STATHIS PSILLOS

INFERENCE TO THE BEST EXPLANATION AND BAYESIANISM*

Comments on Ilkka Niiniluoto's "Truth-seeking by Abduction"

1. INTRODUCTION

Niiniluoto (2003) has offered an incisive and comprehensive review of the recent debates about abduction. There is little on which I disagree with him. So, in this commentary, I shall try to cast some doubts to the attempts to render Inference to the Best Explanation (IBE) within a Bayesian framework.

Lately, there has been a lot of discussion about the place of IBE in Bayesian reasoning. Even Niiniluoto argues that "Bayesianism provides a framework for studying abduction and induction as forms of ampliative reasoning" (2003, 15). There is a tension, however, at the outset. Bayesian reasoning does *not* have rules of acceptance. On a strict Bayesian approach,[1] we can never detach the probability of the conclusion of a probabilistic argument, no matter how high this probability might be. So, strictly speaking, we are never licensed to accept a hypothesis on the basis of the evidence. All we are entitled to do, we are told by strict Bayesians, is a) to detach a conclusion about a probability, viz., to assert that the posterior probability of a hypothesis is thus and so; and b) to keep updating the posterior probability of a hypothesis, following Bayesian conditionalisation on fresh evidence. But IBE is typically seen as a rule of acceptance. In its least controversial form, IBE authorises the *acceptance* of a hypothesis H, on the basis that it is the best explanation of the evidence. Think of the standard IBE-based argument for the existence of middle-sized material objects. According to this, the best explanation of the systematic, orderly and coherent way we experience the world is that there are stable middle-sized material objects which cause our experiences. Presumably, those who endorse this argument do not just assert a conclusion about a probability; they assert a conclusion *simpliciter*. That is, their claim is not that the probability that material objects exist is high, but rather that it is reasonable to accept that they do exist. Hence, there is a tension between Bayesianism and standard renderings of IBE. This might make us wary of attempts to cast IBE in a Bayesian framework. But this is only the beginning of our worries.

Niiniluoto surveys a variety of recent results about the connection between IBE and Bayesian confirmation. They are all invariably instructive. But I want to challenge the *motivation* for attempting this. There are two questions to be asked.

F. Stadler (ed.), *Induction and Deduction in the Sciences*, 83–91.

First, if we were to cast IBE within a Bayesian framework, could we do it? I do not doubt that this can be done (given the flexibility of the Bayesian framework). But I shall raise some worries about the ways that it can done. These worries will usher in the need the raise a *second* question, viz., should we want to cast IBE within a Bayesian framework? This question, I think, is more intriguing than the first.

2. IBE AND BAYESIAN KINEMATICS

The crux of IBE, no matter how it is formulated, is that explanatory considerations should inform (perhaps, determine) what it is reasonable to believe. Now there are several ways to import explanatory considerations into a Bayesian scheme. There is a contentious one, due to Bas van Fraassen (1989). He claims that the right way to cast IBE within a Bayesian framework is to give *bonuses* to the posterior probabilities of hypotheses that are accepted as best explanations of the evidence. That is, *after* having fixed the posterior probability of a hypothesis in a Bayesian way, if this hypothesis is seen as the best explanation of the evidence, then it is entitled to a rise of its posterior probability. It's not hard to see that if one followed this way of updating one's degrees of belief, one would end up with Dutch books. In fact, this is precisely the strategy that van Fraassen himself follows in order to argue that, as a rule of updating degrees of belief, IBE is incoherent. But why should one take van Fraassen's suggestion seriously? As we have seen, the key recommendation of IBE is that explanatory considerations should inform (perhaps, determine) what we reasonably come to believe. So, if one were to cast IBE within a Bayesian framework, one should make sure that explanatory considerations are *part* of the Bayesian kinematics for the determination of the posterior probability of a theory, and not something that should be added on to confer bonus degrees of belief to the end product.

Given the Bayesian machinery, there are two ways in which explanatory considerations can be *part* of the Bayesian kinematics. They should either be reflected in the prior probability of a theory, relative to background knowledge, or (inclusively) in the likelihood of the theory. Niiniluoto shows what conditions should be satisfied vis-à-vis the priors and the likelihoods so that the best explanation is also the best confirmed theory (or that the better explanations receive the better confirmation). But better confirmation (even high confirmation) falls short of rightful acceptance. So, some of the excitement of IBE, as a rule of acceptance, is lost. But we have noted this already. An important further problem is this. Though priors and likelihoods *can* reflect explanatory judgements, it is clear that they fail to discriminate among competing hypotheses with the *same* priors and the *same* likelihoods. This problem is particularly acute for the case of likelihoods. The whole point of insisting on IBE is that it promises rationally to resolve observational ties. When two or more competing hypotheses entail the

very same evidence, then their likelihoods will be the same. Hence, likelihoods cannot resolve, at least some, (perhaps the most significant), observational ties.

