HOW CAN LOVELINESS BE A GUIDE TO TRUTH? INFERENCE TO THE BEXT EXPLANATION AND EXEMPLARS

from: *The Aesthetics of Science: Beauty, Imagination and Understanding*
edited by Milena Ivanova and Steven French, NewYork: Routledge 2020.

URL of publisher's version:
https://www.taylorfrancis.com/chapters/edit/10.4324/9780429030284-7/loveliness-guide-truth-inference-best-explanation-exemplars-alexander-bird

Abstract

According to Peter Lipton's account of Inference to the Best Explanation, a broadly aesthetic property of explanations, their loveliness, is what make explanations good and therefore likely to be true. How can loveliness be a guide to truth? Lipton raises three problems for the claim that loveliness is a guide to truth, but his answers to those problems are not fully convincing. I propose an answer that provides a better response. It is exposure to and training with exemplars of good science that inculcates in scientists a sense of what a good explanation looks like. Explanations are lovely to the extent that they are like these exemplars. This idea has its roots in the work of Kuhn and is confirmed by the work of cognitive psychologists. If the exemplars are chosen for their success, then a realist can claim that loveliness is indeed likely to be correlated with the truth.

1 Introduction—Inference to the Best Explanation and its problems

Science makes frequent use of Inference to the Best Explanation (IBE). While IBE is not the only form of inference to be found in science, it is key to our assessments of the plausibility of many theoretical scientific claims. The best discussion of what IBE is and whether it can give us rational beliefs and epistemic preferences concerning theories remains Peter Lipton's (2004) *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.

What I have called 'explanatory goodness', Lipton calls 'loveliness'. I shall take 'loveliness' to refer to the good-making features of an explanation, whatever they may be. On that construal it is almost trivial that when people make an inference to the best explanation, they are inferring the loveliest explanation. But it is far from trivial to know what loveliness is. The following are substantial questions: What are the lovely making features of an explanatory theory? And how do scientists come to see and respond to these features in a theory? Lipton himself thinks that loveliness is a matter of the understanding that a potential explanation would give, if it were true. Others have different views. According to the view I articulate below, loveliness is context-dependent (which is one reason why we find it difficult to agree on a single, fixed account of what it is).

Both stages in IBE raise important philosophical questions. A crucial problem concerns the first stage.

> **Underconisderation** The actual explanation may not be among the potential explanations we consider and rank. If so, the top ranked explanation will not be true. The truth was not even considered.

IBE cannot generate knowledge if we do not consider and assess all prima facie plausible explanations. Since stage 1 filters out so many logically possible explanations, what confidence can we have that the actual explanation is allowed through? The stage 2 ranking is of no use if the actual explanation hasn't made it through stage 1 on account of the scientists' failure to think of it. Why should we think that we have considered all the relevant possibilities? Kyle Stanford (2006) argues that we do in fact have reason, as demonstrated by the history of science, for thinking that our reasoning about novel theoretical claims suffers from what he calls 'the problem of unconceived alternatives'. We can articulate the problem as a problem about the truth-directedness of scientists' imaginations. The very idea of IBE concedes that the scientific imagination is directed unfailingly on the truth. For multiple hypotheses, products of the imagination, are considered, of which at most one is true. So, mostly, the scientific imagination produces false hypotheses. Given that the scientific imagination does not have a strong link with the truth, we ought not expect it to have the magical ability always or even very frequently to generate the true hypothesis alongside the false ones it generates.

Assuming that the actual explanation *is* among those investigated at stage 2, two problems immediately raise their heads, which Lipton calls 'Hungerford's objection' and 'Voltaire's objection'.

> **Hungerford's objection** Loveliness, being an aesthetic quality or akin to one, is subjective. If loveliness is subjective, then it cannot correlate with the truth, which is objective. Hence the loveliness of a potential explanation cannot be an indicator its truth.

Lipton here borrows Margaret Hungerford's line in *Molly Bawn*, that 'beauty is in the eye of the beholder'. As Kragh (1990: 287) puts it, 'The main problem is that beauty is essentially subjective and hence cannot serve as a commonly defined tool for guiding or evaluating science.' It is unclear whether explanatory loveliness is itself an aesthetic quality. But even if it isn't, it is prima facie very plausible that it is akin to an aesthetic property at least as regards subjectivity, for four reasons. First, we tend to use *aesthetic terminology* to describe the good-making features of explanations and theories, whether Lipton's 'loveliness', or McAllister's 'beauty', or Glynn's 'elegance'. The second reason for thinking that loveliness is an aesthetic quality or is at leat a subjective quality is that it is *ineffable*. It is difficult to say exactly what makes an explanation lovely. We are better able to recognize it than we are able to articulate what it is. Aesthetic qualities are ineffable in this way—many people will know that Mozart's oboe quartet is beautiful but cannot say in virtue of what properties it is beautiful. The ineffability of explanatory loveliness may in part explain our use of aesthetic terminology. The third reason is also behind the use of aesthetic terminology to describe theories. We take a certain kind of *pleasure* on grasping a particularly revealing, unifying, apt, or economical explanation. This pleasure is the same as or is very similar to the kind of pleasure one has on experiencing (with understanding) an artistically valuable poem, sonata, or still life. Finally, we find *variability* across different disciplines or across different times in the same discipline, regarding what counts as lovely. Even when we can agree that something such as simplicity is an aspect of loveliness, simplicity looks to be a different thing in physics and in explanations in physiology. Even within one field, astronomy, what counted as simplicity in one era (circular motion until the time of Kepler) was later no longer a determinant of loveliness. This variation in assessments of loveliness is what one would expect if loveliness is or is like an aesthetic quality, subject to changes in taste.

> **Voltaire's objection** If the loveliest explanation is to be true, then the actual world has to be lovelier than other possible worlds. But we should not expect the actual world to be the loveliest.

