their cultural background. If so, our sense of beauty is too subjective to be any indicator of something objective, such as biological fitness. I am not endorsing this objection—maybe there less variability in perceptions of beauty than one thinks, perhaps there is an unvarying core to our sense of beauty surrounded by a varying penumbra—rather I am pointing to the form of this argument, which it shares with Hungerford's objection.

[5]See Taub (2008) and Timmermann (2013) for discussion of scientific verse in ancient and late medieval periods respectively.

and medieval astronomers may have been moved by the aesthetically pleasing nature of uniform circular motion whereas in the Newtonian era elegance was found not in the motions of planets or other bodies but in the simplicity of the equations governing them. The beauty of modern physics is for many found in something else, the symmetries that the theories embody.[6] And if we go beyond physics, to biology for example, what counts as a desirable character of a theory is different again.

How worrying is Hungerford's problem for IBE? That rather depends on how plausible one finds the accusation of subjectivity in the sense just articulated. Are our judgments of explanatory goodness analogous to judgments of aesthetic value, albeit less transparently subjective? I suggest that the following are prima facie reasons to think that they might be and which need to be addressed by any response to the problem:

i. *Affect* While a scientist's assessment of a theory will typically be informed by a conscious, rational assessment of the theory and its relationship to the evidence, ultimately her judgment of whether the theory is a good or poor explanation will be a matter of affect—her inner feeling about the theory. Is the theory an elegant explanation of the evidence or is it contrived? There is no methodology for making such an assessment—once all the cogitation and thinking is done, that is just a matter of how it feels to the scientist. (This is no different from the arts. A critic may discuss various aspects of a work of art, but whether it is an aesthetically successful work is a further judgment, a matter of the individual's aesthetic response.)

ii. *Pleasure* A key element of the inner response is pleasure (or its absence). In the arts, it is a *positive* affect that drives a positive judgment of aesthetic quality—the work should be satisfying or pleasing. Likewise, in the sciences, a scientist's response to an elegant explanatory theory is one of pleasure, of finding it satisfying.

iii. *Ineffability* Philosophers and others have difficulty in saying what exactly explanatory goodness is; there is no consensus on what it is or what contributes to it. Simplicity, for example, may be thought to be a quality that is widely agreed to be a component of explanatory goodness, but even then philosophers disagree about whether it is an epistemic virtue of explanations or a merely pragmatic one. This difficulty in articulating the nature of explanatory goodness suggests that it had the same ineffability as subjective qualities such as aesthetic beauty.

iv. *Variability* Even when a set of more concrete criteria is proposed, there is variation in what counts as exemplifying those criteria. Furthermore there is variation in the weight or significance attached to the criteria. For example, Kuhn (1977: 321–2) lists five values prized across scientific disciplines and eras—accuracy, consistency, scope, simplicity, and fruitfulness, but he argues that there is variation in what counts are exemplifying these values and how significant they are relative to one another. This variability across time (and sometimes synchronically, as in the case of disagreement during a scientific revolution) is what one would expect if explanatory goodness operates much like taste or fashion.

[6] 'Symmetry denotes that sort of concordance of several parts by which they integrate into a whole. Beauty is bound up with symmetry' (Weyl 1952).

v. *Terminology* When philosophers do attempt to articulate their differing conceptions of explanatory goodness, they tend to use terms that refer to subjective qualities. Lipton talks of 'loveliness', McAllister (1999) of 'beauty', and Glynn (2010) of 'elegance'. These terms as well as others such as 'harmony' may be found in use among scientists also when they articulate the merits of a theory.

In summary, it is a plausible proposal that the central notion of IBE, 'goodness', is assessed by subjective criteria, in which case IBE is on epistemically shaky ground. Hungerford's objection still stands in need of a robust response.

5 Voltaire's objection

Hungerford's objection provides one ground for thinking that judgments of explanatory goodness are unlikely to correlate with the truth. Voltaire's objection provides a different and independent reason for thinking that such a correlation is implausible.

