This is a pre-copy-editing, author-produced version of an article accepted for publication in British Journal for the Philosophy of Science *following peer review.*

Bayesianism and Inference to the Best Explanation

Leah Henderson

ABSTRACT

Two of the most influential theories about scientific inference are Inference to the Best Explanation (IBE) and Bayesianism. How are they related? Bas van Fraassen has claimed that IBE and Bayesianism are incompatible rival theories, since any probabilistic version of IBE would violate Bayesian conditionalisation. In response, several authors have defended the view that IBE is compatible with Bayesian updating. They claim that the explanatory considerations in IBE are taken into account by the Bayesian because the Bayesian either does or should make use of them in assigning probabilities (priors and/or likelihoods) to hypotheses.

I argue that van Fraassen has not succeeded in establishing that IBE and Bayesianism are incompatible, but that the existing compatibilist response is also not satisfactory. I suggest that a more promising approach to the problem is to investigate whether explanatory considerations are taken into account by a Bayesian who assigns priors and likelihoods on his or her own terms. In this case, IBE would emerge from the Bayesian account, rather than being used to constrain priors and likelihoods. I provide a detailed discussion of the case of how the Copernican and Ptolemaic theories explain retrograde motion, and suggest that one of the key explanatory considerations is the extent to which the explanation a theory provides depends on its core elements rather than on auxiliary hypotheses. I then suggest that this type of consideration is reflected in the Bayesian likelihood, given priors that a Bayesian might be inclined to adopt even without explicit guidance by IBE. The aim is to show that IBE and Bayesianism may be compatible, not because they can be amalgamated, but rather because they capture substantially similar epistemic considerations.

1. Introduction

Two of the most influential theories about scientific inference are 'Inference to the Best Explanation', the view that scientific inference is governed by the search for the most explanatory theories, and Bayesianism, the view that scientific inference involves finding theories which are the most probable in the light of the evidence, where the probabilities for the different theories are to be updated by the rule of 'conditionalisation'. At first sight, Inference to the Best Explanation (IBE) and Bayesianism appear quite different. How are they related?

Bas van Fraassen has argued that if IBE and Bayesianism are both taken to provide rules for updating probabilities on a given hypothesis space of theories, the rules are incompatible (Van Fraassen [1989]). According to van Fraassen, any probabilistic version of IBE would violate conditionalisation.

In response, several authors have attempted to defend the position that IBE is compatible with Bayesian updating. To date, the main proposal has been that the Bayesian takes into account explanatory considerations because the Bayesian either does or should make use of explanatory considerations in assigning probabilities (priors and/or likelihoods) to the hypotheses in the hypothesis space (Okasha [2000]; Lipton [2004]; Weisberg [2009]). I call this position 'constraint-based compatibilism', since it is based on the idea that explanatory considerations serve as constraints on the assignment of Bayesian probabilities.

In this paper, I argue that van Fraassen has not established that Bayesianism is incompatible with IBE (section 5). However I also argue that constraint-based compatibilism lacks independent motivation (section 6). I present a case in which a different kind of compatibilist account is possible. Given certain 'natural' priors, the Bayesian takes explanatory considerations into account, even though those explanatory considerations are not responsible for the assignment of the priors. I suggest that this points the way towards a more promising response to van Fraassen based on what I call 'emergent compatibilism' (section 7).

Throughout the paper, I will follow the practice in the literature of using the term 'explanationist' to refer both to an agent who performs IBE and to someone who characterises scientific inference in terms of IBE. A similar double usage will apply in the case of 'Bayesian'.

2. Preliminaries

In their traditional forms, IBE and Bayesianism characterise the assessment of hypotheses in terms of different epistemic attitudes. IBE tells us what hypothesis to select for belief (or perhaps acceptance), whereas Bayesianism operates in terms of degrees of belief. If one regards IBE as wedded to belief rather than degrees of belief, then on some views of the relationship between beliefs and degrees of belief, IBE and Bayesianism may be incompatible from the outset.[i] However, discussions of the relationship between IBE and Bayesianism have generally supposed, following van Fraassen's lead, that it is reasonable for an explanationist to have degrees of belief in different hypotheses. That is, the explanationist assigns greater degrees of belief to hypotheses that provide better explanations of the phenomena. This is the form of IBE which I will consider in this paper.

3. IBE

Better explanation is often cited by scientists themselves as a reason for giving a certain theory extra weight. For example, in *On the origin of species*, Charles Darwin argues for his theory of natural selection on the grounds that it explains 'large classes of facts' in a more satisfactory manner than its rivals, in particular special creationism (Darwin [1859]). A number of philosophers have also advocated that scientific inference is driven by the search for the best explanation. The significance of this claim largely depends on what it means for a theory to provide a 'better explanation' or the 'best explanation' of the phenomena (accounts are provided by Harman [1965]; Thagard [1978], Thagard [1992]; Niiniluoto [1999]; Lipton [2004]). It is often suggested that a theory provides the best explanation if it possesses the optimal combination of explanatory virtues in relation to the phenomena, where explanatory virtues include considerations such as simplicity, unification, scope, and fruitfulness. In discussions of the relationship between IBE and Bayesianism, analysis of IBE has typically not gone much further than this. However, deeper analysis is required to fully understand how IBE can be represented in Bayesian terms, and I undertake this task in section 7.1.

4. Bayesianism

Suppose that the hypothesis space under consideration is $\mathbf{H}$ and it contains a collection of candidate hypotheses T. A minimalistic version of Bayesianism, which I will call 'Min Bayes', has two tenets. First, the *coherence* postulate says that degrees of belief in the different hypotheses should be represented by a probability distribution (or density) over $\mathbf{H}$. Second, updating of probabilities is conducted according to the rule of *conditionalisation.* According to this rule, when evidence or data D is obtained, the probabilities are updated by replacing the ‘prior’ probability in hypothesis $T \in \mathbf{H}$, $p(T)$, with a ‘posterior’ probability which is given by the conditional probability, $p(T \mid D)$. The conditional probability $p(T \mid D)$ can be computed using Bayes’ rule:

$$p(T \mid D) = \frac{p(D \mid T)\, p(T)}{p(D)} \quad (1)$$

Here $p(D \mid T)$ is the ‘likelihood’ of T given D, and $p(D)$ is the initial probability of observing the evidence.[ii]

Bayesians divide along a spectrum between two extremes. At one extreme are ‘subjective’ Bayesians, who hold that Min Bayes is the complete story about rational constraints on inductive inference. According to this position, any assignment of probabilities which satisfies coherence and conditionalisation represents a reasonable state of degrees of belief. At the other extreme are ‘objective’ Bayesians who argue that there is just one rationally acceptable probability function, given the available background belief. Objective Bayesian positions have involved adopting normative constraints on the probabilities such as the Principle of Indifference, or the Maximum Entropy Principle. [iii]
In between these two extremes are Bayesian positions which adopt some constraints on the probabilities but without necessarily restricting to just one rationally acceptable probability function. For example, Abner Shimony has proposed a ‘tempering’ condition on prior probabilities (Shimony [1970]), and many Bayesians would also allow our knowledge of chance to constrain degrees of belief.

I now turn to a consideration of the different views of the relationship between IBE and Bayesianism which have been proposed, starting with van Fraassen’s incompatibilist view.

5. The incompatibilist view: IBE contradicts Bayesianism

According to van Fraassen, IBE and Bayesianism have incompatible updating rules. This is because, as he sees it, the explanatory considerations in IBE are something additional to the considerations of the Bayesian. Van Fraassen sees Bayesianism as taking into account ‘the initial probabilities and the series of outcomes’ when forming a posterior distribution. The explanationist, on the other hand, also takes into account ‘explanatory success’ (Van Fraassen [1989], p. 166) and adds extra probability to hypotheses which have this feature.

Van Fraassen casts his argument in the form of a dialogue between an ‘orthodox Bayesian’ whom he calls ‘Peter’, and an itinerant Preacher of IBE. The Preacher tries to convert Peter to infer to the best explanation. The discussion takes place surrounding the particular example of an 'alien die' (that is, a die which is discovered on another planet). There are a number of hypotheses $T_1 \ldots T_n$ concerning the bias of the die. The hypothesis T_i says that the chance of throwing an ace on any given toss is $\frac{i}{n}$. Peter assigns equal prior probabilities to the different bias hypotheses, and assigns likelihoods in accordance with his knowledge of chances. That is, he takes $p(ace \mid T_i) = \frac{i}{n}$. He throws the die and observes a series of aces. He then forms a posterior probability distribution by conditionalisation on this evidence. The Preacher comes and tells Peter that hypotheses of high bias are a better explanation of throwing a series of aces than hypotheses of lower bias. Therefore, the Preacher says, 'you should raise your credence in the more explanatory hypotheses. What? exclaims Peter. More than I would anyway? Yes says the Preacher' (Van Fraassen [1989], p. 166). Thus on van Fraassen's view, following IBE involves following a different rule than simple conditionalisation.