This objection says that the IBE enthusiast has an unjustified Panglossian faith that the actual world is the loveliest of all possible worlds. Even if loveliness is objective, there will be many worlds where it does not correlate with truth. So why think that loveliness and truth correlate in ours? Lipton regards Voltaire's objection as a version of Hume's problem of induction: there are worlds where apparent regularities are not true regularities, so why think that the apparent regularities are true regularities on our world? Voltaire's problem may be even worse. For induction to be a useful guide to the truth, it need only be that the actual world has a high enough degree of regularity. It does not have to be the most regular of worlds. Voltaire's problem is more demanding. For IBE to lead frequently to the truth, it has to be that the actual world is the loveliest of possible worlds, or at least among the most lovely worlds. And that just seems implausible.

In other work (Bird 2020) I argue that Lipton's own solutions to these problems are only partially satisfactory and that the sceptical worries the three objections raise remain. In this paper I provide my own response to those objections. That response appeals to the idea that a community's standards of explanatory goodness are acquired in the process of scientific training and learning which uses *exemplars*, drawing on a proposal by David Walker (2009, 2012).

2 Exemplars

In this section I articulate what is meant by an 'exemplar' and explain why thinking with exemplars is an important part of cognition in science. While the exemplar idea originates with Kuhn, its importance extends well beyond its place in Kuhn's (earlier) philosophy of science.

2.1 Exemplars in Kuhn's thought

In the 'Postscript 1969' to the second edition of *The Structure of Scientific Revolutions* Thomas Kuhn explicates in detail what he means by the contentious notion of 'paradigm'. There is a sense of 'paradigm' concerning which Kuhn (1970: 10) remarked, 'By choosing it, I mean to suggest that some accepted examples of actual scientific practice—examples which include law, theory, application, and instrumentation together—provide models from which spring particular coherent traditions of scientific research.' To express this meaning of 'paradigm' Kuhn also used the term 'exemplar', thereby distinguishing paradigms-as-exemplars from paradigms in a broader, more sociological sense. Exemplars are, Kuhn (1970: 187) said, the most novel and least understood aspect of his book. Exemplars are social in that they are shared exemplary illustrations and puzzle-solutions, and it is that by being shared that they can explain the development of a social activity such as science. But the key innovative feature of the exemplar idea is not so much the social as the psychological function of exemplary puzzle-solutions.

Kuhn's central claims about the functioning of exemplars are as follows:

> (i) the processes of scientific cognition are driven by perceived similarity to exemplary puzzles and their solutions (exemplars);
> (ii) the ability to perceive such similarities is acquired by training with exemplars;
> (iii) that ability is primarily an ability to recognise patterns and relevant similarities;
> (iv) it is therefore not an ability that is mediated by following rules.

Focussing on (i), the principal processes of scientific cognition in question are:

> (a) selecting puzzles;
> (b) solving puzzles;
> (c) assessing the quality of a proposed puzzle-solution.

It had been widely thought that in order for science to be rational scientists must follow certain rules of rationality, at least in the so-called context of justification. Optimally philosophers of science should be able to reconstruct the apriori rules of scientific method by which our theories are justified and hence accepted. Kuhn himself rejected the sharp distinction between the context of discovery and the context of justification since exemplars play a role in both (see (i) above). Moreover, justification and the explanation of scientific change do not depend on scientists following, even unconsciously, the rules of scientific method. Rather, what drives theory-choice is perceived similarity to an exemplar. This is most clear during normal science but is true also even in revolutionary science.

Kuhn (1970: 36) compared normal science to solving puzzles such as crossword puzzles. One feature of puzzle-solving to which Kuhn sought to draw our attention is the fact that one can learn to do crossword puzzles quicker and more easily simply by practising them. Although one can write very effective computer programmes that play chess by implementing an algorithm, good human players recognise patterns in positions in a non-algorithmic way that is acquired by past exposure to many similar positions. Kuhn (1970: 47) also likens practice with scientific puzzles to finger exercises on a musical instrument, emphasizing that the use of exemplars to solve problems is rather more like a skill than like the application of rules.

The most obvious cases of learning by repeated exposure and practice in science will be found in the mathematical sciences where students learn from a textbook and lectures and are given problems to solve. The easier problems will be very similar to the problems that are solved in the textbook and lectures. Harder problems will be less obviously similar. Practice with the former will give a student an improved ability that will in due course allow them to tackle the more difficult problems. As suggested, what makes a difficult problem more difficult will be not just increased complexity but also the fact that it is less clearly similar to a problem the student has seen before. What, according to Kuhn, repeated practice with exemplary puzzle-solutions provides is an ability to spot a similarity between a new puzzle and one that the student has seen before.

We are familiar with the idea of a non-rational power of recognising similarity in the perceptual sphere, for example, recognising the similarity between faces of members of the same family, or hearing that two tunes are similar. It is also true that we can recognise similarities between more abstract patterns and structures and this similarity recognition is likewise a non-rational one.[1] Of course, to say that this capacity is non-rational is not to say that it is irrational. It is clearly not irrational to be able to spot the similarity between a mother and her daughter. Rather what is meant is that this capacity is not mediated by a process of following rules.

Kuhn's purposes in drawing our attention to the existence and function of exemplars are (i) to undermine the then prevailing logical empiricist view that theory choice is governed by apriori rules of rationality (a view exemplified by Carnap's inductive logic or Popper's falsificationism); (ii) to explain certain 'revolutionary' episodes of theory-change—these involve changes in the governing exemplars; and (iii) to explain an interesting aspect of such changes, namely incommensurability. In fact Kuhn himself, from the mid-1970s seemed to lose interest in the exemplar idea and instead focussed on more philosophical accounts of incommensurability. At the same time, anti-realist ideas, absent from the first (1962) edition of *The Structure of Scientific Revolutions*, came more to the fore.