Let our standards of explanatory goodness be such that we prefer explanations that have quality S over those that do not, and those that have a high degree of S to those that have a low degree. For IBE to be reliable the world must be such that high-S explanations are more likely to be true than low-S explanations, and in particular that explanations with the maximum degree of S are frequently true. Call such a world an 'S-world'. Voltaire's objection contends that we do not have any reason, for supposing that the actual world is an S-world. Imagine a community similar to that in the preceding section where scientists prefer theories expressed in dactylic hexameters to those in heroic couplets. However, they do so not for any aesthetic preference for the former, but simply because this is a well-established standard of this community. Hungerford's objection no longer applies: whether a verse is in dactylic hexameters is an *objective* feature of the verse. So there could be a correlation between theories in dactylic hexameters and the truth. But why should there be? Clearly we would not expect any such correlation—we do not believe that the actual world is a dactylic-hexameters-world. But what reason to we have for believing that the actual world is an S-world, where S-standards are the standards we actually use?

Lipton does not present a solution to Voltaire's objection as such but argues that Voltaire's objection in effect reduces to Hume's problem of induction. The way I have expressed Voltaire's problem in the preceding paragraph suggests this. Our use of IBE assumes that our world is a high-S world, just as our use of enumerative induction assumes that our world is a largely uniform one. But it is difficult to see how we can justify such assumptions without appealing either to the very same assumption or at least something similarly intractable.

However, in my view this undersells Voltaire's objection. By giving the objection this name, Lipton implies is suggesting that the objector claims that IBE proponent is like Dr Pangloss in Voltaire's *Candide*. Dr Pangloss holds the Lebnizian view that the actual world is the best of all possible worlds, "Dans ce meilleur des mondes possibles, tout est au mieux". Here 'best' means best in terms of moral and physical good and evil. Voltaire satirizes the view by pointing to the evidence for the contrary, such as the terrible earthquake at Lisbon in 1755. However, there is another objection to such a view. Note that there are so many possible worlds which differ from our own world to a lesser or greater degree. So the proposal that the actual world is

the *best* of all these is to propose that the actual world is very special indeed. It is not merely an adequate world, nor even a fairly good world, it is the *best* world. And that just seems implausible. It would be an enormous fluke that we live in the best of possible worlds whereas our very similar counterparts in other possible worlds do not. The IBE enthusiast is likewise betting on the world being the best of all possible world in terms of *explanatory* goodness. Which is also to make the actual world a very special world, and is likewise implausible. There must be many worlds in which the correct explanations are often ranked second or third best or even lower. So why is the actual world not one of them, but is instead this very special world? If our S-standards are set *a priori* (as in Voltaire's day our moral standards were held to be), the it would be a fluke that our world meets the S-standards to the highest or even a very high degree, when most worlds do not.[7]

There is therefore a difference from Hume's problem. What does a world need to be like in order for inductive projection in that world to be truth-conducive? Roughly a world should be such that in a good proportion of cases, observed regularities persist; or, a little more precisely, the world should be such that our sampling procedures should not render our samples unrepresentative of the populations from which they are drawn. Let us call such worlds *projectible*. They contrast with unprojectible worlds where although we observe regularities, they break down in future or other hitherto unobserved or unsampled cases. (We can ignore as irrelevant largely irregular worlds where there are few partial or complete interesting regularities to be observed at all.) The supposition that the actual world is a projectible world does not seem to be a Panglossian supposition. It requires the actual world to be different from the unprojectible worlds, but it does not require the world to be special or better than *all* (or almost all) the rest. The projective inductivist can hold that the actual world is just one of many projectible worlds and is nothing special in this respect. On the other hand, IBE does require the actual world to be special. When we infer that the *best* explanation is true, then for that inference to be correct, it must be the case that the actual world is the *best* by our explanatory standards. Walker (2012: 66) puts the point in terms of laws: 'The defender of IBE must show that, of all the lovely law-governed possible worlds, the actual world has the loveliest laws.' That requires the actual world to be unique. Perhaps being close to best will be enough for warranted belief or some knowledge. But just being good (but not close to being the best) by those standards will not be enough. So, compared to projective induction, IBE puts a rather stronger requirement on the way the world must be in order for its inferences to lead to knowledge or justification.

If that is right, then Voltaire's objection may remain even if we feel we have a solution to Hume's problem. Hume's problem kicks in only with the demand that in using projective induction we must also show that the world is a projectible world. Reliabilist and other externalist epistemologists will reject that demand, asserting that can have projective knowledge or warrant if we happen to inhabit a projectible world, and make a correct projective inference—it is not necessary to give a further justified reason for thinking that we do indeed inhabit such a world (Mellor 1991). Reliabilism can also help IBE, but only to some extent. If we do indeed inhabit a world where the best explanations are true, then IBE will be reliable, and so we can have warrant, rational belief, and perhaps even knowledge, by use of IBE. In both cases the reliabilist is able to defeat the sceptic who asserts 'knowledge and ratio-