Van Fraassen goes on to use this as an argument that IBE is an unsuitable rule of inference, since it will fall foul of the dynamic Dutch-book arguments which justify Bayesian conditionalisation (Van Fraassen [1989], pp. 166-169). Implicitly then, he denies that there is any other way for explanatory considerations to play an epistemic role than by boosting the probabilities of explanatory hypotheses beyond what they would have received by conditionalisation. He thus presents a kind of dilemma. Either explanatory considerations play this boosting role and produce an updating rule that conflicts with conditionalisation, or they play no role at all in a Bayesian context.

5.1 Criticism of the incompatibilist view

The problem with van Fraassen's view is that the boosting interpretation appears unmotivated from the outset and it is not supported by van Fraassen's own example. Although it is logically possible for the probability distribution assigned by the Bayesian and by the explanationist to diverge in the way van Fraassen suggests, nothing about the example he gives inclines us to think they do. In fact, it seems rather easy to understand the alien die example without conflict between IBE and Bayesian conditionalisation. As more and more aces are thrown, the explanationist regards high-bias hypotheses as increasingly clearly the best explanation. And the Bayesian also assigns higher probability to the high-bias hypotheses. Why should we think that IBE involves assigning any more probability to the 'explanatory' hypotheses than conditionalisation would? Peter may not need to do anything differently in order to conform with IBE.

If we reject the boosting interpretation of IBE, are we forced, as van Fraassen implies, to regard IBE as having no epistemic significance? We might rather regard van Fraassen's model as a misinterpretation of IBE in the probabilistic context.[iv] But then the challenge is to find an interpretation of IBE in the probabilistic context which does not render it incompatible with conditionalisation, but which still allows explanatory considerations to play an epistemic role.

6. Constraint-based compatibilism

So far, compatibilists have proposed marrying Bayesianism and IBE in the following way. The explanatory considerations involved in IBE need not provide extra boosts to the posterior probability as van Fraassen suggests. Rather explanatory considerations play a role in guiding or determining the assignment of Bayesian priors and likelihoods. The basic idea is that explanatory considerations constrain how Bayesian probabilities are assigned, thus I call this position ‘constraint-based compatibilism’.

Constraint-based compatibilism appears to have several attractions. For proponents of IBE, it is attractive because it preserves a key role for explanatory considerations in inference, while allowing the use of the ‘powerful and well-studied formal framework’ that Bayesianism provides (Weisberg [2009], p. 136). The view is also supposed to be beneficial to Bayesianism, by filling in alleged gaps in the Bayesian framework.

There are different versions of the view depending on whether the role of explanatory considerations is seen as purely descriptive or also normative. Lipton argues for the descriptive claim that people actually do find candidate hypotheses and assign priors and likelihoods to them based on explanatory considerations (Lipton [2004]). Part of Lipton’s idea here is that it may be easier for us to follow explanationist heuristics than to perform abstract probabilistic calculations that ‘we are not very good at’ (Lipton [2004], p. 120). Thus, for Lipton, explanatory considerations provide essential assistance in running the Bayesian machinery. For Weisberg, on the other hand, the role of explanatory considerations is explicitly normative (Weisberg [2009]). His recommendation is that we *should* use explanatory considerations to constrain the priors, and thus that we should adopt a novel form of objective Bayesianism.

An important question for the constraint-based compatibilist is whether they are trying to marry IBE with a subjective or with a more objective form of Bayesianism. If it is with subjective Bayesianism, the explanatory considerations are to be primarily responsible for determining the priors and likelihoods. If, on the other hand, it is with a more objective form of Bayesianism, there are already some restrictions, whether as a result of formal principles or less formal reasoning, on what counts as reasonable initial probabilities. Explanatory considerations are to serve as further constraints which supplement those probability-constraining principles already accepted by the Bayesian.

Lipton tends to argue that the Bayesian would be very much at a loss regarding the assignment of probabilities, without explanationist help. Lipton argues that this is because we find dealing with ‘abstract probabilities’ difficult, and explanationist heuristics are psychologically easier to cope with. Thus IBE is to step in and provide the means for assigning probabilities that the Bayesian would otherwise lack. On this point of view, the marriage is between IBE and a subjective form of Bayesianism.

Weisberg, on the other hand, explicitly keeps his options open on this issue. He doesn't specify whether we should marry IBE with subjective or objective Bayesianism. He says that explanationist thinking should be used to 'fix the prior probabilities either in conjunction with, or in place of existing objectivist principles. Ideally, explanationist considerations would complement existing objectivist principles like the Principle of Indifference' (Weisberg [2009], p.137).

Constraint-based compatibilists have also made different suggestions about the exact role that explanatory considerations are to play with respect to the assignment of probabilities. Lipton proposes that 'explanatory considerations of unification, simplicity and their ilk' help to determine the priors (Lipton [2004], p. 115). Lipton also argues that since priors depend on the results of previous empirical investigations, to the extent that explanatory considerations are involved in these investigations they will also have a role to play in determining the priors on a given hypothesis space (Lipton [2004], p. 115). According to Lipton, explanatory considerations are used to determine likelihoods as well as priors. He says that 'one way we judge how likely D is on H is by considering how well H would explain D '. In particular, the idea is that considering the mechanism linking cause and effect in a potential causal explanation may help us to form a judgment of how likely the cause would make the effect (Lipton [2004], p. 114).

Weisberg gives three examples to illustrate ways in which explanatory considerations should constrain the priors. In one case, he proposes that higher prior probability is to be assigned to more systematizing theories, where systematizing is matter of trading simplicity with strength as in Lewis' account of laws (Weisberg [2009], p. 139).

In another example, Weisberg suggests that explanatory considerations constrain the prior probabilities indirectly via constraints on the conditional probabilities $p(T \mid D)$. The idea is that if T provides a better explanation of D than T', then we should assign $p(T \mid D) > p(T' \mid D)$ (Weisberg [2009], pp. 138-139). Then inequalities of this kind provide constraints on the priors, perhaps when combined with objectivist constraints on the likelihoods,

Weisberg's final example is one in which explanatory considerations provide assistance in application of the Principle of Indifference. There are traditional objections to the Principle of Indifference which stem from inconsistency on apparently equally valid choices of hypothesis space such as those related by reparametrisation. Weisberg's thought is that these objections might be to some extent overcome if explanatory considerations help to determine the appropriate hypothesis space to apply the Principle to. Specifically, we should choose the hypothesis space where the hypotheses count as 'potential explanations' of the relevant data. Michael Huemer (Huemer [2009a]) makes a similar proposal about how the explanationist may aid the objective Bayesian. [v]His view is that what counts as a potential explanation is based on a relation of 'explanatory priority'. Specifically the 'explanans must be in some sense *prior* to (or more basic than, or more fundamental than) the explanandum' (Huemer [2009a], p. 352). Explanatory priority, according to Huemer, may arise due to causal priority, temporal priority, a part-whole relation, an in-virtue-of relation, or supervenience (Huemer [2009a], pp. 352-353). Although part of the notion of explanatory priority is empirical, Huemer claims that 'at least *some* facts about explanatory priority can be known a priori' (Huemer [2009a], pp. 364).

Lipton also argues that explanatory considerations may play a role in the 'context of discovery' by helping to generate a 'short-list' or hypothesis space of plausible hypotheses which provide potential explanations of the data. According to Lipton, we determine what hypothesis provides a potential explanation according to our background beliefs. Since those background beliefs were themselves determined by IBE on previous evidence, explanatory considerations involved in prior empirical investigations help to determine the candidate hypotheses (Lipton [2004], pp. 116-117, 148-151). For example, suppose we have a background belief that people's emotions cause their political beliefs.[vi] Then we may generate a list of different possible emotional states as the hypotheses to consider in explaining political beliefs. However, the background belief that emotions cause political beliefs was itself gained empirically as the result of IBE on earlier data. Thus, explanatory considerations involved in IBE help to determine what kind of hypotheses may serve as potential explanations of new data. Lipton again argues that this is a way in which explanatory considerations may fill in the Bayesian account in an area where Bayesianism itself has little or nothing to say. [vii]

6.1 Criticism of constraint-based compatibilism

If the proposal of the constraint-based compatibilist is to marry IBE with subjective Bayesianism, then there are no further norms supplied by the Bayesian which might compete with the proposed explanationist constraints. On the more descriptive version due to Lipton, the proposal is that explanatory considerations actually do step in and determine the choice of priors and likelihoods. On a normative version, the recommendation is that explanationist constraints on priors and likelihoods should be adopted.

