Bayesianism and inference to the best explanation*

Valeriano IRANZO

Received: 2006.04.10
Final Version: 2008.02.23

BIBLID [0495-4548 (2008) 23: 61; pp. 89-106]

ABSTRACT: Bayesianism and Inference to the best explanation (IBE) are two different models of inference. Recently there has been some debate about the possibility of "bayesianizing" IBE. Firstly I explore several alternatives to include explanatory considerations in Bayes's Theorem. Then I distinguish two different interpretations of prior probabilities: "IBE-Bayesianism" (IBE-Bay) and "frequentist-Bayesianism" (Freq-Bay). After detailing the content of the latter, I propose a rule for assessing the priors. I also argue that Freq-Bay: (i) endorses a role for explanatory value in the assessment of scientific hypotheses; (ii) avoids a purely subjectivist reading of prior probabilities; and (iii) fits better than IBE-Bayesianism with two basic facts about science, i.e., the prominent role played by empirical testing and the existence of many scientific theories in the past that failed to fulfil their promises and were subsequently abandoned.

Keywords: Bayesian epistemology, inference to the best explanation, confirmation, frequentism, prior probability, explanatory value.

Scientists are very often confronted with alternative accounts of experimental data. Decisions concerning which one should be preferred usually involve several values. Quantitative confirmation theories focus on probability. In particular, they conceive the confirmation of a hypothesis as an increase of its probability due to the evidence. On the other side, when comparing rival hypotheses, scientists use to consider their respective explanatory merits. Sometimes they prefer hypothesis h1 instead of hypothesis h2 because h1 is better qua explanation than h2, even though h1 does not enjoy any significant predictive success yet. Eventually, h1 should exhibit its predictive power, but this further condition, sometimes considered as a necessary one for acceptance of hypothesis, should not depreciate the important role played by explanatory value in scientists' assessments of hypotheses. A good example is Einstein's General Relativity. Before enjoying a substantial predictive record of success, it was widely considered a valuable theory which deserved careful scrutiny, even though this assessment was largely based on explanatory merits like unification and simplicity. Hence, the superiority of h1 over h2 concerning explanatory value can be a good reason —for scientists, not only for some philosophers of science— to investigate the consequences of the former in detail.

In this paper I will address one general worry raised by the foregoing remarks, namely, the supposed connection between explanatory merit, on one side, and confirmation and probability on the other. I will restrict myself to Bayesianism and "inference to the best explanation" (IBE, hereafter). Recently there is some debate about

* This work has been funded by the Spanish Ministry of Science and Education (Ramón y Cajal Programme) and also by the Generalitat Valenciana (research project GV 06/22).

the possibility of integrating them (van Fraassen 1983, Douven 1999, Hon and Rakover 2001, Niiniluoto 2004, Psillos 2004). IBE is a model of inference that stresses the role played by explanatory considerations in non-deductive inference in general. It does not seem out of place, then, to allude to IBE in the scientific context, given that such sort of considerations effectively guide scientific judgment. Besides, the formal apparatus provided by probability theory could perhaps make IBE more precise. In fact, vagueness has been one of the criticisms levelled at it.

1. "Bayesianizing" inference to the best explanation

Peter Lipton, one of the most prominent advocates of IBE, alludes to such properties as scope, simplicity, unification, mechanism, and precision in order to distinguish good explanations from bad ones (Lipton 2004). We will focus, however, on the general property of being a good explanation and I will use the expressions 'explanatory merit' or 'explanatory value', instead of Lipton's favourite expression 'loveliness', to refer to that general property.