2.2 Exemplars beyond Kuhn

The value of the exemplar idea can be divorced from Kuhn's later (albeit moderate) anti-realism. Indeed, it is entirely consonant with an approach to scientific cognition that is naturalistic and realist in that it holds that we are capable of latching onto the truth and that we have this capacity in virtue not of grasping apriori rules of rationality but in virtue of possessing certain reliable modes of thinking that themselves may be innate or may be the products of experience and learning. Such ca-

[1] See Kuhn (1974) for an account of learning the categories of objects by learning to see similarities and differences, which he thinks is an instance of the same kind of thing as learning to use exemplars in research. Note that Kuhn (1974: 310) remarks that this process can be modelled on a computer; *cf.* related remarks in *Structure* (1970: 192).

pacities are therefore themselves open to scientific investigation and evaluation.[2] Facial recognition, as just mentioned, is a simple example of such a capacity, largely an innate capacity. The ability to assess the strength and potential of a chess position or the ability to recognize the composer of a previously unheard piece of music are learned capacities of this sort. Kuhn regards the ability to see the similarity between a new scientific problem and a previoiusly solved one as another learned recognitional capacity. Kuhn's idea has been taken up in this spirit by a number of philosophers (e.g. Andersen et al. 1996, Nickles 2003), in particular those who link the idea of exemplar to a wider capacity for pattern recognition and analogical thinking (e.g. Margolis 1987, Bird 2005) At the same time, there is evidence from more recent cognitive psychologists (Leake 1998, Dunbar 1996, 1999, Gentner et al. 2001) that confirm the view that scientists as well as others do reason by drawing analogies between the problems they face and previously solved problems (i.e. exemplars).

According to this research, scientists look for solutions to their scientific problems by looking for analogies between their problems and ones that have been solved before. Indeed a scientist will see the world through the lenses of her exemplars, in the sense that she will classify and conceptualise the world without the mediation of an inference. For example, the treehopper *Cyphonia clavata* has an extraordinary growth that looks very much like a tree ant. An evolutionary biologist will *immediately* see this as an adaptation, without going through a thought process involving 'these are the reasons ..., *a*, *b*, *c*, ... why this body-structure is most plausibly an adaptation to deter predation'. A pre-Darwinian naturalist would not have conceptualised what he saw as an adaptation, but might have, again without the mediation of inference, seen it as an instance of Divine Providence. A consequence of the fact that cognition using exemplars can be a wholly or partly unconscious process, like facial recognition, is that the subject can fail to be aware of all the factors influencing the cognitive process. It may simply not be possible to give a complete reconstruction of the process in terms of a rational inference. So such cognitive capacities are like intuition in that they are partly or wholly unmediated by conscious reasoning. On the other hand they are unlike intuition in that they are learned. They are 'second-nature', or, more prosaically, 'quasi-intuitive cognitive capacities' as I (2007) call them. And they are plausibly what Duhem (1914) had in mind when he talked of the 'good sense' that scientists use to prefer one theory to another.[3]

[2] It is notable that Kuhn (1970) cites many experimental psychologists, such as Bruner, Postman, Hastorf, Stratton, Whorf, and the Gestalt psychologists, i.e. Wertheimer, Koffka, and Köhler, in supporting his claims.

[3] Duhem's description of good sense ('le bon sens') matches this account of quasi-intuitive capacities in several respects. Good sense give a scientist those 'motivations which do not proceed from logic and yet direct our choices'. 'Good sense exists in every scientist to some degree ... and it can be cultivated and sharpened by training and practice', says Ivanova (2010: 60). There are, however, important differences. Duhem does think that good sense can be acquired from exemplars. But the exemplars, according to Duhem, are not instances of great science, but are great scientists (i.e. French scientists plus Newton and Huygens). According to Ivanova good sense is the collection of intellectual virtues possessed by an ideal scientist. This good sense does not vary from field to field or from era to era. Whereas what scientists learn from Kuhnian exemplars is localized to the field from which the exemplars are taken, and will change over time, e.g. as a result of revolutionary change. A final difference that is relevant to my argument (and not one that Kuhn would have endorsed) is that quasi-cognitive intuitive capacities need not be virtuous, in that they need not lead to better theories. Learning with bad exemplars can make one disposed to prefer flawed theories.

Cognizing with exemplars is consistent with realism. An apt analogy with an exemplar that is the correct solution to its own problem can be conducive to finding the correct solution to a new problem. But this consistency with realism does not exclude the possibility of systematic error. If the analogy is inapt or the exemplar is itself an error, then thinking with exemplars will lead to false beliefs (I give some examples below). So this theory of scientific cognition may permit realism but it certainly does not guarantee it. However, in the context of defending IBE from the three sceptical problems—Underconsideration, Hungerford's problem, and Voltaire's problem—that, I will argue, is enough.

2.3 Exemplars and loveliness

We are now in a position to see how the concept of an exemplar can provide a solution to the three problems facing IBE. The key idea in what follows is that scientists acquire their standards of explanatory goodness thanks to their exposure to exemplars. Although I introduced the idea by referring to Kuhn, the basic idea goes back to Aristotle, that we acquire our values through certain kinds of training, rather than by the learning of rules. We acquire virtuous dispositions and learn how to recognise the virtues in others by moral training, by witnessing the virtuous actions of others and imitating such actions, repeatedly acting as the virtuous person does. In science, training with exemplars is similar: the young scientist engages in academic exercises of problem solving with examples—from textbook questions to apprenticeship as a doctoral student. The young scientist thereby acquires the capacity to see new problem situations as similar to exemplary ones, to deploy theoretical and experimental tools in an analogous fashion, and to judge others' attempts to do the same. In particular she acquires the capacity to see theoretical problem situations as calling for a certain kind of explanation and correspondingly learns to judge how well a proposed explanation meets the demands of some problem situation. In a related vein Walker (2012) provides a paradigm-focussed understanding of Lipton's notion of explanatory loveliness.[4]

The discovery of Coulomb's law provides an illustration of the effect of exemplars on scientific cognition and on judgments of explanatory goodness in particular. In the context of Newtonian mechanics and the success of Newton's law of gravitation it was natural for a physicist confronted with an unstudied source of force to want to find its law also. So the exemplar of gravitation generates the problem for electrostatics. Furthermore it directs us towards a solution: by analogy with Newton's law, an inverse square law naturally suggests itself; indeed Coulomb was not the first to propose an inverse square law. Franz Aepinus had proposed it without any experimental evidence and Charles Stanhope attempted a demonstration that was widely accepted despite being flawed. Coulomb employed a torsion balance in his experiment, much the same, though differing in scale, as the torsion balance devised by John Michell and which Henry Cavendish used 'to weigh the Earth'. Coulomb's experiment was not universally considered to have proven his case: other scientists had difficulty in replicating his results and some proposed different laws on the basis of results using the same apparatus; modern scholars differ on whether Coulomb re-