As we have seen, one of the proposals made by constraint-based compatibilists is that explanatory considerations such as simplicity, unification or systematizing power guide or determine the prior probabilities. Prior probabilities are supposed to represent the epistemic situation before taking the new data into account. Therefore the considerations which govern the choice of prior must be either a priori, or based on background beliefs from previous investigations. Thus, how well hypotheses explain the past data can help determine the prior probabilities. However, explanatory considerations like simplicity and unification also depend on the new data. We assess the best explanation with respect to all the data, past and present. When we say for example that a theory provides the simplest explanation, we mean that it explains all the data in the simplest way. Similarly when a hypothesis is unifying, it unifies not just past data, but also the new data. Therefore, it is not at all clear that people do, or even can, use the explanatory considerations like simplicity and unification which are involved in determining the best explanation of all the data to constrain the prior, as Lipton suggests.[viii]

Another possibility is Weisberg's idea that IBE imposes constraints on the conditional probabilities $p(T \mid D)$ and hence indirectly on the priors. This is in principle possible, but employing constraints of this kind would not be as straight-forward as Weisberg makes it seem. It could be for example that with respect to data D, T provides a better explanation than T', and hence we should constrain $p(T \mid D) > p(T' \mid D)$. Given certain other constraints on the likelihoods, this inequality implies some constraints on the priors. But with different data, the inequality may be reversed and thus a different constraint on the priors may result. To put constraints on the priors, one would need to hypothetically consider each possible way that the evidence could turn out, and then find prior probabilities which will agree with the explanationist assessment no matter which evidence they are conditionalised on. Although it is hard to be definitive about psychological feasibility without considering empirical evidence, it is not obvious that such a procedure possesses the psychological ease that motivated Lipton to turn to the explanationist for help.[ix]

One might still stipulate, as Weisberg does, that we *should* constrain priors in such a way as to agree with IBE. As Weisberg says, this means giving 'normative primacy' to IBE and effectively recommending a new form of objective Bayesianism. This is a possible response to van Fraassen's challenge, but without independent motivation, it appears more like a stipulation designed to ensure that Bayesianism is compatible with IBE, rather than an explanation of why it is. It has the consequence that IBE is merely accommodated in, rather than explicated by the Bayesian framework. And it also raises the question of why this new form of objective Bayesianism should be preferred over previous attempts to provide objective Bayesian norms.

The other suggestion that constraint-based compatibilists have made is that traditional forms of objective Bayesianism are not to be abandoned altogether, but rather supplemented with explanatory considerations. Thus, IBE should supply supplemental constraints to be fitted in alongside more traditional objective Bayesian constraints in areas where these prove insufficient. The main concrete suggestion here has been that explanatory considerations may step in to the alleged hole in the Bayesian framework in the context of discovery.

Lipton's proposal along these lines is not very compelling because he has not succeeded in identifying a genuine 'hole' in the Bayesian account which explanatory considerations might fill. It is reasonable to suggest, as Lipton does, that how well hypotheses explained past data can help to determine both the hypothesis space and the prior probabilities. However, Lipton is not right to present this as a case where IBE plays a role in generating the hypothesis space where Bayesianism is silent. In terms of the role of background beliefs and previous empirical investigations, the Bayesian is no worse off than the explanationist. The explanationist argues that we determine what counts as a potential explanation according to background beliefs themselves determined by IBE. Similarly, the Bayesian determines which hypotheses seem plausible according to background beliefs or theories, which gained a high posterior probability in Bayesian evaluations in the past. Thus, one might say that Bayesian conditionalisation plays a role in the generation of hypotheses for consideration, just as much as IBE does. In fact, in any two-stage process comprising generation of alternatives and selection amongst them where the generation process depends on some products of previous selection, one might say that the selection criteria are to some extent involved in the generation process. The selection process could be IBE or Bayesianism--or even natural selection, as a biological analogy Lipton draws makes clear (Lipton [2004], pp. 150-151). It seems more appropriate to argue that the role of previous empirical investigation in determining the hypothesis space is essentially similar for Bayesianism and for IBE, rather than that IBE governs or completes the Bayesian account.

On the other hand, Huemer and Weisberg point to an area in which some objective forms of Bayesianism based on the Principle of Indifference have well-recognised difficulties. They suggest that explanatory considerations are to complement objective Bayesianism by helping to determine the hypothesis space to which the Principle of Indifference should be applied. However, to provide a compatibilist reply to van Fraassen, it is not sufficient to just show that explanatory considerations may have a place in a Bayesian setting. One also needs to establish that the proposed combination of objective Bayesian constraints and explanatory considerations amounts to a probabilistic version of IBE.

To see this, let us return to van Fraassen's example with the die. The starting point for that example is a hypothesis space of different hypotheses about the bias of the die. It is assumed that Bayesian Peter gives each hypothesis equal probability. Thus the case is quite consistent with the recommendation that we apply the Principle of Indifference to a hypothesis space in which hypotheses are potential explanations of the data. The hypotheses about the bias of the die are potential explanations of the particular outcomes of throws of the die. On Huemer's view this is because objective chances are explanatorily prior to particular events (Huemer [2009a], p. 359). However, even given this hypothesis space, there is still a question of whether IBE contradicts conditionalisation. Van Fraassen's challenge concerns the relationship between IBE and Bayesian conditionalisation in the context of justification, taking a certain hypothesis space as a starting point. Even if explanatory considerations do help to determine that hypothesis space, we still need an argument to establish that the assessment of the explanationist on that hypothesis space will be fully taken into account by a reasonable Bayesian who conditionalises on the evidence. In the next section, I will offer such an argument.

7. Emergent compatibilism

According to van Fraassen, in order to take explanatory considerations into account, the Bayesian would have to do something different from what she would be inclined to do otherwise. Thus, the rule of IBE conflicts with the Bayesian rule of conditionalisation. In response, compatibilists have argued that no such conflict arises, and that one can have a probabilistic version of IBE which does not violate conditionalisation. The exact form of this response depends on what version of Bayesianism one has in mind.

On the one hand, if one considers subjective Bayesianism, there is no strong sense of what the Bayesian would 'do otherwise'. There are many permissible options. In this case, compatibility can be achieved if the Bayesian follows the explanationist lead and assigns priors and/or likelihoods which lead to preference for the best explanation. I have suggested that the challenge here is primarily to provide an independent motivation for this recommendation.

On the other hand, if the Bayesianism in question is of a more objective variety, the main suggestion so far is that explanatory considerations should serve as further constraints which are to supplement whatever other objective norms are in place. I have argued that this response does not succeed in establishing that a Bayesian version of IBE is possible. In my view, a better approach is to ask whether a Bayesian who adopts constraints on her probabilities which are reasonable on her own terms, would end up favouring more explanatory theories. If so, this would give us an alternative conception of the relationship between IBE and Bayesianism, on which IBE would 'emerge' without the Bayesian doing anything differently from what she would anyway. I call this kind of position 'emergent compatibilism'.

In the following, I consider some examples which suggest that emergent compatibilism may be a viable option. The approach that I will take is first to provide a more in-depth analysis of what is involved in IBE by focusing on specific cases (section 7.1), and then to show they are taken into account by a Bayesian making natural assumptions (section 7.2).

7.1 Analysis of IBE

What kinds of considerations go into determining the 'best explanation'? Discussions of IBE, particularly Lipton ([2004]), have suggested that IBE should be spelled out primarily as 'Inference to the Loveliest Explanation' where the hypothesis which provides the 'loveliest' explanation is 'the one which would, if correct, be the most explanatory or provide the most understanding' (Lipton [2004], p. 59). Loveliness is then further spelled out in terms of explanatory virtues such as simplicity, unification and so on. It is distinguished from 'likeliness', in the sense of how likely the hypothesis is to be true. However, as Lipton admits, a 'defensible' version of IBE cannot be simply Inference to the Loveliest Explanation, since IBE does also depend to some extent on a non-explanatory notion of likeliness. For example, a conspiracy theory might be an excellent explanation, if true--that is, it might be very lovely--but it will not be judged to provide the best explanation if it is also wildly implausible or far-fetched. How plausible, or likely, the hypothesis is at the outset is an additional consideration which acts as a kind of 'sanity check' on the explanation. As a starting point then, there are two components involved in taking one hypothesis to be a better explanation of certain evidence than another. Either the hypothesis should explain the data better, if true, or it should be initially more plausible, or both.