Lipton maintains that explanatory merit enjoys a double role in inductive inference. The first one is *psychological*:

> According to IBE, our inferential practices are governed by explanatory considerations. Given our data and our background beliefs, we infer what would, if true, provide the best of the competing explanations we can generate of those data (so long as the best is good enough for us to make any inference at all). (Lipton 2004, p. 56)

In our everyday life we have a tendency to infer the hypothesis that is the best, *qua* explanation, from the pool of available hypotheses. In scientific research, in its turn, experimental tests and procedures of gathering empirical information are devised in order to decide which hypothesis should be accepted. It could be said that, in principle, scientists are interested in selecting the true —or the most probable— of those hypotheses (let us put aside for the moment other factors which could be involved). We may grant that all of them enjoy some previous plausibility; otherwise, any efforts to test them would be pointless. But the issue for IBE is not simply that when deciding which hypotheses deserve to be investigated in detail scientists are guided by considerations about their respective explanatory successes. The fact is, rather, that scientists' assessments on the plausibility of those competing hypotheses are definitely based on judgments about how good they are as explanations. Shortly, as a consequence of their explanatory import, hypotheses *are considered* more or less probable.

Nonetheless, even though Lipton focuses on the descriptive side of IBE, explanatory merit also plays an *epistemic*, normative, role for him:

> We may characterize the best explanation as the one which would, if correct, be the most explanatory or provide the most understanding: the 'loveliest' explanation. [...] The version of inference to the best explanation we should consider is Inference to the Loveliest Potential Explanation. Here at least we have an attempt *to account for epistemic value in terms of explanatory virtue.* This version claims that the explanation that would, if true, provide the deepest understanding *is the explanation that is likeliest to be true.* (Lipton 2004, pp. 59, 61; my emphasis)

"Explanatory loveliness" is an epistemic value insofar as it increases probability. Certainly, our inductive inferential practice is fallible, and there is no perfect match between likeliness and explanatory value. Sometimes the likeliest explanation is not very enlightening, and an explanation can also be appealing without being likely. But, if IBE-theorists are right, other things being equal, good explanations are more probable than bad ones: "After all, IBE is supposed to describe strong inductive arguments, and a strong inductive argument is one where the premises make the conclusion likely" (Lipton 2004, p. 60). As a result, IBE is not compatible with any account of the theory-choice problem which allocates explanatory value within the realm of non-epistemic —purely pragmatic— virtues. After this brief sketch of IBE's essential claims, let us turn to Bayes's Theorem.

Bayes's Theorem is a deductive consequence of the axioms of probability. According to an epistemological reading of it, this theorem provides an algorithm to calculate the degree of support that a particular bit of evidence *e* confers to a hypothesis *h*. Here is the simplest epistemological version of this theorem (see Gillies 2000, chaps. 3-4, and Joyce 2003 for details):

$$p(h/e) = \frac{p(e/h) \cdot p(h)}{p(e)} = \frac{p(e/h) \cdot p(h)}{p(e/h) \cdot p(h) + p(e/\neg h) \cdot p(\neg h)}$$

$p(h/e)$ is the conditional probability of h, given that e occurs.

$p(h)$ is the *prior* probability of h.

$p(e/h)$ is the *likelihood* of e on h.

$p(e)$ is the *expectedness* of e.

In situations where there are some rival incompatible explanations h_i, h_j, h_k, ... , h_n, the appropriate formula is:

$$p(h_i/e) = \frac{p(e/h_i) \cdot p(h_i)}{\sum_{i=1}^{n} p(e/h_i) \cdot p(h_i)}$$

It is worth emphasizing that "prior probability" does not necessarily mean "a priori probability." Prior probability is the probability of *h before e* is obtained and that's all. The theorem is silent on the way *h* gets its prior probability. So it does not rule out the possibility that *h*'s prior could be based on some evidence previous to *e*.[1] Now, a particular bit of evidence e_1 confirms *h iff* when e_1 occurs the probability of *h* is increased, that is, *iff* $p(h/e_1) > p(h)$. In other words, e_1 confirms *h iff* the posterior, the conditionalized probability, is higher than the prior probability. When we get further evidence, e_2, we should take $p(h/e_1)$ as the prior probability; the likelihood now would be $p(e_2/h)$ and the corresponding expectedness is $p(e_2)$. Then we would calculate $p(h/e_2)$. This sort

[1] That is the reason for including a special term to refer to the background knowledge:

$$p(h/e \wedge b) = \frac{p(h/b) \cdot p(e/h \wedge b)}{p(e \wedge b)}$$

The term *b* draws our attention to the fact that probability assessments are always relative to our previous knowledge. Granted the point, the version in the main text suffices here.