But things get worse if we base our hopes on likelihoods. One thought, explored by Niiniluoto (2003, 22) might be to equate the best explanation with the hypothesis that enjoys the highest likelihood. But, as we are about to see, this is deeply problematic. In fact, as the so-called base-rate fallacy shows, likelihoods are relatively mute. If explanatory considerations enter the Bayesian story via likelihoods, then so much the worse for the explanatory considerations.

3. LIKELIHOODS AND THE BASE-RATE FALLACY

One way to introduce the base-rate fallacy is via the so-called Harvard Medical School test. Here are some details. A test for the presence of a disease has two outcomes, 'positive' and 'negative'. Let's call them *e* and *not-e*. Let a subject (Joan) take the test and let H be the hypothesis Joan has the disease. The test is highly reliable: it has zero *false-negative rate*. That is, the likelihood that the subject tested negative given that she *does* have the disease is zero (i.e., prob(not-e/H) = 0). Consequently, the *true-positive rate*, i.e., the likelihood of being tested positive given that she *does* have the disease is unity, (prob(e/H) =1). But the test also has a very small *false-positive rate*. That is, the likelihood that the subject is tested positive though she *doesn't* have the disease is, say, 5% (prob(e/not-H) = .05). Now, Joan takes the test and she tests positive. In the standard literature of the base-rate fallacy, given the above details, the following question is asked: what is the probability that Joan has the disease given that she tested positive? That is, what is the posterior probability prob(H/e)?

If we try to answer this question in a Bayesian framework, then it is clear that there is some crucial information missing: we are not given the incidence rate (base-rate) of the disease in the population. In other words, we are not given the prior probability of the hypothesis that the subject has the disease before she takes the test, i.e., prob(H). If this incidence rate is very low, e.g., if only 1 in 1,000 in the population has the disease, then it can be easily shown that it is very unlikely that Joan has the disease given that she tested positive: prob(H/e) would be less than .02.

It is notorious that when this problem is posed to experimental subjects, they tend, with overwhelming majority, to answer that the probability that Joan has the disease given that she tested positive is very high – very close to 95%. The so-called base-rate fallacy is that experimental subjects who are given the problem above tend to neglect base-rate information (that is, they tend to neglect the prior probabilities),[2] *even when* they are given this information explicitly. Several conclusions have been drawn from it and the relevant literature is massive. (Characteristically, one of the conclusions is that ordinary people are not Bayesians; another one is that ordinary people do not reason rationally because they do not follow Bayes's rule.) Suppose we asked the experimental

subjects: what is the best explanation of the evidence? That is, given that Joan tested positive in a highly reliable test, what is the best explanation of this fact? Now, this is *not* a question about probabilities. It is more like a question about what it is reasonable to accept about this particular case. Hence, it would not, it seems, be unreasonable for them to argue that the best explanation of the evidence is that Joan has the disease. But let's leave all this to one side.[3] The point I want to focus on is not whether and in what sense the base-rate neglect is indeed a fallacy. My point is simply that the base-rate fallacy (no matter how one reads it) shows that it is incorrect just to equate the best explanation of the evidence with the hypothesis that has the highest likelihood. As we saw above, it turns out that, if we consider just the likelihood of a hypothesis, and if we think that this is the way to determine the best explanation, then there is no determinate answer to the question 'what is the best explanation of the evidence?'. A very small prior probability can dominate over high likelihood and lead to a very small posterior probability. Let me put the point in a more conspicuous way. If we try to cash IBE within a Bayesian framework by focusing on likelihoods (that is, by saying that the best explanation is the hypothesis with the highest likelihood), then intuitive judgements of best explanation and judgements of Bayesian confirmation may well come apart.