To fill in this account further, we obviously need to say more about what it means for a hypothesis to explain the data better, if true. IBE can operate on different types of hypotheses or theories, and this can make a difference to what kinds of considerations are involved. I first consider the case where the hypothesis space consists of specific hypotheses which do not leave any details or parameters unspecified. Discussion of the relationship between IBE and Bayesianism so far has tended to focus on simple examples such as these. Van Fraassen's example of the alien die is a case in point. Here the hypotheses simply concern the value of a parameter, namely the bias of the die. In section 7.1.1, I consider the case of specific hypotheses, and then in the following section 7.1.2, I consider IBE on a hypothesis space of more complex theories.

7.1.1 IBE on specific hypotheses

I will argue that in cases where the hypotheses are specific, a hypothesis h explains the data better than h', if true, just when h would make the data more expected than h'. In judging which hypothesis renders the data most understandable, we consider nothing more than which hypothesis renders it most expected.

Consider van Fraassen's alien die example. Here we have a number of different hypotheses about the bias which provide potential explanations. Suppose a sequence of heads is thrown. I claim that the hypothesis which makes this outcome most understandable if true is just the hypothesis, which, if true, makes it most expected. The judgement of how expected the outcome is need not be seen as governed by explanatory considerations. In order to determine which hypothesis provides the best explanation, we combine the judgements about the expectedness of the outcome, given the hypotheses, with judgements about how plausible the hypotheses are. Thus IBE would involve two considerations in cases like this: a hypothesis provides a better explanation than another if i) it would make the explanandum more expected if true and/or ii) it is initially more plausible.

Samir Okasha has made a similar point.[x] He offers the following example:
'A mother takes her five-year old child to the doctor. The child is obviously in some distress. On the basis of the mother's information, the doctor forms two competing hypotheses: that the child has pulled a muscle, and that he has torn a ligament; call these H_1 and H_2 respectively. A keen advocate of IBE, the doctor examines the child carefully, and decides that H_2 offers the better explanation of the observed symptoms ...' (Okasha [2000], p. 703).

Okasha suggests that the doctor might justify her reasoning by appeal to two considerations. First, that 'preadolescent children very rarely pull muscles, but often tear ligaments'. And second, that the symptoms, 'though compatible with either diagnosis, are exactly what we would expect if the child has torn a ligament, though not if he has pulled a muscle'. The involvement of these two considerations can be seen by slightly modifying the example. Suppose that the symptoms were only somewhat, rather than exactly to be expected, if the ligament hypothesis were true. Then, intuitively, we would regard the ligament hypothesis as providing a less good explanation. Similarly, suppose that pulling ligaments is actually quite uncommon. Again, intuitively, we would then regard the ligament hypothesis as providing a less good explanation.

On some accounts of explanation, such as Hempel's, it is a requirement on the explanans that it confer a high probability on the explanandum (Hempel [1965]). Some have argued that this is a mistake and suggested that a hypothesis can explain a phenomenon even though it does not make that outcome likely. For example, one can explain recovery from a disease by the fact that the patient was given a particular treatment, even if giving the treatment didn't make it expected that she would recover (see, for example, Salmon [1989], Chap.s 2-3). This type of case does not undermine my claim. Although an explanation need not make the explanandum expected in absolute terms, nonetheless, an explanation would be *better* if it did make the explanandum *more* expected. If the patient were given two treatments, the better explanation of her recovery would still be the treatment which made her recovery more expected, everything else being equal.

7.1.2 IBE on general theories

As we saw in section 3, explanatory virtues like simplicity and unification have been taken as very central to the notion of IBE. However in the examples just considered, they appeared to play no role. I suggest that some of these further features of IBE have been obscured by the tendency to focus on simple examples. In many realistic scientific situations, the theories in the hypothesis space are not just simple unstructured hypotheses. Rather a 'theory' consists of some 'core' claims together with a range of more specific assumptions or 'auxiliaries', which specify particular versions of the theory. In what follows, I will argue that a central factor in the explanatory considerations comprising ‘loveliness’ is the relative responsibility of the core and the auxiliaries in providing an explanation of the phenomena, and that this provides the basis for virtues like simplicity and unification.

Duhem was one of the first to point out that in any test of a given theory, its central or core claims are combined with auxiliary hypotheses which make the theory more specific in a given situation (Duhem [1906]). The idea that any single theory is comprised of both core claims and a set of auxiliaries which evolve over time in response to empirical evidence has been taken up by a number of authors, notably Imre Lakatos (Lakatos [1970]) and John Worrall (Worrall [1999]; Worrall and Mayo [2009]). The core claims of a theory are whatever cannot be altered without it becoming a different theory. Changing the auxiliary hypotheses, on the other hand, produces different more specific versions of the same theory. Often the role of an auxiliary hypothesis is to pin down the value of a ‘free parameter’ in the core of the theory.

For example, it seems appropriate to take Newton’s second law as a core part of Newton’s theory. When Newton’s theory is applied to the planets, it is coupled with specific claims about the positions and masses of the planets which produce specific Newtonian models of the planetary system. These claims are auxiliary assumptions, but they are necessary for producing a theory which is specific enough to be tested against observations in a given domain. Another well known example of core and auxiliary parts of theories comes from 18th century optics.[xi] The core claim of the ‘emission theory’ of light has been taken to be the claim that light consists of ‘projectiles hurled with extreme speed’ (Duhem [1906], p. 189). This was supplemented with specific auxiliary claims about the nature of the particles and forces in different circumstances. On the other hand, the core claim of the ‘wave theory’ of light was taken to be the claim that light consists of ‘vibrations whose waves are propagated within an ether’ (Duhem [1906], p. 189). Again, this needed to be filled in with auxiliary hypotheses giving the details of the nature of the ether and the waves themselves, including parameters such as the wavelength in different specific situations.

These examples indicate the kind of distinction between core and auxiliary which we have in mind. Generally speaking, specifying the cores of different candidate theories is part of the process of coming up with a hypothesis space of candidate theories, or, in other words, it is part of the context of discovery. It is an interesting and difficult question what scientific and/or philosophical criteria determine the appropriate choice of hypothesis space, however I will not attempt to answer this here.[xii] In this discussion, I will set these issues aside, and assume that the hypothesis space, including the appropriate cores of theories, has already been determined.

How good of an explanation a theory provides, if true, is not just determined by the special case of the theory which makes the phenomena most expected. In many cases, several general theories may be capable of making the phenomena highly expected, when combined with certain auxiliary hypotheses. My proposal is that in such situations, one theory provides a lovelier explanation than another the more the core claims of the theory by themselves allow the phenomenon to be explained and the less the explanation is reliant on auxiliary hypotheses.

I will illustrate the basic proposal with the example of how the Copernican and Ptolemaic theories explain the retrograde motion of the planets. Although the treatment that I give is simplified, it is still more complex and realistic than is typical of the literature on IBE, and allows us to see the significance of different levels of theory in the IBE process.

7.1.3 Copernicus vs. Ptolemy

The Copernican and Ptolemaic theories share a common 'background theory' that uniform circular motion is the natural state of motion for heavenly bodies. Thus, it was deemed desirable to explain the planetary appearances in terms of models where planets move uniformly in circles. The two theories differ in their core claims about how the circles are arranged. I take the core claim of the Copernican theory to be the claim that the sun is at the centre of the planetary orbits, and the core claim of the Ptolemaic theory to be the claim that the earth is at the centre. In each case, there are many ways to construct more specific models of the planetary system by filling in the details of number and order of planets, sizes and rates of orbits and so on.[xiii]

There are a number of phenomena to be explained, but I will focus here on the phenomenon of retrograde motion. Every night a visible planet appears in a slightly different position in the sky than it did the night before. Relative to the stars, the planet generally moves in an eastward direction, though every now and again it reverses direction and moves westward for a while before moving eastwards again. When the planet doubles back, it is said to be in 'retrograde motion'. The motion of a planet can be measured either with respect to the stars, or with respect to the sun. The time it takes for a planet to return to the same longitude with respect to the stars is called its 'sidereal period' P. The time it takes to return to the same position with respect to the sun is called its 'synodic period' S. For example, a 'superior' planet, such as Mars, Jupiter or Saturn, will at times be observed to be at an angle of 180° from the sun. Then it is said to be 'in opposition' to the sun. The synodic period is the time between successive oppositions. There is a certain observed regularity between the sidereal and synodic periods

$$(3) \quad \left| \frac{1}{P} - \frac{1}{E} \right| = \frac{1}{S}$$

where E is the solar year, that is, the time it takes for the sun to return to the same position relative to the fixed stars. It is also observed that retrograde motion for a superior planet occurs always when the planet is seen in opposition to the sun.