[7] IBE might be reliable if the actual world is not the best world, but is very similar to it. The objection remains, since it would require the actual world to be one of a small and very special set of all possible worlds.

nal belief are not *possible*'. But Voltaire's objection raises a worry that seems to be consistent with accepting this reliabilist refutation of scepticism. For the Voltairean may respond, "I accept that knowledge from IBE is possible, since it is possible that we inhabit the best of all possible worlds. But it is just *so implausible* that we live in that optimal world, that we ought to conclude that as a matter of contingent fact, it is highly unlikely that IBE is reliable; and it is therefore unlikely that IBE gives us knowledge or even rational belief."

6 What is needed to defend IBE?

In this section I consider what kind of response is needed that would defend IBE again the three objections. Note that we are in the territory of cognitive psychology. For we are asking about our human ability to think up (potential) explanations. We are asking whether we are able to exercise our imaginations in a way that bears an appropriate relationship to the truth. What, then, do the objections to IBE tell us about explanatory goodness, such that it can be a guide to the truth? Hungerford's objection tells us that explanatory goodness must be objective: it must be a quality that could correlate with the truth. Voltaire's objection requires that we can explain why the actual world is a lovely one, a world where 'good' explanations tend to be true. That suggests that our standards of explanatory goodness cannot be apriori but are somehow answerable to the way the world is. Finally, Underconsideration shows that the methods by which we select our hypotheses for consideration can direct us, in a good proportion of cases, to hypotheses that are likely to be true, thereby making it likely that the true explanation is among those we consider. Hypotheses selection cannot be a random walk in the logical space of possibilities, but must be directed towards the actual explanations.

When we say that explanatory goodness must be objective, and so not subjective, we need to be clear what we mean by 'subjective'. In the relevant sense, a property Φ is subjective when the truthmaker for some a's being Φ is a subject's affect in response to a, rather than some intrinsic property of a itself. When Mrs Hungerford wrote that 'beauty is in the eye of the beholder', she meant that whether some person (Eleanor Massereene in *Molly Bawn*) is beautiful is a matter of the attitude of the beholder, not an objective property of the person beheld. As Hume (1757) puts it, 'Beauty is no quality in things themselves: It exists merely in the mind which contemplates them; and each mind perceives a different beauty.' On the other hand, some objective, intrinsic properties of things might be detected by subjective experiences. So one way to understand secondary properties is as intrinsic dispositional properties of things to produce responses in observers. An instance of this view says that 'red' denotes an objective, intrinsic quality of objects, a disposition to cause certain ('normal') observers to have 'red' experiences.[8] So while responding to Hungerford's objection rules out explanatory goodness being a subjective property in the former sense (like beauty, according to Hungerford and Hume), it does not need to rule out explanatory goodness having the connection with subjectivity or response-dependence that redness has (when understood as a secondary quality).

[8] In this territory lies the murky issue of response-dependent concepts and whether and in what sense they denote subjective or objective properties. See Rosen (1994), Wedgwood (1997), and Haukioja (2013).

Heuristics and gut feelings

I believe that it is possible to find an account of explanatory goodness that responds satisfactorily to the demands of the three objections. Goldstein and Gigerenzer (2002)) discuss the recognition heuristic, a cognitive mechanism often exemplified in answering the following question: 'Which city has the larger population, Detroit or Milwaukee?' German students more frequently answered this question correctly than American students, despite knowing less about the cities than the Americans. In fact many of the German students had not heard of Milwaukee. And this is why they did better. Unconsciously they are using the recognition heuristic, which one might articulate thus: If you recognize the name of one city but not that of the other, then infer that the recognized city has the larger population. Because the heuristic is unconscious, its phenomenology is like that of intuition. When the heuristic is at work, the subject responds 'Detroit, surely' without knowing why she thinks that is the correct answer. Gigerenzer (2007: 16) uses the term 'gut feeling' to describe a judgment '1. that appears quickly in consciousness. 2. whose underlying reasons we are not fully aware of, and 3. is strong enough to act on'.

The relevance of gut feelings for my response to the problems of explanatory goodness is this. Gut feelings have an element of subjectivity: they are judgments based on feelings or intuitions. They are nonetheless judgments that are aimed at determining an objective fact (e.g., which of two named cities is larger). They are able to connect the subjective feeling with the objective fact because the cognitive mechanism behind the judgment, the unconscious heuristic, is able to access what the subject knows (e.g., recognizing the name of one city, but not the other). So it is possible for a cognitive mechanism to deliver judgments about objective facts with some degree of reliability, but which are also subjective in the weaker sense discussed above. A solution to the problems of explanatory goodness should therefore not be impossible.