[4]Although I do not agree with all the details of Walker's account of loveliness. For example, he regards the 'dormitive virtue' explanation of opium's capacity to induce drowsiness as unlovely (2012: 64). In the context of IBE that would mean that this explanation should be rejected as be unlikely to be true. But it is true, albeit only minimally informative. So although it may in one sense be fail to be a 'lovely' explanation, that is not the sense of 'lovely' relevant to IBE.

ally could have achieved the results he claimed he did (Heering 1992, 1994; Falconer 2004; Martínez 2006). Nonetheless, Coulomb's proposed law was accepted by the majority, despite its empirical shortcomings. It is apparent that as far as Coulomb and many of his contemporaries were concerned, the very fact of the analogy with the law of gravitation played a powerful role in influencing their judgment in accepting the law. That role cannot be thought of as an evidential role: it is difficult to see how the fact that Newton's law is an inverse square law is *evidence for* Coulomb's law, particularly in the context of IBE, since the form of Newton's law is not explained by Coulomb's law having that form (nor vice versa). Nor would it be true to the psychology of the scientists; the evidence they appealed to was what was known of electrical phenomena, in particular the various experimental results. Rather, the common form of Coulomb's proposed law and Newton's well-attested older law made the former a better explanation of the evidence.

In the following sections I argue that this exemplar-based account of our judgments of explanatory goodness resolves the three objections to IBE. (As I have emphasized, to respond satisfactorily to these objections is to remove prima facie reasons to reject IBE and so scientific realism; they are not intended to be arguments for the conclusion that one must rationally accept IBE and realism.) The key features of judgments of explanatory goodness, being based on learning with exemplars, are:

- The judgments are not entirely a matter of rational, conscious reflection. Since they are instances of recognizing patterns and similarities, there will be an unconscious element to the cognitive process.

- Consequently, the subject will be able to make a judgment without being able to articulate fully the basis on which the judgment is made—they may be able to say little more than 'it feels right' or 'that is how it strikes me'.

- The factors that do influence judgments will be dependent on the nature of the exemplars used in learning. Hence standards of explanatory goodness will not be apriori and may be specific to the discipline or sub-discipline from which the exemplars come, and to their stage of development.

3 Exemplars and Hungerford's objection

Hungerford's objection says that our evaluative standards are too subjective to correlate with truth. Loveliness looks very much like an aesthetic quality, and therefore subjective.

We should note that 'subjective' can be ambiguous. A quality can be subjective in the sense that it is response dependent, where the response is a private state. And it can be subjective in the sense of its holding is merely a matter of opinion. These will often go together. 'Chacun à son goût'—we might think that what food we find delectable and which unpleasant is usually just matter of taste, and there is no objective deliciousness or disagreeableness that these could correspond to. On the other hand, some qualities are subjective in the first sense but not the second—colour experiences are, arguably, like that. It is common to think that aesthetic qualities are of the type that are subjective in both senses: all there is are our inner aesthetic experiences; there is no fact about their being correct, aesthetic taste being a matter of personal opinion only. The line from Hungerford's *Molly Bawn*, 'beauty is in the eye of the beholder', expresses that view. But Hume and Kant both thought otherwise, holding that there are objective standards of taste to which the subjective aesthetic response of a suitable critic will correspond. Let's call qualities that are subjective in both senses 'merely subjective'.

Various features of explanatory loveliness do suggest that it is a subjective quality and indeed a merely subjective quality like (or among) the aesthetic qualities, at least as Hungerford's line takes them to be. In brief, these features were: i. Aesthetic terminology: writers use subjective qualities to describe explanatory goodness; ii. Ineffability: there is little or no agreement on what explanatory goodness is; Pleasure: we take pleasure in certain explanations, a pleasure that is at least very similar to aesthetic pleasure; iv. Variability: even when we focus on particular qualities that may contribute to explanatory goodness we find that there is variation in what counts as exemplifying these qualities; there is also no rule for weighting the qualities that contribute to explanatory goodness.

The correct response to this objection is to point out that these facts, while prima facie suggestive of mere subjectivity, do not entail it. Indeed they are also what one should expect if the process of learning to produce and evaluate explanations proceeds via training with exemplars rather than through the acquisition of explicit rules.[5] For example, one may learn what prudence is by learning to recognize examples of prudent actions and by taking such actions oneself. Such practical knowledge need not be manifested in an ability to say precisely what features of a given action make it prudent and how they differ from the properties of actions that are judged to be cowardly or avaricious. Prudence may nonetheless be an entirely objective feature of people and their actions. Many cognitive capacities are aimed at determining the presence of some objective property, yet the possessors of such skills are unable to articulate in any detail how they exercise such powers.

For example, a music lover who is not a trained musicologist may be able to identify the composer of a piece of music she has not heard before. She may nonetheless be able to give little concrete justification or explanation of her judgment beyond 'it just sounds like Mendelssohn'. Here we have an example of a cognitive skill detecting an objective property. Other examples include mastery of grammar: most English speakers will regard 'the big red barn' as grammatically preferable to 'the red big barn' without being able to articulate any rule of grammar that they are following in making this judgment—the first just *feels* right, and the second does not (Kihlstrom 1987: 1447). Experts at speed chess will be able to evaluate moves and the resulting chess positions as weak or strong with little or no conscious assessment. Many of these cognitive skills, like the more familiar capacity to recognize a face, are pattern-recognition capacities. In all these cases the detection of the relevant objective property or properties proceeds via a process that is partly or wholly unconscious. Consequently, one would not expect possessors of such capacities to be able to articulate what qualities they are responding to. So what is *not* occurring is the conscious detection of relevant qualities, and then a process of reasoning leading to an assessment. Since the subject cannot articulate some or all of the key determinants of their judgment, that process of judgment will be ineffable to the

[5] Walker's (2012: 68–9) response to this objection is to argue that subjectivity is really relativity to a scientific community. Exemplars and explanatory goodness do indeed vary between such communities: exemplars and what counts are a good explanation differ between evolutionary biology and geology. But I do not think that this answers the problem—indeed Walker's solution looks like a communitarian version of Lipton's solution. Instead of focussing on the scientific community *per se*, we should focus on the psychology of individuals that working with exemplars induces. That psychology will indeed be shared by members of the same community, because they work with the same exemplars.

subject. At the conscious level all that may occur is the response to the problem: a feeling that this is right or good. That feeling will be a positive affect amounting in some cases to pleasure.