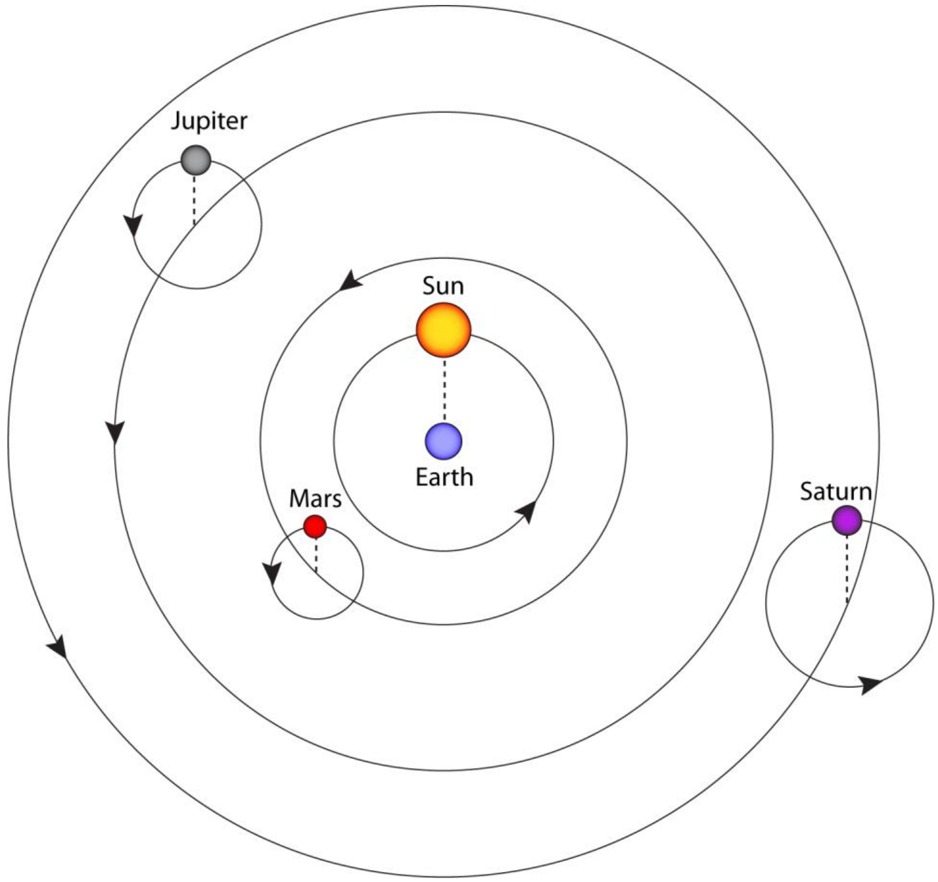

Figure 1: In the Ptolemaic theory, retrograde motion of the planets is explained by epicycles.

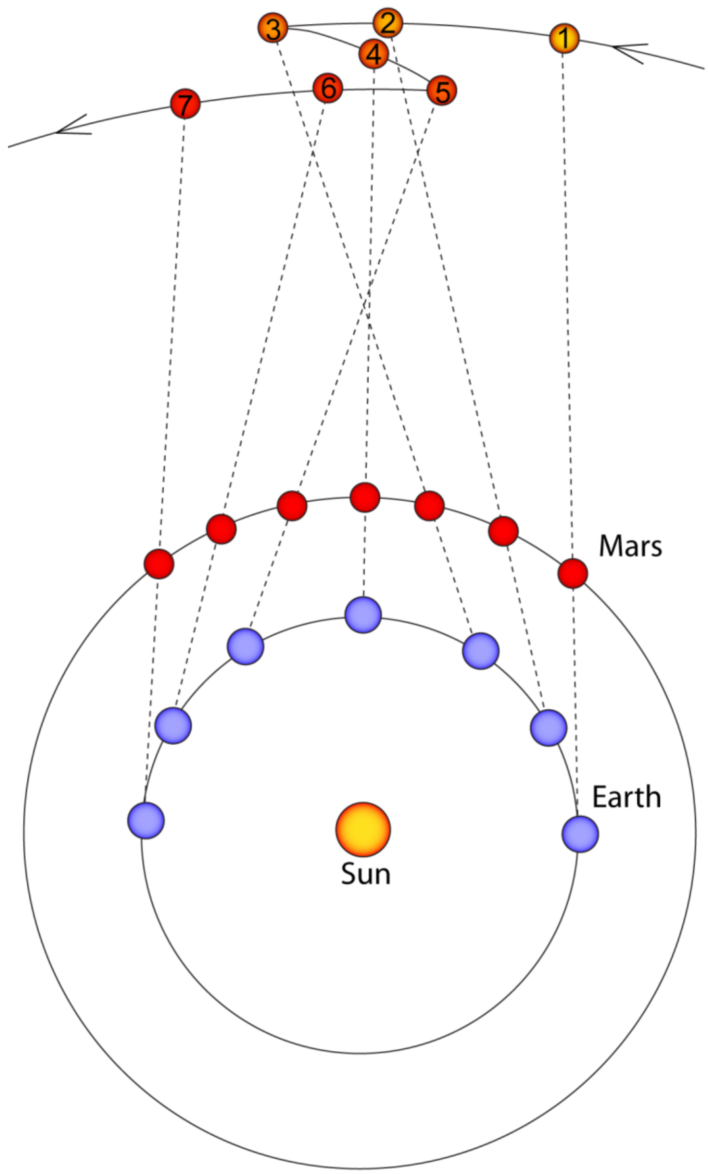

Figure 2: In the Copernican theory, the explanation of the observation of retrograde motion is that the outer planet is overtaken by the faster moving inner planet.

Consider first how these phenomena associated with retrograde motion are explained by the Ptolemaic theory. Each planet is taken to move in a small circle, or ‘epicycle’, whose centre moves in a larger circle (the ‘deferent’), in the same direction around the earth. When traversing the epicycle, at some points the planet is moving backwards with respect to the deferent - when this happens, according to the theory, retrograde motion occurs (see Figure 1). The sidereal and synodic periods are given by the period of the deferent and the epicycle respectively. These two parameters can then be adjusted in such a way as to ensure that equation 3 holds. It can also be ensured that retrograde motion of a superior planet occurs at opposition to the sun by adjusting the epicycle rates in such a way that a line from the planet to the centre of its epicycle is always parallel to the line from the earth to the sun (see Figure 1).

Retrograde motion is explained quite differently by the Copernican theory. At certain times planets overtake each other in their orbits around the sun. When, for example, the earth overtakes a planet which is further from the sun, the planet will appear to be moving ‘backward’ for a certain amount of time: this is the retrograde motion (see Figure 2). With both the earth and the planets orbiting the sun, it becomes inevitable that there will be overtaking of one planet by another as long as the planets are not moving at exactly the same angular velocity. In fact, Copernicus assumes that planets closer to the sun move more rapidly than planets further away. Then the observed relationships between the synodic and sidereal periods are a mathematical consequence of the sun-centred theory.[xiv] It also follows that retrograde motion of a superior planet occurs when the planet appears in opposition to the sun since overtaking happens when the planet, sun and earth lie on the same line.

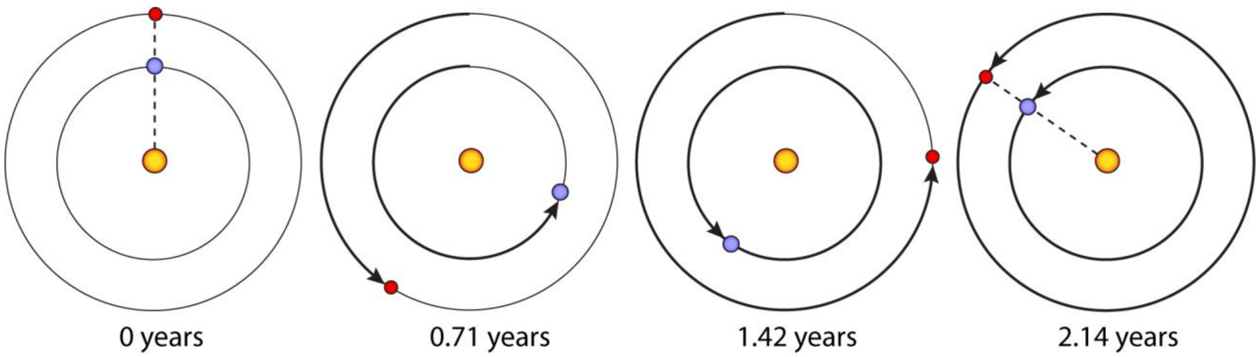

Figure 3: When 2.14 years have elapsed since the last opposition, Earth and Mars will be in opposition again. In the meantime, Earth has completed one more complete orbit around the sun than Mars.