Surely more needs to be said at this stage and I cannot say it now. Let me just distinguish between two issues. One is: can we equate the best explanation with the hypothesis that has the highest likelihood? I have just shown that we cannot. The other issue is: can we accept a hypothesis as the best explanation of the evidence if its posterior probability is low? This is a tough question. But I do not want to give a straightforward negative answer to it. Of course, it's unlikely that Joan has the disease given that she tested positive, if we know that the base-rate of the disease in the population is very low. But unlikely things happen and we don't want to say that it's outright unreasonable to believe that an unlikely thing has happened (especially if this best explains the evidence).[4] In any case, we are not always (or most typically) in situations where we have definite probabilities available. Nor can reasonable belief be equated with highly probable belief. There is more (and perhaps less) to reasonable belief than high probability.[5]

The point about likelihoods I have just made generalises. Consider what is called the Bayes factor, i.e., the ratio of likelihoods prob(e/not-H)/prob(e/H). One might try to connect IBE with likelihoods as follows. If the Bayes factor is small, then H is a better explanation of the evidence *e* than not-H. For, the thought will be, there are two ways in which the Bayes factor can be minimised: either when *e* is very unlikely when H is false or when *e* is very likely when H is true. Now, we can see that a version of Bayes's theorem is this:

$$\text{prob(H/e)} = \text{prob(H)}/\ \text{prob(H)} + f\,\text{prob(not-H)},$$

where f is the Bayes factor, i.e., prob(e/not-H)/prob(e/H). Wouldn't we expect that the smaller the Bayes factor is, the greater is the posterior probability of the

hypothesis? Wouldn't we thereby find a way to accommodate IBE, via the Bayes factor, within Bayesianism? Well, as above, what really happens depends on the prior probability. The Bayes factor, on its own, tells almost nothing. I say 'almost nothing' because there is a case in which the prior probability of a hypothesis does not matter. This is when the Bayes factor is *zero*. Then, no matter what the prior prob(H) is, the posterior probability prob(H/e) is one. So, it's only when just one theory can explain the evidence (in the sense that the likelihood prob(e/not-H) is zero) that we can dispense with the priors. That's a significant result. But does it show that IBE is accommodated within Bayesianism? In a sense, it does. But this sense is not terribly exciting. If there was only one potential explanation, then it would be folly not to accept it. But this case is really exceptional. We are still left with the need to distinguish between grue and green!

The moral so far is double. On the one hand, likelihoods cannot capture the notion of a good (the best) explanation. Put in a different way, even if likelihoods could, to some extent, carry the weight of explanation, they couldn't carry all of this weight on their own. On the other hand, we need to take into account the prior probabilities before we draw safe conclusions about the degree of confirmation of a hypothesis.

4. EXPLANATION AND PRIOR PROBABILITIES

What then remains of the Bayesian kinematics as an (indispensable) entry point for explanatory judgements is the prior probabilities. Now, it is one thing to say that priors are informed by explanatory considerations and quite another thing to say that they *should* be so informed. No-one would doubt the former, but subjective Bayesianism is bound to deny the latter. So, we come to the crux. There are two ways to think of IBE within a Bayesian framework. The *first* pays only lip service to explanatory considerations. For all the work in degree-of-belief updating (or, as some Bayesians say, in maintaining internal coherence in an agent's belief-corpus) is done by the usual Bayesian techniques and, perhaps, by the much-adored appeal to the washing out of priors. It may be admitted that the original assignment of prior probabilities might be influenced by explanatory considerations but the latter are no less idiosyncratic (from the point of view of the subjective Bayesian) than specifying the priors by, say, consulting a soothsayer. If we think this way, IBE, in a loose sense, is rendered consistent with Bayesianism, but it loses much of its excitement. It just amounts to a *permission* to use explanatory considerations in fixing an initial distribution of prior probabilities.

The other way to think of IBE within a Bayesian framework is to take explanatory considerations to be a *normative* constraint on the specification of priors. This is the way I would favour, if I were to endorse the Bayesianisation of IBE. It would be an exciting way to bayesianise IBE. (Better put, it would be an

exciting way to explanationise Bayesianism – forgive me the bad English words.) For it would capture the idea that explanatory considerations should be a rational constraint on inference. We might still be short of acceptance, since all we end up with is a degree of belief (no matter how high), but it would, at least, be a degree of *rational* belief. This move would also show how the resolution of observational ties is not an idiosyncratic matter. For some theories would command a higher initial rational degree of belief than others and this would be reflected, via Bayesian kinematics, in their posterior probability.[6]

But don't we all know that the story I have just outlined is, to say the least, *extremely* contentious? It would call for an objectivisation of Bayesianism and this is something that we, presumably, know it cannot be done. Whence do the explanatory virtues get their supposed rational force? And how are they connected with truth? I think these are serious worries. I am not sure they are compelling. For instance, I think there can be an a posteriori argument to the effect that theories with the explanatory virtues are more likely to be true than others (cf. my 1999, 171-6). And there is also an argument to the effect that judgements of prior probabilities should aim to improve the coherence of our system of beliefs and that the explanatory virtues improve such coherence (cf. my 2002). But showing all this would be an uphill battle. It would call, to say the least, for a radical departure from the standard Bayesian criteria of rationality and belief-revision. So, the project of accommodating IBE within Bayesianism would involve a radical rethinking of Bayesianism. And not many people are, nowadays, willing for such a radical rethinking.