In addition, the assessment of goodness may be a matter of balancing a number of criteria. When assessing the comparative strength of two chess positions, a chess master will take into account not only the value of the pieces, but their effectiveness in combination, positional strength, as well as psychological and other factors (e.g. whether the position is of a kind that is well understood or is unfamiliar, whether it favours aggressive or positional play, etc.). So although it may be possible to say what factors go into making a position strong for a player, it will be impossible to say explicitly how those factors should be weighted and combined; that is consistent with strength being objective and the master being able to make an accurate assessment of relative strength of two positions. Kuhn identifies five values (accuracy, consistency, scope, simplicity, and fruitfulness) employed in the valuation of a piece of science. But this does not amount to an account of explanatory goodness since there may be no algorithm for weighting these values. While the goodness of an explanation may be an objective property in each case, what makes that explanation good may differ between cases, and may be a complex combination of characteristics. Consequently the attempt to articulate standards of goodness may be fruitless—the standards are ineffable. Yet training with exemplars may permit a scientist to recognize them nonetheless.

This also explains the 'terminology' feature of judgments of explanatory goodness. If the processes are ineffable and judgments are made according to what feels right, then they will have much the same phenomenal character are aesthetic judgments. In which case it is hardly surprising that terminology is borrowed from aesthetics to describe the subject's response.

A further reason why we might expect it to be difficult to articulate standards of goodness even though those standards are measuring objective features of explanations is that there are different standards for different sciences. Different branches of science are governed by different paradigms: the exemplars of theoretical physics differ from those of inorganic chemistry and those of cell biology. Therefore, on the view being put forward here, there will be different standards of explanatory goodness set in these different disciplines and even within their sub-disciplines. There is not a single set of tacit criteria of explanatory goodness, but multiple sets, and so in that sense there is not a single inferential practice of IBE, but instead there are multiple inferential practices. What counts as 'simple' may not be the same for all scientific disciplines: in theoretical physics it might be the mathematical simplicity of the laws; in physiology it might be the simplicity of a causal process; in evolutionary biology it might be the simplicity of an evolutionary path, and so forth. Explanatory goodness and the general qualities (simplicity etc.) that contribute to goodness will therefore be multiply realizable. Hence one would not expect to be able to articulate a single informative account of what counts as explanatory goodness for all science, even though in each case what is being assessed is objective. Thus variability is what one would expect if standards of goodness are exemplar-driven, given that different branches of science employ different exemplars. Walker (2012: 67) provides a different dimension of variability—over time within a field. Using the example of astronomy he points out that the high explanatory goodness attached to circular orbits was changed to accommodate elliptical orbits. This might appear to a Hungerford objector as a change in taste. We ought rather to see it as a change in explanatory standards that is responsive to the greater success of Keplerian cosmology over Copernican. As a result of that success the former displaces the latter as our preferred exemplar and so as the source of our standards in cosmology.

So the exemplar-based explanation of our ability to evaluate explanatory goodness would predict that even though explanatory goodness is objective, our response to it will have features in common with an aesthetic response and so may appear subjective. That said, I suggest that the exemplar explanation is a better explanation of the data than the hypothesis that our judgments are subjective, because although there is variability between different fields as regards the assessment of goodness, there is rather less variability within fields. The stability of our evaluations of explanatory goodness suggests, though does not prove, that it is indeed an objective feature of explanations. If it were subjective, and genuinely like aesthetic judgments employed in literature or fine art, then one would expect greater variation and disagreement in our judgments. The fact that there is greater agreement in our judgments of explanatory goodness than in our attempts to articulate what it is, suggest that it is an objective property that we come to recognize through learning with exemplars.[6]

4 Exemplars and Voltaire's objection

in this section I argue that the exemplar approach to understanding explanatory goodness also provides a straightforward response to Voltaire's objection. I then compare this account to James McAllister's 'aesthetic induction'.

4.1 Responding to Voltaire's objection

Put bluntly, Voltaire's objection suggests that it would be a fluke if, given all the possible worlds there are, the actual world is the best by some given set of standards. It would indeed be a fluke if one were antecedently to lay down a set of standards, and then to ask whether the actual world meets them better than all the others. In effect, that is what happens with our standards of moral and physical good and evil, since those standards are established for reasons that one would not expect to correlate with the way the actual world is. Hence Dr Pangloss's conviction that the world is the best in moral and physical terms seems absurdly optimistic.

But the standards of explanatory goodness are not like that. Our exemplars are chosen for their success. Let us imagine that in a particular mature science, M, success correlates with truth. So in M there will be a number of exemplars which are true or have a high truth-content. These exemplars form the basis by which junior scientists entering M acquire their sense of what a good explanation is. Since standards of explanatory goodness are acquired from successful exemplars, and since in M success correlates with truth, it is perfectly plausible that these standards of explanatory goodness themselves are indicative of truth.