Both the Copernican and the Ptolemaic theories have specific arrangements of orbits and values of parameters which make retrograde motion highly expected. And in both cases, the explanation of the phenomena involves some additional hypotheses not contained in the core of the theory itself. However, the two cases differ in the extent to which the additional hypotheses are essential to the explanation. In the Copernican explanation, the observation of retrograde motion and the coincidence between retrograde motion and opposition for superior planets appear as a direct consequence of placing the sun at the centre. The observed equation 3 also follows directly from this core claim together with the auxiliary hypothesis that outer planets move more slowly. In the Ptolemaic case, on the other hand, the explanation of the occurrence of retrograde motion necessitates extra hypotheses of planetary epicycles. Without the epicycles, retrograde motion would be unexplained by the Ptolemaic theory. Just placing the earth at the centre of the planetary orbits is insufficient to explain either equation 3, or the coincidence of retrograde motion with opposition. It is only by making further specific hypotheses about the epicycle rates that these observations can be accounted for. Thus, the Copernican theory provides the lovelier explanation because accepting its core claim that the sun is at the centre of the planetary orbits allows the phenomenon of retrograde motion to be explained with less dependence on auxiliary hypotheses.

So far, we have only considered the explanation of the phenomenon of retrograde motion. The situation becomes more complicated when taking into account all the phenomena which require explanation. Copernicus' full model also, like Ptolemy's, depended on a number of epicycles. Part of the reason for this was the background assumption that the correct way to construct models was using circular orbits. Later scientists came to consider the question of whether the planetary phenomena might be best explained by models based on elliptical orbits, rather than circular orbits. Again, the explanatory advantage of elliptical models was that they eliminated the need for extra auxiliary hypotheses in explaining the phenomena. Thus, the basic assumptions which serve as common background for the comparison of theories by IBE may themselves be subject to IBE when a theory presenting a rival set of assumptions comes on the scene.

7.14 Explanatory virtues

This case suggests that part of what it means for a theory to provide a lovelier explanation of certain phenomena is that the explanation based on the theory relies to a greater extent on the theory's core claims rather than auxiliary hypotheses. How does this fit in with the common way of characterising IBE in terms of 'explanatory virtues'?

Copernicus himself claimed the advantage of simplicity for his theory, saying that he was able to find an arrangement of circular orbits that explained the phenomena 'with fewer and much simpler constructions than were formerly used' (Copernicus 1878, p. 58). It has been suggested by Paul Thagard that the dependence on core claims rather than auxiliary hypotheses is just what the explanatory virtue of simplicity amounts to (Thagard [1978]). According to Thagard, 'the explanation of facts F by a theory T requires a set of given conditions C and also a set of auxiliary hypotheses A ', where conditions C are accepted independently of T or F, whereas 'an *auxiliary hypothesis* is a statement, not part of the original theory, which is assumed in order to help explain one element of F, or a small fraction of the elements of F'. Thagard proposes that 'simplicity is a function of the size and nature of the set A needed by a theory T to explain facts F'. This way of looking at simplicity fits with the widely expressed intuition that adding adjustable parameters makes a theory less simple (Popper [1959]; Lakatos [1970]; Rosenkrantz [1977]; Howson and Urbach [2006]), since specifying the values of adjustable parameters in a theory may be seen as a special case of auxiliary hypotheses.

Another virtue that the Copernican theory has over the Ptolemaic theory is unification. The Copernican theory correlates phenomena, such as the synodic and sidereal period of a given planet, or the periods of different planets, which are unrelated according to the Ptolemaic theory (Myrvold [2003]). This feature may also be understood in terms of dependence of the explanation on core claims of the theories as opposed to auxiliaries. The Ptolemaic theory relies on auxiliary hypotheses which are 'tuned' to the particular domain. For example, it uses hypotheses about the period of the deferent of the planet to explain the observations of the planet's sidereal period, and another set of hypotheses about the period of the epicycle to explain observations of the synodic period. In the Copernican theory, by contrast, because one does not have a separate parameter for each of these phenomena, the phenomena are unified by the structure and parameters which are a common part of the explanation for both.

Our examples suggest then that simplicity and unification are very closely related explanatory virtues, both corresponding to explaining certain phenomena in terms of core elements of the theory as opposed to auxiliary hypotheses. Of course, the concept of unification only applies when the phenomena to be explained come from various domains, whereas simplicity applies both where the phenomena come from different domains as well as where there is just one phenomenon to be explained. Simplicity and unification may be regarded as different ways of referring to a basic feature of better explanation. However, they are both virtues corresponding to how well a theory explains a given set of phenomena. Another aspect of better explanation is providing an explanation of more phenomena or more various phenomena. This virtue, which is sometimes called 'scope' or 'consilience' typically trades off against the virtue of simplicity or unification, since it may be necessary to bring in more specific auxiliary hypotheses in order to allow the theory to provide explanations across a wider range of domains. The explanatory performance of a theory may also be judged over time. Thus some theories have the virtue of 'fruitfulness', meaning that they are able to provide simple explanations of an increasingly wide range of phenomena as more evidence is collected.

7.1.5 Summary

I suggest that there is a certain conceptual 'skeleton' of IBE which governs how it assigns degrees of belief. There are two considerations involved in IBE: first, the 'loveliness' of the hypothesis with respect to the phenomena to be explained--that is, how well the hypothesis would explain the phenomena, if it were true--and second, the initial plausibility of the hypothesis. In cases where the hypotheses in question are simple and unstructured, how lovely the hypothesis is simply amounts to how expected it would make the evidence if it were true, whereas in cases where the hypotheses are general theories comprising core claims and auxiliaries, how lovely a hypothesis is depends also on the extent to which it provides an explanation in terms of its core claims without relying on auxiliary hypotheses.

7.2 Bayesian account

I will now consider how the Bayesian may take into account the same central considerations that make up the skeleton of IBE which I have isolated in the previous section. For the Bayesian, the posterior probability assigned to a particular theory depends on its prior probability and its likelihood. The prior takes into account the same considerations of initial plausibility as are involved in IBE. Similarly, I will now argue that in the examples I have considered, the likelihood takes into account the same considerations as the 'loveliness' component of IBE.

In the case of simple hypotheses such as dice throwing, the parallel is clear. We saw that in such cases loveliness involves little more than expectedness of the evidence given the hypothesis, which is exactly what the Bayesian likelihood also represents.

For more general theories such as the Copernican and Ptolemaic theories, we have seen that the loveliness does not simply depend on a theory having the best-fitting special case. Both the Copernican and the Ptolemaic theories arguably fit the data on retrograde motion very well given suitably arranged auxiliaries. The further factor involved in the explanatoriness of the theories is how dependent the explanation they provide is on the core of the theory as opposed to the auxiliaries. The Copernican theory provides the better explanation of the phenomena because it explains more from its core and is less dependent on auxiliaries than the Ptolemaic theory.

It has been well-recognised that the Bayesian likelihood may, under certain seemingly natural choices of priors, reflect explanatory virtues, particularly simplicity (Rosenkrantz [1977]; Jefferys and Berger [1992]; MacKay [2003]; Huemer [2009b]; Henderson et al. [2010]). To see this, consider the Bayesian likelihood for general theories like the Copernican and Ptolemaic theories. This is not just the likelihood of the best-fitting special case of the theory, but is an average of the likelihoods across all the specific models that the theory allows. Let the free parameters of the general theory T be θ. Then the likelihood of T, given D, is:

$$p(D|T) = \int p(D|\theta,T)p(\theta|T)d\theta$$

In the average, the likelihoods of specific models $p(D|\theta,T)$ are weighted by the prior probability density conditional on the theory, $p(\theta|T)$. Let us suppose for concreteness that the data in question in the Ptolemaic and Copernican case are the observations of the sidereal and synodic periods of the planets.

Depending on how the prior over parameters $p(\theta|T)$ is chosen, the Bayesian likelihood for the Copernican theory may be higher or lower than the likelihood for the Ptolemaic theory. However, there is a strong intuition that in the absence of any information that parameter values fall in a particular region of the parameter space, it will be most natural to spread the prior probability density out over the whole parameter space. The thought is, in part, that it would be unreasonable to concentrate the prior in any particular subregion of the parameter space without good reason to do so, since in so doing one might miss out on assigning probability to the correct value. The key point here is that the underlying motivation for spreading the prior may be quite independent of considerations of which of the two theories one regards as more explanatory.