Cognitive psychologists from Bartlett (1932) onwards have described a number of mental mechanisms, such as schemata, scripts, frames, and analogies, that are in the same broad category as the heuristics that generate gut-feeling judgments. As with the recognition heuristic, the use of these mechanisms is partly or wholly unconscious and draws on our knowledge and experience of the world to generate judgments and decisions. A further important feature of such mechanisms is that they are typically acquired through experience. While Gigerenzer is concerned to show that some heuristics are innate, being the product of evolution alone, others are developed in the course of our interactions with the world (the recognition heuristic could not be entirely innate, for example). So the judgments we form are not strictly intuitive, since they are learned responses—they are what I have called 'quasi-intuitive (Bird 2005b, 2007a)'. They are nonetheless phenomenologically like intuition in that the lead at least partly unconsciously to a judgment of the form 'this feels right'.

Exemplars

The mechanism on which I draw in answering the problems of explanatory goodness is the *exemplar*. The term 'exemplar' is used by Kuhn (1970) to articulate one of the two senses in which he used the broader term 'paradigm'. An exemplar is an exemplary solution to a scientific problem that serves as a model for subsequent science in the relevant field. Young scientists acquire their understanding of exem-

plars in large part by practising problem-solving with them—at first simple problems, then more complex ones. In so doing they acquire the ability to see a new problem as fundamentally similar to an exemplary puzzle and therefore requiring the same kind of solution (e.g. a vibrating string as similar to a pendulum and therefore as exhibiting simple or damped harmonic motion). This 'seeing' is essentially a matter of pattern recognition (where the pattern may be an abstract rather than a visual one). Pattern recognition, an ability typically acquired through repeated exposure and practice, is always in part and sometimes entirely an unconscious process as well as a cognitive one (think of recognizing a face, or correctly identifying the composer of an unfamiliar piece of music).

Kuhn (1977: 321–2) also tells us that in acquiring understanding through training with exemplars, a scientist also acquires the shared values of her field. As mentioned above, Kuhn holds that there are, in the abstract, values that are shared across all of science (primarily accuracy, consistency, scope, simplicity, and fruitfulness). Nonetheless, what features count as exemplifying these values, as well as how they are to be weighed against each other, varies from one field to another. My proposal is that this is how a scientist's sense of the quality of an explanatory theory is acquired. Indeed, I suggest that the best way to understand explanatory goodness is as a summary judgment of the value of a theory, to which the particular values mentioned contribute. Thinking about explanatory goodness in this way allows us to reply to the three objections.

Exemplars and Hungerford's objection

Above I argued that Hungerford's objection, that explanatory goodness is too subjective to correlate with truth, appears plausible because (i) we judge it on the basis of affect, an inner response to a theory; (ii) we take pleasure when we encounter it; (iii) it is ineffable; (iv) it shows variability across different fields and times; and (v) we use aesthetic terminology to describe it. While these are strongly suggestive of a subjectivity shared with aesthetic qualities, they are consistent also with the weaker kind of subjectivity that can correlate with objective properties. As we saw, the gut feeling of a judgment formed by an unconscious heuristic has the latter kind of weak subjectivity. Judging the explanatory goodness of a potential explanation, informed by training with exemplars, is like a gut feeling in this respect, which explains feature (i). The feeling of rightness is also a positive one. The exemplar is held up as a paradigmatic example of how science ought to be; recognition that a hypothesis is like the exemplar provides intellectual pleasure—(ii). Because such judgments are based on a hypothesis 'feeling right', only partially informed by conscious reflection, it is difficult for a subject to articulate the basis of this judgment and to say what explanatory goodness is—(iii). Because in different fields and at different periods, there are different paradigms—exemplars—in operation, what counts as a good explanation will vary over time and from one scientific field to another—(iv). These aspects, (i)–(iv) are also found in aesthetic qualities, and aesthetic qualities are the most prominent instances of properties displaying (i)–(iv), so it is not surprising that we use aesthetic terminology to describe them.