The key to this response is the idea that our standards of explanatory goodness are fixed by features of the actual world—by properties of theories that are actually successful. The exemplar model, put to work in another possible world, would lead to different standards of explanatory goodness, standards that would correlate with

[6] I would remark parenthetically that a possible response to Hungerford's objection is just to deny Hungerford's claim. Perhaps there is more objectivity in aesthetic judgment than she suggests, as Aristotle and Hume held, and has been argued by some recent psychologists (see Grammer et al. 2003; Dutton 2003; Falk and Balling 2010).

the truth in that world.[7] Hence the worry that it would be a fluke if the actual world is the best of all possible worlds explanation-wise can be set aside. The exemplar model tells us that the actual world is the best of all possible worlds (or close to it) *by actual-world standards*—and that is entirely to be expected.

This exemplar-based solution depends on two things:

(i) success does indeed correlate with the truth; and
(ii) in discerning explanatory features of exemplars we are discerning properties that play some role in the success of theories.

Do we know these things to be true?

Before responding to that question, it is important to remark that the exemplar-based approach adopted here does not provide any guarantee that exemplars will set standards that are truth-conducive. A period of normal science may well proceed on the basis of false theories that promote exemplars of explanatory goodness that are not truth-conducive. For example, much medieval science proceeded on the basis that a powerful analogy provides a good explanation. Alchemy employed four elements and four primary qualities. Analogously medicine employed four humours and related these to four temperaments and to personality types as well as the four ages of man. These were all also linked to the four seasons. One can extend the structure so that physiology is explained in terms of four principal organs. One can see how, given such models, parallel explanations in terms of a four-fold division of qualities or kinds, would be seen as explanatorily powerful. Analogies were often religious or quasi-religious. So trial by water was thought to be effective because the accepted best explanation for why the accused floats in water is that water, being an element, is pure and so would reject (and so cause to float) the body of a person who has an impure soul. More recently, the replication crisis in social psychology suggests that its exemplars are not truth-conducive. For example, one highly publicized study (Carney et al. 2010) maintained that adopting high-power poses causes increased feelings of power and tolerance to risk, alongside changes in biomarkers (raised testosterone and lowered cortisol). This was explicitly modelled on research by Strack et al. (1988) that showed that contraction of the 'smile muscle' (zygomaticus major) increases pleasure and enjoyment. While both studies found support for their hypotheses, both later failed replication tests. So modelling a new hypothesis on an earlier, successful hypothesis does not guarantee truth and may in fact do the opposite, encouraging falsity.

So in such cases the tacit criteria of explanatory goodness at work and acquired from exemplars of this sort will not lead to the truth. On the other hand, if the criteria are acquired from exemplars that are true or mostly so, then those criteria *can* be indicative of the truth. Arguably, in some branches of science, such highly truth-like exemplars came to be accepted and so truth-conducive criteria of explanatory goodness came to be employed as a result of the scientific revolution of the seventeenth century. Under such circumstances, further highly-truth-like hypotheses will come to be accepted and will add to the stock of exemplars that in turn reinforce (or gradually refine or develop) the tacit criteria of explanatory goodness.

If one were to show that IBE, as applied in some particular scientific field, really is truth-conducive, then one would indeed have to provide strong arguments for (i) and (ii), that is we would need to show that we are in the virtuous state of affairs

[7] Walker (2012: 72) puts it well regarding explanatory loveliness, 'There is no trans-world standard; loveliness is changeable enough to establish and improve a connection with truth in each possible world where IBE is applied.'

described at the end of the last paragraph. This is the approach taken by Walker (2012: 69). It is to enter into rather more general issues surrounding scientific realism. These are well-known and it is not necessary to pursue them here. In responding to Voltaire's problem (and the other two discussed), I am not aiming to convert the sceptic (or even neutral observer) to realism, but to avert a challenge that prima facie would be a reason for the realist to surrender her position. In the light of that goal, it suffices to make the following remarks. So long as (i) and (ii) are in fact true, then IBE will be reliable, and so, according to the epistemological reliabilist, will generate rational belief. In order to rebut scepticism about IBE that is enough (in the reliabilist's eyes). Propositions (i) and (ii) are at least plausible and not prima facie especially improbable. We have defused three objections that, if correct, would have shown that they are false or highly implausible. Lipton is right that we do not need to worry about such general issues of scepticism, and the epistemological responses to them—if on the one hand reliabilism is a satisfactory epistemology, then the current comments suffice; if it is not, then the problems are sufficiently widespread that worrying about the details of IBE is irrelevant. However, in articulating Voltaire's objection, I pointed out that it had an aspect that made it seem especially worrying: the reliabilist solution would be of little avail if the conditions for reliability are ones that seem *very unlikely* to be fulfilled. What the exemplar-based solution does is resolve this additional worry. Contrary to Voltaire's objection, the reliability of IBE does not require the world to be special after all.

4.2 McAllister's aesthetic induction

The response to Voltaire's objection that I advocate above has important similarities to what McAllister (1999: 77–81) calls the 'aesthetic induction'.[8] According to McAllister, scientists make aesthetic judgments of theories and together with the application of empirical criteria these play a role in which theories are accepted. This occurs because scientists perceive the presence of certain aesthetic qualities in empirically successful theories and perform an aesthetic induction, projecting the observed correlation of empirical and aesthetic properties. Thus the aesthetic properties can rationally play a role in theory choice, because they are indicators of the empirical properties of theories that scientists value in their pursuit of their scientific goals (e.g. empirical adequacy, truth).

Both this aesthetic induction and the currently proposed idea of exemplar-based judgments of explanatory goodness share the idea that epistemic assessment involves an element that is not purely empirical and which operates by placing an epistemic value on certain qualities of theories that have been found in earlier theories that are regarded as successful. There are significant differences that need to be borne in mind:

> The aesthetic induction is straightforwardly an enumerative induction whereas the exemplar-based approach articulates an element of inference to the best explanation.
>
> The aesthetic induction provides a reason to believe a theory that is independent of and parallel to the empirical evidence for the theory, whereas in the exemplar case these are not distinct reasons. An exemplar-based judgment of explanatory goodness contributes to the assessment of empirical evidence.

[8] And I owe a debt to McAllister's (1999) book in developing my thinking on this topic.

That assessment concerns whether a theory is a good explanation *of the data*, and so tells us whether the data is good evidence for the theory.