If one does spread the priors in this way, it turns out that the Copernican theory is favoured by the Bayesian for the same reason as it is favoured by the explanationist. In the Copernican theory, both the sidereal and the synodic periods depend on the parameters corresponding to the periods of the deferents. There is a specific relation (expressed by equation 3) that the synodic and sidereal periods are expected to stand in, and this holds in all the specific cases of the Copernican theory. The relation is expected on the basis of the core claims of the theory, and not on the basis of any particular parameter settings. In the Ptolemaic theory, on the other hand, the relation in equation 3 only holds for some parameter settings, but not for others, since it is not determined by the core of the theory. In the Copernican theory, the prior probability for specific models is only distributed over cases in which equation 3 holds, whereas the Ptolemaic prior is spread out over a larger parameter space that includes many possibilities which do not fit the relation. This is because the Ptolemaic theory has extra free parameters corresponding to the periods of the epicycles which can be tuned to account for observations of any synodic periods. Thus it can account for any relationship between the synodic and sidereal periods. Since the data do in fact conform to the relationship expressed in equation 3, most of the additional possibilities allowed by the Ptolemaic theory have poor fit and serve only to bring down the average likelihood of the Ptolemaic theory compared to the Copernican theory. Thus, with prior assignments that may be thought natural, the Bayesian likelihood favours the Copernican theory because it does not rely on auxiliaries in the form of adjustable parameters to as great a degree as the Ptolemaic theory.

In order that the Bayesian favours the Copernican theory, it is not necessary that a higher prior probability be assigned to the Copernican theory than to the Ptolemaic theory. The two theories could be assigned the same prior, yet the Ptolemaic theory would still be penalised in the likelihood.

This example illustrates an alternative way in which IBE and Bayesianism may be related compatibly to that suggested by the constraint-based compatibilist. Here the explanatory considerations do not drive the inference, but rather emerge from the Bayesian account. On this picture, the Bayesian account serves as an explication of what is going on in IBE.

In the broader discussion in which van Fraassen's incompatibilist position is situated, the normative status of IBE is what is in question. Van Fraassen uses the alleged incompatibility of IBE with conditionalisation to argue that IBE is not a normatively compelling rule. A Bayesian explication of IBE may help us to focus our concerns about ways in which IBE may be problematic from a normative point of view. In particular, in order for the Bayesian account to vindicate the normative status of IBE, we would need to be able to justify fully the choice of the 'natural' prior. However, norms to which we might appeal to support it, like tempering, or the maximum entropy principle, have proved quite controversial within the Bayesian tradition. I will not attempt to discuss this further here, but what I have said so far gives an indication of how emergent compatibilism may provide the basis for some specific ways of tackling the problem of IBE's normative standing.

By contrast, the constraint-based compatibilist, at least in the normative version, merely begs the question against van Fraassen's conclusion. The constraint-based compatibilist finds a way to preserve compatibility between IBE and conditionalisation, but only by presupposing that IBE is normatively compelling. For only then can the constraint-based compatibilist recommend that IBE place normative constraints on the assignment of Bayesian probabilities.

Emergent compatibilism would not make IBE redundant. IBE and Bayesianism might be regarded as different levels of description of the inference process, where IBE describes the gross phenomena of inferential thinking, whereas the quantitative detail which gives rise to these explanatory features is captured by the Bayesian model. As a suggestive analogy, consider the different levels of description provided by thermodynamics and statistical mechanics, where thermodynamics describes thermal phenomena at the macroscopic level and statistical mechanics provides a lower-level description of the molecular underpinnings. Thermodynamic concepts like heat, entropy and energy are not made redundant by statistical mechanics. Rather they are taken to refer to certain properties of collections of molecules, properties which may themselves be represented partially in terms of probabilities. Similarly, explanatory concepts like simplicity, scope and unification are not made redundant by the Bayesian account. They refer to certain properties of the relationship between a theory and the evidence, properties which may in turn be represented in terms of probabilities.

8. Conclusion

Most discussion of the relationship between IBE and Bayesianism has focused on how to relate their different methods of representing and updating probabilities on hypotheses in a given hypothesis space. In this context, van Fraassen has suggested that IBE and Bayesianism are incompatible because IBE involves additional explanatory considerations which would have to be accommodated in a probabilistic setting by adding bonus probability onto more explanatory hypotheses beyond that they would have received by conditionalisation. I have argued that van Fraassen's own motivating example does not support this interpretation of IBE.

The alternative view is that IBE and Bayesianism are compatible: that is, a probabilistic version of IBE does not contradict conditionalisation. The problem is to explain why this should be so. To date, the main compatibilist argument has been that explanatory considerations help to determine the assignment of priors and/or likelihoods. I call this position 'constraint-based compatibilism'. I have argued that constraint-based compatibilism does not provide a wholly satisfactory response to the incompatibilist challenge.

Instead I have presented examples which suggest an alternative reason that the explanatory considerations involved in IBE are taken into account by the Bayesian. The first task was to determine what these considerations are. I suggested the following conceptual skeleton of IBE. A theory provides a better explanation of certain evidence than another theory to the extent that a) the theory itself has more initial plausibility and b) it is 'lovelier', that is, compared to the other theory, it would explain the evidence better if it were true. I then showed how in certain realistic examples, the Bayesian, making assignments of probabilities on her own terms, takes these considerations into account. If the theory is more plausible, it gets a higher prior, and if it is lovelier, it gets a higher likelihood. Cases of the latter connection fell into two categories. For cases of specific hypotheses, there appeared to be little more to the loveliness of a hypothesis than how expected it makes the phenomena. Thus the Bayesian would in such cases assign the higher likelihood to the lovelier hypothesis. For cases of general theories, such as those of Copernicus and Ptolemy, I proposed that the lovelier theory is that whose explanation relies more on the core of the theory, as opposed to the auxiliaries. With certain independently motivated choices of priors, the Bayesian likelihood favours the lovelier theory, because of the way it averages over specific cases that the theory allows. It was not necessary to assume, as the constraint-based compatibilist does, that explanatory considerations help to determine priors or likelihoods.

This study points the way toward an alternative strategy that compatibilists could take in response to van Fraassen's incompatibilist challenge. This is an 'emergent compatibilist' view according to which explanatory considerations emerge from the Bayesian framework rather than providing constraints on the assignment of probabilities. The methods would be compatible, not because they can be amalgamated, but rather because they capture substantially similar epistemic considerations. Emergent compatibilism would allow the two methods to be mutually illuminating. IBE provides a way to describe in qualitative terms what is going on in Bayesian inference, with the Bayesian account providing a systematic way to explain how explanatory virtues are preferred in IBE.

Department of Philosophy
Baker Hall 135
Carnegie Mellon University
Pittsburgh PA 15213
Email: leahh@andrew.cmu.edu

Acknowledgements

I would like to thank Selim Berker, Alex Byrne, Ned Hall, Kareem Khalifa, Bernhard Nickel, Agustín Rayo, Jan-Willem Romeijn, Jonah Schupbach, Robert Stalnaker, Joshua Tenenbaum, Judith Thomson, Roger White, John Worrall and Steve Yablo for making useful comments on earlier versions of this paper. I have benefitted from clarifying discussions with Teddy Seidenfeld in making the final revisions. I would also like to thank two anonymous referees for this journal for their constructive criticisms.

References

Christenson, D. [2001]: 'Preference Based Arguments for Probabilism'. *Philosophy of Science*, **68**, pp. 356--376.

Copernicus, N. [1878]: 'Commentariolus', in E. Rosen (*ed*.), *Three Copernican Treatises*, New York: Columbia University Press, pp. 55--90.

Darwin, C. [1859]: *On the Origin of Species by Means of Natural Selection*, New York: Collier, 6th edition.

Duhem, P. [1906]: *The Aim and Structure of Physical Theory*, Princeton: Princeton University Press.

Harman, G. [1965]: 'The Inference to the Best Explanation'. *Philosophical Review*, **74**, pp. 88—95.

Harman, G. [1970]: 'Induction: A Discussion of the Relevance of the Theory of Knowledge to the Theory of Induction (with a Digression to the Effect That Neither Deductive Logic Nor the Probability Calculus Has Anything to Do with Inference)', in M. Swain (*ed*.), *Induction, Acceptance and Rational Belief*, Dordrecht, Holland: D. Reidel.

Hempel, C. G. [1965]: *Aspects of Scientific Explanation*, New York: Free Press.

Henderson, L., Goodman, N. D., Tenenbaum, J. B. and Woodward, J. F. [2010]: 'The Structure and Dynamics of Scientific Theories: A Hierarchical Bayesian Perspective', *Philosophy of Science*, **77**, pp. 172--200.

Howson, C. and Urbach, P. [2006]: *Scientific Reasoning: The Bayesian Approach*, Chicago and La Salle, Illinois: Open Court.