Exemplars and Voltaire's objection

Voltaire's objection was, put simply, that if our standards of explanatory goodness are apriori, then it would be an implausible coincidence that our world, given all

the ways a world could possibly be, meets those standards so closely that very frequently good explanations are true. The lesson to draw from that objection is that standards of explanatory goodness are not apriori. And if standards of explanatory goodness are gained from training with exemplars, then indeed they are not apriori. Of course, that our standards are gained from exemplars does not guarantee that they do correlate with the truth. For if our exemplars are themselves badly mistaken, the standards of explanatory goodness gained from them will not be truth-conducive—much medieval science was in this position. However, if the exemplars of a field are true or highly truthlike, then the standards of explanatory goodness they generate will be truth-conducive. And since it would not be an implausible coincidence that such exemplars are true, then it does not require an implausible coincidence that true theories have a high degree of explanatory goodness.

Exemplars and Underconsideration

Underconsideration raises the challenge of explaining why it is that our methods of theory selection at Stage 1 make it likely that the true theory is among those selected for consideration. Kuhn emphasizes the role that exemplars play in discovery, the process of generating hypotheses. Training with exemplars enables scientists to see the world in a particular way, which is to say that some features of the world become scientifically salient and others less so. It also means that certain explanatory schemata suggest themselves. In the Aristotelian paradigm motion is salient, and an Aristotelian scientist will naturally seek an explanation of motion in terms natural tendencies in the object itself. Whereas a Newtonian scientist will not find rectilinear motion itself salient, but will regard changes in motion or non-rectilinear motion as salient, and so she will look for the explanations of these in forces governed by general laws. So the explanations we reach for are not randomly chosen from among all possible explanations, but will be ones that are similar or analogous to explanations with which we are already familiar. When our exemplars are erroneous, as in the Aristotelian case, then we probably will not consider the correct explanation. But when our exemplars are correct, then they make it likely that the true explanation will be among those we in fact consider.

7 Conclusion

Inference to the Best Explanation is the form of reasoning employed by many of our best scientific theories. (It is also central to the No Miracles Argument for scientific realism, though I regard that as less important.) If, therefore, there are general arguments that suggest that IBE cannot be reliable, then such arguments are a threat to scientific realism. Peter Lipton articulates three such arguments. In this paper I have explained why I do not find Lipton's own responses to those arguments convincing. The problems he raises still need solutions.

In my opinion, we do not learn from sceptical arguments that we do not have knowledge. For our conviction that we do indeed have knowledge (of the external world, of some science, etc.) should be greater than our confidence that the sceptical philosophical arguments are sound. Instead, by looking for and finding correct responses to the sceptical arguments, we learn something about the nature of knowledge and of our knowledge-producing processes. So in this paper I have tried to indicate what we learn about IBE from Lipton's three problems. We learn

that explanatory goodness is subjective only in the weak sense. It is a property that although it reveals itself to us as an inner state or feeling of rightness, is nonetheless responsive to the world, so that it could correlate with the truth. Secondly, our standards of explanatory goodness are not apriori. Instead they are themselves responsive to the world, which it is why it is not an implausible coincidence that true theories should be meet those standards. Finally, we learn that the way in which we choose our hypotheses for investigation should not be a random selection from the space of possible hypotheses but should instead be directed towards the hypotheses with higher chances of being true.

It is, however, one thing to say that a correct account of explanatory goodness should show that it is only weakly subjective and that our standards of explanatory goodness should be responsive to the world, and quite another thing to show how this is in fact the case. I have suggested that one place to look for am answer is Kuhn's idea that the processes of scientific cognition are driven by exemplars. Exemplars are our paradigms of good science—they show what good science should look like and so become the source of our scientific values—our standards of explanatory goodness. Because our deployment of exemplars and so of our standards of explanatory goodness are quasi-intuitive—like intuition, but learned from experience—our sense of explanatory goodness is subjective, but only weakly. And in the not implausible circumstance that our exemplars are correct scientific hypotheses, it will not be a coincidence that explanatory goodness is indicative of the truth.

It might seem incongruous that I am using Kuhnian ideas to explain how it is possible for explanatory goodness to be truth-conducive. In *The Structure of Scientific Revolutions* Kuhn is very interested in the individual and social psychology of scientific change—there are more references to psychologists than philosophers. Although he later expressed anti-realist views about truth, in the first edition of *Structure*. Kuhn does not discuss truth nor does he say anything else that directly implies anti-realism, and the book was well received among practicing scientists. It is only after criticism from Lakatos and others that Kuhn came to be thought of as an anti-realist. And Lakatos himself had a quite specific, Popper-inspired, conception of what scientific rationality amounts to. In particular, Kuhn's interest in the psychology of scientific thought was particularly excoriated by Lakatos. Times have since changed (even if the suspicion that *Structure* is anti-realist has not). What Lakatos (1970: 178) ridiculed as Kuhn's appeal to 'mob psychology' would now be regarded as naturalistic social epistemology.