The basis of the aesthetic induction is the empirical qualities of prior theories. That the latter are conducive to the goals of science can (allegedly) be known by apriori analysis of those goals. In contrast, the basis of the exemplar-based judgments is just that those theories are held by the community to be exemplars. As mentioned above, those exemplars may be poor guides to achieving the goals of science.

Exemplar-based judgments of explanatory goodness involve largely tacit cognition of the relevant properties whereas McAllister presents the aesthetic induction as if it were an explicit inductive inference. That said, I think McAllister's account would be equally good if the aesthetic induction were tacit.

The aesthetic induction concerns explicitly aesthetic properties whereas exemplar-based judgments of explanatory goodness encompass a wider range of properties. In responding to Hungerford's objection I argued that while aesthetic terminology might often be used, this need not indicate that scientists and others are detecting purely aesthetic properties—if the properties in question were general aesthetic ones such as 'beauty', then we might expect more disagreement than we in fact find. That said, there might be richer, more finely-grained aesthetic qualities on which there is less expert disagreement. [9] And it should be noted that McAllister's understanding of an 'aesthetic' quality is liberal, and includes fit with metaphysical commitments, which will also be part of exemplar-based judgments.

These differences tell in favour of the exemplar-based approach. The aesthetic induction assumes that we can clearly distinguish empirical from non-empirical (e.g. aesthetic) criteria and that we can assess a theory's merits with regard to each separately. While in some cases there may be aesthetic qualities of a theory that can be assessed independently of evidence, more generally it is not possible to make such a distinction, at least if explanationist considerations are central to our practices of theory confirmation.[10] For according to explanationist views (such as IBE) how well evidence e supports theory T depends on how good an explanation of e is provided by T. So we cannot assess a theory on empirical grounds (how well the evidence supports it) independently of assessing it on supposedly non-empirical grounds (its explanatory goodness).

5 Exemplars and the problem of Underconsideration

One of the functions of an exemplar is to help us solve puzzles by generating solutions. So when a puzzle requires an answer in the form of an explanation of certain evidence, exemplars will help us generate potential explanations. The problem of Underconsideration questions whether we are likely to generate the true explanation among the potential explanations we consider. Exemplars solve this problem as follows. We are faced by a new puzzle, P, with data that requires explanation. Let

[9] To give a musical parallel, critics may disagree as to whether a piece is 'beautiful' or 'great' but may agree that it creates a dramatic tension by means of a sustained crescendo over a repeated rhythmic pattern.

[10] McAllister does list explanatory power among the empirical criteria of theory assessment.

us assume (a) that, frequently, relevantly similar puzzles have similar (true) solutions; and (b) the extant, accepted exemplars in this field are themselves frequently true explanations of their data. Our exemplar-based approach tells us that we seek solutions to P by looking first for past scientific problems to which we have accepted solutions. We then model our potential solutions to P on these accepted solutions—this process might generate more than one potential solution. Assumption (b) tells us that these accepted solutions are often true. Assumption (a) tells us that the true solution to P is likely to be similar to one of these accepted solutions, which is to say that the true solution is likely to among those we have considered by this process.

Assumptions (a) and (b) are not unreasonable, and so there is no reason to suppose that Underconsideration is an inevitable problem. Lipton (2004: 162) himself notes "Theory generation is highly constrained by background, and insofar as the background approximates the truth, we should not be so surprised that our powers of generating the truth are substantially better than guesswork." Lipton's reference to 'background' is in the context of background *auxiliary* hypotheses, and it remains unclear how true background auxiliary hypotheses could point our imaginations in the direction of true hypotheses. But if we think of 'background' as referring to our (true) exemplars, then it becomes immediately clear how such constraints are truth-conducive. (Lipton's own answer to Underconsideration is that since ranking at stage 2 is dependent on background theories (auxiliary hypotheses), reliable ranking implies that often background theories are often true. So Underconsideration cannot be a problem that generally afflicts our theories. While this argument is valid, the sceptical proponent of the problem of Underconsideration can doubt it its premise by happily denying that our stage 2 ranking is known to be reliable.)

I am arguing that if we are in a position that our models are true hypotheses, then modelling our new hypotheses on those will direct our attention to potential explanations that are likely to contain the true explanation among them. Our search in the space of hypotheses is not a random walk. For example, in 1981 researchers found a clustering of highly unusual symptoms, a rare form of pneumonia, *Pneumocystis carinii* (now known as *Pneumocystis jiroveci*), and a rare form of cancer, Kaposi's sarcoma, among homosexual men in the United States. This syndrome, eventually called AIDS, elicited four hypotheses from researchers:

(A1) Recreational drugs.
(A2) Overload of the immune system from familiar sexually transmitted diseases and other infection.
(A3) Bacterial infection—infection by a bacterium, probably hitherto unknown.
(A4) Viral infection—infection by a virus, probably hitherto unknown.

These explanations were suggested by researchers' experiences with disease and illness. For example, immunodeficiency was already known to be a consequence of familiar infections, such as chicken pox. Perhaps severe immunodeficiency might be caused by multiple infections, such as young gay men experienced. A natural pair of hypotheses for any new syndrome are viral and bacterial infection, since so many other diseases care caused by viruses and bacteria. And of course in this case the viral hypothesis was shown to be correct by by Luc Montagnier in 1983. And so, if by good science or good fortune we have been able to generate a correct explanatory hypothesis in the past and its success leads to it becoming an exemplar that informs future searches for explanations, then there is a high chance that the true answers to those subsequent explanatory puzzles will be generated. In that way, if we ally the exemplar model of the cognitive processes of science with scientific realism, we can see that the former can generate a virtuous cycle of true answers to explanatory puzzles generating further true answers to similar explanatory puzzles.

6 Conclusion

A common conception of cognition says that our reasoning processes, if rational, should be explicable by reference to apriori rules. Rational deductive thought can be explained as conforming to a priori rules of deductive logic, as articulated initially by Peano, Frege, and Russell. Carnap and others hoped to extend this to inductive reasoning. Popper, holding that there could be no such logic of induction, maintained that rational inference in science must therefore be deductive (hence his falsificationism). Bayesians claim instead that good scientific reasoning can be understood as instantiating Bayes's rule, itself a theorem of the apriori axioms of probability.