Huemer, M. [2009a]: 'Explanationist Aid for the Theory of Inductive Logic'. *British Journal for the Philosophy of Science*, **60**, pp. 345--375.

Huemer, M. [2009b]: ‘When is Parsimony a Virtue?’. *Philosophical Quarterly*, **59**, pp. 216--236.

Jaynes, E. T. and Bretthorst, G. L. [2003]: *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press.

Jefferys, W. H. and Berger, J. O. [1992]: ‘Ockham’s Razor and Bayesian Analysis’, pp. 64--72.

Jeffrey, R. C. [1992]: *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.

Jeffreys, H. [1998]: *Theory of Probability*, Oxford: Oxford University Press, 3rd edition. First edition 1939.

Kaplan, M. [1981]: ‘A Bayesian Theory of Rational Acceptance’, *Journal of Philosophy*, **LXXVIII**, pp. 305--330.

Kyburg, H. [1961]: *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University 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*, Cambridge: Cambridge University Press.

Lakatos, I. and Zahar, E. [1975]: ‘Why Did Copernicus’ Research Program Supersede Ptolemy’s?’, in R. S. Westman (*ed*.): *The Copernican Achievement*, Berkeley: University of California Press, pp. 354--384.

Lewis, D. [1986]: ‘A Subjectivist’s Guide to Objective Chance’, in *Philosophical Papers Volume II*, Oxford UK: Oxford University Press, pp. 83--113.

Lipton, P. [2004]: *Inference to the Best Explanation*, London, UK.: Routledge, 2nd edition.

MacKay, D. J. C. [2003]: *Information Theory, Inference and Learning Algorithms*, Cambridge: Cambridge University Press.

Makinson, D. [1965]: ‘The Paradox of the Preface’, *Analysis **25***, pp. 205--207.

Myrvold, W. [2003]: ‘A Bayesian Account of the Virtue of Unification’, *Philosophy of Science*, **70**, pp. 399--423.

Niiniluoto, I. [1999]: ‘Defending Abduction’. *Philosophy of Science*, **66**, S436--S451.

Okasha, S. [2000]: ‘Van Fraassen’s Critique of Inference to the Best Explanation’, *Studies in History and Philosophy of Science*, pp. 691--710.

Pollock, J. and Cruz, J. [1999]: *Contemporary Theories of Knowledge*, Maryland: Rowman and Littlefield.

Popper, K. R. [1934], *The Logic of Scientific Discovery*, London, UK: Hutchinson and co.

Rosenkrantz, R. D. [1977]: *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*, Dordrecht, Holland / Boston, USA: Synthese Library.

Salmon, W. C. [1989]: *Four Decades of Scientific Explanation*, Minneapolis: University of Minnesota Press.

Seidenfeld, T. [1979]: ‘Why I Am Not an Objective Bayesian; some Reflections Prompted by Rosenkrantz’, *Theory and Decision*, **11**, pp. 413--440.

Shimony, A. [1970]: ‘Scientific inference’, in: R. G. Colodny (*ed.*), *The Nature and Function of Scientific Theories: Essays in Contemporary Science and Philosophy*, Pittsburgh: University of Pittsburgh Press, pp. 79--172.

Sober, E. [1990]: ‘Let’s Razor Ockham’s Razor’, in: D. Knowles (*ed*.), *Explanation and its Limits*, UK: Cambridge University Press, pp. 73--93.

Sturgeon, S. [2008]: ‘Reason and the Grain of Belief’, *Nous*, **42**, pp.139--165.

Thagard, P. [1992]: *Conceptual Revolutions*, Princeton, New Jersey: Princeton University Press.

Thagard, P. R. [1978]: ‘The Best Explanation: Criteria for Theory Choice’, *Journal of Philosophy*, **75**, pp. 76--92.

van Fraassen, B. C. [1980]: *The Scientific Image*, Oxford: Oxford University Press.

Van Fraassen, B. C. [1989]: *Laws and Symmetry*, Oxford, UK: Clarendon Press.

Weisberg, J. [2009]: ‘Locating IBE in the Bayesian Framework’, *Synthese*, **167**, pp. 125--144.

Williams, G. C. [1966]: *Adaptation and Natural Selection: A Critique of some Current Evolutionary Thought*, New Jersey: Princeton University Press.

Worrall, J. [1999]: 'New Evidence for Old', in P. Gardenfors, J. Wolenski, and K. Kijania-Placek (*eds*.): *In the Scope of Logic, Methodology and Philosophy of Science*, Dordrecht: Kluwer, pp. 191--208.

Worrall, J. and D. G. Mayo [2009]: 'Theory Confirmation and Novel Evidence', in D. G. Mayo and A. Spanos (*eds*.), *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, Cambridge: Cambridge Books, pp. 125--169.

[i] One might, for example, think that only one or other type of attitude is appropriate for epistemology. Jeffrey [1992], Christenson [2001], and Howson [2006], for instance, favour degrees of belief, whereas Harman [1970] and Pollock and Cruz [1999] favour full belief. On some views, also, there are difficulties in connecting full belief with degrees of belief, due to the lottery and preface paradoxes (Kyburg 1961; Makinson [1965]; Kaplan 1981; Sturgeon [2008]).

[ii] The probability $p(D)$ is essentially a normalisation constant which ensures that the posterior distribution is a valid probability distribution which sums to 1. It can be computed from the 'likelihood function' which is a function of the hypotheses T, giving the likelihood for each T given particular D, as:

$$(2) \qquad p(D) = \sum_{T \in \mathbf{H}} p(D \mid T)\, p(T)$$

(the sum is replaced by an integral if the hypotheses T are continuously varying quantities).

[iii] Objective Bayesian positions are set forth in Jeffreys [1998], Jaynes and Bretthorst [2003] and

Rosenkrantz [1977]. They are widely criticised, for example in Seidenfeld [1979].

[iv]Okasha alleges that van Fraassen has simply produced an 'idiosyncratic' interpretation of IBE in the probabilistic context (Okasha [2000], p. 703). However, in my view, van Fraassen's interpretation is not so much idiosyncratic as based on the way IBE is put to use in certain contexts as an argument against anti-realism. One of the key arguments against the underdetermination of theories by evidence is that explanatory considerations provide extra evidence for the truth of a theory which go beyond any evidence we may have for a theory's empirical adequacy (Van Fraassen [1980], pp. 153-155). On this view, IBE would be expected to contain extra elements which cannot be fitted into a standard Bayesian story about confirmation. Hence these may be interpreted as adding bonus probability onto the Bayesian probabilities.

[v] Huemer, unlike Weisberg, does not present his view as a response to van Fraassen's incompatibilism.

[vi]Lipton himself does not offer any examples. The example I offer here is based on Huemer [2009], pp. 364-365, who uses it in service of a different point.

[vii]Okasha also says that Bayesianism is 'silent' in the context of discovery (Okasha [2000], pp.706-707)

[viii]Lipton's position appears to be inconsistent. On the one hand, he regards simplicity and

unification as essential ingredients of 'loveliness', or how explanatory a hypothesis is with respect to certain evidence (Lipton [2004], p. 122). But on the other, he acknowledges that the explanatory considerations involved in loveliness cannot constrain the priors since they are relative to the new evidence (Lipton [2004], p. 113).

[ix] Lipton turns to the explanationist, because he claims that the Bayesian does not provide a 'recipe we can readily follow' (Lipton [2004], p. 113).

[x]In some places, Okasha puts forward a view like this. In other places, his remarks suggest constraint-based compatibilism, such as when he talks about using explanatory considerations 'as an aid for calculating the priors and likelihoods' (Okasha [2000], p.703).

[xi]This example is discussed by Duhem [1906] and extensively by Worrall (Worrall [1999]; Worrall and Mayo [2009]).

[xii] As we have seen in section 6, some role for 'explanatory considerations' in this domain is not precluded.

[xiii]Lakatos [1975] analyses the case of Ptolemy and Copernicus in terms of the 'hard core' of the research programmes and the use of ad hoc assumptions.

[xiv] At opposition, the earth, sun and planet all lie on a line. A superior planet will again be in opposition when it is ‘overtaken’ by the earth at a time when the earth has completed one more orbit around the sun than the planet has (see Figure 3). The planet is moving at an angular velocity of $\frac{360}{P}$ and the earth is moving at angular velocity $\frac{360}{E}$. When opposition occurs again, the planet has moved through an angle $\frac{360}{P}S$, and this is equal to the angle through which the earth has moved $\frac{360}{E}S$ minus 360^0 (see Figure 3). Equating these angles yields the observed relation expressed in equation 3. Similar considerations yield the relationship for inferior planets.