Finally, we should be clear about what this 'defence of scientific realism' amounts to. Even if correct, it does *not* mean that we have shown that some strong thesis of scientific realism is correct: that the successful theories of science are true or highly truthlike. Rather, this paper has a weaker, more defensive intent. If the three problems were correct, then one could not take a realist attitude to any theory supported by IBE (let alone to science in general). So this defence removes that sceptical threat. It does not thereby show that a successful scientific theory *is* true. Indeed, as we have seen, the exemplar story is consistent with exemplars being false and providing a misleading standard of explanatory goodness. Nonetheless, it is important to know that our scientific processes *can* be truth-conducive. where they actually are depends on the details of each particular case. That's good enough and is as much realism as we should expect from philosophy.

References

Bartlett, F. 1932. *Remembering*. Cambridge: Cambridge University Press.

Bird, A. 2005a. Abductive knowledge and Holmesian inference. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, pp. 1–31. Oxford: Oxford University Press.

Bird, A. 2005b. Naturalizing Kuhn. *Proceedings of the Aristotelian Society* **105**: 109–27.

Bird, A. 2007a. Incommensurability naturalized. In L. Soler, H. Sankey, and P. Hoyningen-Huene (Eds.), *Rethinking Scientific Change and Theory Comparison*, Volume 255 of *Boston Studies in the Philosophy of Science*, pp. 21–39. Dordrecht: Spinger.

Bird, A. 2007b. Inference to the only explanation. *Philosophy and Phenomenological Research* **74**: 424–32.

Bird, A. 2010. Eliminative abduction—examples from medicine. *Studies in History and Philosophy of Science*: 345–52.

Gigerenzer, G. 2007. *Gut Feelings. Short cuts to better decision making*. New York, NY: Penguin.

Glynn, I. 2010. *Elegance in Science*. Oxford: Oxford University Press.

Goldstein, D. G. and G. Gigerenzer 2002. Models of ecological rationality: The recognition heuristic. *Psychological Review* **109**: 75–90.

Grammer, K., B. Fink, A. P. Møller, and R. Thornhill 2003. Darwinian aesthetics: sexual selection and the biology of beauty. *Biological Reviews* **78**: 385–407.

Haukioja, J. 2013. Different notions of response-dependence. In M. H. B. Schnieder and A. Steinberg (Eds.), *Varieties of Dependence*, pp. 167–90. Munich: Philosophia Verlag.

Hume, D. 1757. Of the standard of taste. In *Essays Moral, Political, Literary*. References to the edition by Eugene F. Miller. Indianapolis, IN.: Liberty Fund, 1987.

Kragh, H. 1990. *Dirac: A scientific biography*. Cambridge: Cambridge University Press.

Kuhn, T. S. 1970. *The Structure of Scientific Revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Kuhn, T. S. 1977. Objectivity, value judgment, and theory choice. In *The Essential Tension*, pp. 320–39. Chicago, IL: University of Chicago Press.

Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*, pp. 91–195. Cambridge: Cambridge University Press.

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

McAllister, J. 1999. *Beauty and Revolution in Science.* Ithaca, NY: Cornell University Press.

Mellor, D. H. 1991. The warrant of induction. In *Matters of Metaphysics*. Cambridge: Cambridge University Press.

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

Rosen, G. 1994. Objectivity and modern idealism: What is the question? In M. Michael and J. O'Leary-Hawthorne (Eds.), *Philosophy in Mind*, pp. 277–319. Dordrecht: Kluwer Academic Publishers.

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

Taub, L. 2008. *Aetna and the Moon.* Corvallis, OR: Oregon State University Press.

Timmermann, A. 2013. Scientific and encyclopaedic verse. In *A Companion to Fifteenth-Century English Poetry*. Cambridge: D. S. Brewer.

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

Walker, D. 2012. A Kuhnian defence of inference to the best explanation. *Studies in History and Philosophy of Science* **43**: 64–73.

Wedgwood, R. 1997. The essence of response-dependence. *European Review of Philosophy* **3**: 31–54.

Weyl, H. 1952. *Symmetry.* Princeton, NJ: Princeton University Press.