Cognitive scientists and naturalistically inclined philosophers recognise that this conception of rational inference is too narrow. Margolis (1987), for example, argues that pattern recognition is central to human thinking and judgment, including in science. The case made for the role of exemplars in science supports this idea. The human capacity for pattern recognition is a cognitive capacity—it helps us make correct judgments about the world. Yet it cannot be construed as implementing a rule. There is nothing logical (nor illogical) in recognizing a face; nor in recognizing one scientific problem as like another.

Furthermore, since pattern recognition is a basic cognitive skill, not reducible to the implementation of a rule or the product of some other kind of reasoning, we are typically unable to articulate (from the first-person perspective) what it is in virtue of which we exercise this skill successfully. One often sees a pattern, and as far as the subject is concerned, that is that. In this respect, pattern recognition is phenomenologically like the appreciation of aesthetic qualities: cognitive and aesthetic responses can be elicited immediately, without conscious ratiocination, leading to a conviction that this is correct/beautiful, that can be difficult to justify verbally. (Although neither requires conscious reasoning, both can be enhanced by conscious thought.) These first person similarities explain why scientific cognition using exemplars is often described using aesthetic terminology.

Instances of great art can be held up as exemplars of aesthetic qualities (beautiful, sublime, tragic, romantic, graceful, and so on) and can thereby set the standard by which other works are judged to have these qualities. In the sciences likewise, great achievements—paradigms in Kuhn's terms—exemplify (among other things) our standards of good science. In aesthetics realism is highly contentious, whereas in the philosophy of science realism, although contended, is the majority view. The exemplars of science may be false or they may be true. When true, the standards exemplified by an exemplar will be truth-conducive. If they are false, then we may find false theories more attractive (arguably this was the case for much pre-modern science). It cannot be the *philosopher's* job to say that scientists are now in the former, 'good' scenario of learning our values from true exemplars. And in any case, some fields of science may be in a good scenario and others in a bad scenario. But the philosopher can argue, as I have tried to do, that the good scenario is at least *possible*: learning from exemplars *can* inculcate in us standards of explanatory loveliness that will give us a preference for explanatory theories that are more likely to be true.

References

Andersen, H., P. Barker, and X. Chen 1996. Kuhn's mature philosophy of science and cognitive psychology. *Philosophical Psychology* **9**: 347–63.

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

Bird, A. 2007. 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. 2020. Scientific realism and three problems for inference to the best explanation. In W. J. Gonzalez (Ed.), *New Approaches to Scientific Realism*. De Gruyter.

Carney, D. R., A. J. C. Cuddy, and A. J. Yap 2010. Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science* **21**: 1363–8.

Duhem, P. 1914. *La théorie physique, son objet et sa structure* (2nd enlarged ed. ed.). Paris: Chevalier et Rivière. (Reissued, Vrin, 1981) Translated P. Wiener *The Aim and Structure of Physical Theory* 1954 Princeton NJ: Princeton University Press.

Dunbar, K. 1996. How scientists really reason. In R. Sternberg and J. Davidson (Eds.), *The Nature of Insight*, pp. 365–95. Cambridge, MA: MIT Press.

Dunbar, K. 1999. How scientists build models: In vivo science as a window on the scientific mind. In L. Magnani, N. J. Nersessian, and P. Thagard (Eds.), *Model-Based Reasoning in Scientific Discovery*, pp. 85–99. New York, NY: Kluwer/Plenum.

Dutton, D. 2003. Aesthetics and evolutionary psychology. In J. Levinson (Ed.), *The Oxford Handbook of Aesthetics*, pp. 693–705. New York, NY: Oxford University Press.

Falconer, I. 2004. Charles Augustin Coulomb and the fundamental law of electrostatics. *Metrologia* **41**: S107–S114.

Falk, J. H. and J. D. Balling 2010. Evolutionary influence on human landscape preference. *Environment and Behavior* **42**: 479–93.

Gentner, D., K. Holyoak, and B. Kokinov 2001. *The Analogical Mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press.

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.

Heering, P. 1992. On Coulomb's inverse square law. *American Journal of Physics* **60**: 988–94.

Heering, P. 1994. The replication of the torsion balance experiment: The inverse square law and its refutation by early 19th-century German physicists. In C. Blondel and M. Dörries (Eds.), *Restaging Coulomb. Usages, Controverses et Réplications autour de la Balance de Torsion*, Volume IV of *Biblioteca di Nuncius*, pp. 47–66. Florence: Leo S. Olschki.

Ivanova, M. 2010. Pierre Duhem's good sense as a guide to theory choice good sense as a guide to theory choice. *Studies in History and Philosophy of Science* **41**: 58–64.

Kihlstrom, J. F. 1987. The cognitive unconscious. *Science* **237**: 1445–52.

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. 1974. Second thoughts on paradigms. In F. Suppe (Ed.), *The Structure of Scientific Theories*, pp. 459–82. Urbana: University of Illinois Press. Page references to reprint in Kuhn 1977.

Leake, D. 1998. Case-based reasoning. In W. Bechtel and G. Graham (Eds.), *A Companion to Cognitive Science*, pp. 465–76. Oxford: Blackwell.

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

Margolis, H. 1987. *Patterns, Thinking, and Cognition. A theory of judgment*. Chicago, IL: University of Chicago Press.

Martínez, A. A. 2006. Replication of Coulomb's torsion balance experiment. *Archive for History of Exact Sciences* **60**: 517–63.

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

Nickles, T. 2003. Normal science: From logic to case-based and model-based reasoning. In T. Nickles (Ed.), *Thomas Kuhn*, pp. 142–77. Cambridge: Cambridge University Press.

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

Strack, F., L. L. Martin, and S. Stepper 1988. Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology* **54**: 768–77.

Walker, D. 2009. *A Kuhnian Defence of Inference to the Best Explanation*. Ph. D. thesis, University of Bristol.

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