5. A DILEMMA

The way I have described things leads us to a dilemma. *Either* accommodate (relatively easily) IBE within Bayesianism but lose the excitement and most of the putative force of IBE *or* accommodate an interesting version of IBE but radically modify Bayesianism. I guess we all agree that Bayesianism is the best theory of confirmation available. But at least some of us are unwilling to think that Bayesianism is the final word on the matter, since we think that there is more to rationality (and to scientific method) than Bayesians allow. Those of us who are friends of IBE might then have to reject the foregoing dilemma altogether. This would bring us back to the second question I raised in section 1, and which I took to be the more interesting one: *should we want to cast IBE within a Bayesian framework?*

I cannot start answering this question in this paper. I hope to have sketched why there are reasons to take it seriously. I will conclude with a note on how a *negative* answer to it can be motivated. IBE is supposed to be an *ampliative* method of reasoning. It is supposed to deliver informative hypotheses and theories, viz., hypotheses and theories which exceed in content the observational data, experimental results etc. which prompt them. This content-increasing and

pect of IBE is indispensable, if science is seen, at least *prima facie*, as an activity that purports to extend our knowledge (and our understanding) beyond what is observed by means of the senses. Now, Bayesian reasoning is *not* ampliative. In fact, it does not have the resources to be ampliative. All is concerned with is maintaining synchronic consistency in a belief corpus and (for some Bayesians, at least) achieving diachronic consistency too. Some Bayesians, e.g., Colin Howson (2000), take probabilistic reasoning to be a mere extension of deductive reasoning, which does not beget any new factual content.

There might be two related objections to what I have just said. The first might be that Bayesian reasoning allows for ampliation, since this can be expressed in the choice of hypotheses over which prior probabilities are distributed. In other words, one can assign prior probabilities to ampliative hypotheses and then use Bayesian kinematics to specify their posterior probabilities. The second (related) objection may be found in what Niiniluoto says at some point, viz., that "the Bayesian model of inference helps to show how evidence may confirm hypotheses that are abductively introduced to explain them" (2003, 20). Here again, the idea is that abduction might suggest ampliative hypotheses, which are then confirmed in a Bayesian fashion. If we elaborate and combine the two objections in an obvious way, they imply the following: ampliative IBE and non-ampliative Bayesian reasoning might well work in tandem to specify the degree of confirmation of ampliative hypotheses.[7]

I too have toyed with this idea and still think that there is something to it. In an earlier piece I noted:

> although a hypothesis might be reasonably accepted as the most plausible hypothesis based on explanatory considerations (abduction), the *degree of confidence* in this hypothesis is tied to its degree of subsequent confirmation. The latter has an antecedent input, i.e., it depends on how good the hypothesis is (i.e., how thorough the search for other potential explanations was, how plausible a potential explanation is the one at hand etc.), but it also crucially depends on how well-confirmed the hypothesis becomes in light of further evidence. So, abduction can return likely hypotheses, but only insofar as it is seen as an integral part of the method of inquiry, whereby hypotheses are further evaluated and tested (2000, 67).

But we can also see the limitations of the idea under discussion. For what, in effect, is being conceded is that IBE (or abduction) operates *only* in the context of discovery, as a means to generate plausible ampliative hypotheses and to distil the best among them. Then, the best explanation is taken over to the context of justification, by being embedded in a framework of Bayesian confirmation, which determines its credibility. I think the friends of IBE have taken IBE to be both ampliative *and* warrant-conferring at the same time. It is supposed to be an ampliative method that confers warrant to the best explanation of the evidence. So, if we are concerned with giving a precise degree of confirmation to the best explanation, or if we subject it to further testing, then we can indeed embed it in a Bayesian framework. But something would have gone amiss if we thought that

the best explanation was *not* reasonably acceptable before it was subjected to Bayesian confirmation. To see how this reasonable acceptance could be analysed would lead us beyond the confines of this commentary. But I have started working out the details elsewhere (cf. 2002). For, I think we can profitably cast the issue of the warrant of IBE within the theories of justification which connect justification with the absence of defeaters, theories which were made popular by John Pollock (1986).

NOTES

* Many thanks to Ilkka Niiniluoto, Peter Lipton, Maria-Carla Galavotti and the participants of the Workshop "Induction and Deduction in the Sciences" for many useful comments on an earlier draft.

1. Niiniluoto (personal communication) has rightly pointed out that there are two big strands within Bayesianism. One of them (Levi, Hintikka) promotes the idea of inductive acceptance rules, and hence, advocates ampliative inferences. The other branch (Carnap, Jeffrey, Howson) rejects acceptance rules and considers only changes of probabilities. It is this latter strand of Bayesianism that I take issue with at this point.
2. It is very debateable that we should equate the base-rate with the prior probability. But nothing hangs on this in the use I make of the base-rate fallacy.
3. There is a lot of recent re-evaluation of the base-rate fallacy (cf. Koehler 1996). One point that seems worth making, though I am not sure I want to endorse it in full, is that the base-rate fallacy relates specifically to probabilistic reasoning, where reference classes are to be taken into account. We may or may not be good at seeing the need to take into account reference classes in order to draw conclusions about probabilities concerning individual cases. But it's not clear to me why we *should* take into account reference classes when we look for best explanations of the evidence. An explanation can be reasonable to accept (even true), even though it is unlikely. In a sense, what matters for the explanation of an individual event is not what the other members of the reference class we put it in do, but rather what the details of the individual case we are interested in are.
4. Consider the following. A Geiger-counter detects a certain type of particle by registering a click. The particles are very rare so that the probability that the counter clicks is very low. But suppose it does click. Is it not unreasonable to believe that it registered a particle, especially if it's highly reliable in doing this?
5. That there is more to reasonable belief than high probability is argued at length by Achinstein (2001, chapter 7). He takes high probability to be necessary for rational belief but he denies that it is sufficient. One of his explicit additional requirements is that there must be an explanatory connection between a hypothesis and its evidence.
6. There is an interesting idea in Niiniluoto's paper (2003, 21) that needs to be noted. One may call "systematic power" the explanatory and predictive power of a hypothesis relative to the total *initial* evidence *e*. Then, one may use this systematic power to determine the probability prob(H) of the hypothesis H. Prob(H) will be none other than the *posterior probability of H* relative to its ability to explain and predict the initial evidence *e*, that is relative to its systematic power. In this sense, it can be argued that the prior probability of a hypothesis does depend on its explanatory (that is, its systematic) power.
7. This line is pressed in a fresh and interesting way in Lipton's (2001). Lipton tries to show how a Bayesian and an Explanationist can be friends. In particular, he shows how explanatory considerations can help in the determination of the likelihood of a hypothesis, of its prior probability and of the relevant evidence. I think all this is fine. But it should be seen not as an attempt at peaceful co-existence, but rather as an attempt to render Bayesianism within the

Explanationist framework, and hence as an attempt to make Bayesianism an objectivist theory of confirmation.

REFERENCES

Achinstein Peter, *The Book of Evidence*, New York: Oxford University Press 2001.
Howson Colin, *Hume's Problem*, New York: Oxford University Press 2000.
Koehler Jonathan J., "The Base Rate Fallacy Reconsidered: Descriptive, Normative and Methodological Challenges", *Behavioural and Brain Sciences* 19, 1996, pp. 1-53
Lipton Peter, "Is Explanation a Guide to Inference", in: G. Hon & S. S. Rakover (Eds.), *Explanation: Theoretical Approaches and Applications*, Dordrecht: Kluwer 2001, pp. 93-120.
Niiniluoto Ilkka, "Truth-Seeking by Abduction" *this volume*.
Pollock John, *Contemporary Theories of Knowledge*, Rowan & Littlefield 1986.
Psillos Stathis, *Scientific Realism: How Science Tracks Truth*, London: Routledge 1999.
Psillos Stathis, "Abduction: Between Conceptual Richness and Computational Complexity" in: A. K. Kakas and P. Flach (Eds.), *Abduction and Induction: Essays in their Relation and Integration*, (Applied Logic Series, Vol. 18), Dordrecht: Kluwer 2000, pp. 59-74.
Psillos Stathis, "Simply the Best: A Case for Abduction" in: Fariba Sadri & Anthony Kakas (Eds.), *Computational Logic: From Logic Programming into the Future*, LNAI 2408, Berlin-Heidelberg: Springer-Verlag 2002, pp. 605-25.
van Fraassen Bas, *Laws and Symmetry*, Oxford: Clarendon Press 1989.

Department of Philosophy and History of Science
University of Athens
Panepistimioupolis (University Campus)
Athens 15771
Greece
psillos@phs.uoa.gr