of iterated conditionalization is the Bayesian way of learning from experience, since probability assignments are modified as empirical evidence increases.[2]

In regard to the scientific context, Lipton asserts: "The notion of explanatory loveliness should help to make sense of the common observation of scientists that broadly aesthetic considerations of theoretical elegance, simplicity, and unification are a guide to inference" (Lipton 1991, p. 68). Now the question is how explanatory considerations could be included in Bayes's Theorem. It should be remarked that this is not an issue of concern only for IBE theorists. At the outset of this paper I alluded to the role played by explanatory merit in theory choice. So, even Bayesians who do not feel sympathetic at all with IBE could have a reason to include it in the Bayesian formula.[3]

Anyway, if explanation has any epistemic import, as IBE-theorists claim, it must have an effect on probabilities. Let us think of two hypotheses (h_1, h_2) which greatly differ in their respective explanatory merits: while h_1 is a good explanation, h_2 is notably *ad-hoc*. How to include explanatory merit in the Bayesian algorithm so that the final probabilities attached to h_1 and h_2 reflect their different explanatory value? There are several alternatives:

(1) *Increasing/decreasing conditional probabilities*. Firstly we calculate the conditional probabilities on e. Then we give an extra weight (Δ) to explanatory value. So,

$$p_f(h_1) = p(h_1/e_1) + \Delta \; p_f(h_2) = p(h_2/e_1) - \Delta$$

where p_f is the probability of the hypotheses given both the available confirmatory evidence e_1 *and* their explanatory quality.

It should be noticed here that although explanatory merit is taken into account, properly speaking, it is not included *in* the Bayesian formula. Recall that if hypotheses had further untested observational consequences, new empirical evidence could change the value for p_f by means of the iterated application of Bayes's Theorem for $p(h/e_1)$, $p(h/e_1 \wedge e_2)$, ... But, according to (1), after the required calculations are done, the prize for explanation is *extrinsically* added. It is no surprise that such sort of modification on conditional probabilities is exposed to a dynamic Dutch-Book argument (van Fraassen 1989, chapter 7). A dynamic Dutch-Book argument purports to show that a bookie can make a series of bets, offered at distinct times, against you which guarantee that you will lose some money whatever happens. The conclusion is that an

[2] A clear advantage of Bayesianism over the hypothetic-deductive model is that the latter is a particular case of Bayes's theorem. Thus, if h deductively entails e, then $p(e/h) = 1$. In that case $p(h/e) = p(h) / p(e)$. Since h entails e, $p(h) \leq p(e)$. Then, successful deductive explanation of e by h guarantees that e confirms h, that is, $p(h/e) > p(h)$ for a wide array of situations where $p(h) > 0$ and $p(e) < 1$. Concerning refutation, if $p(e/h) = 1$, $p(\neg e/h) = 0$. But then, if $\neg e$ occurs, $p(h/\neg e) = 0$. So, when $\neg e$ occurs, h has been conclusively refuted. A further complication is that scientists usually invoke auxiliary assumptions in order to neutralize counterevidence (see Salmon 1996).

[3] Decision theory is a further alternative for Bayesians who intend to leave room for explanatory value. Probability and explanatory merit could be considered as different utilities, and scientists' preferences should be rationalized in terms of maximizing expected value (see, for instance, Maher 1993). However, IBE's advocates would reject this point of view because if probability and explanatoriness are set apart, the latter is entirely deprived of confirmational value.

updating rule like the "delta policy" is incoherent because it violates the axioms of probability. Since some advocates of IBE have accepted this conclusion, I will not pursue option (1). In fact, there are less controversial alternatives for them, as we will see in the following paragraphs.[4]

(2) *Increasing/decreasing prior probabilities*. The prior probability is the Bayesian expression for the initial plausibility of a hypothesis. Taking for granted a definition of explanatory value that includes theoretical elegance, simplicity..., it follows that the initial plausibility of hypotheses has to do with their explanatory merit. Thus, good explanations should enjoy higher prior probabilities than bad ones. Then we would calculate the subsequent effect of evidence and we would obtain a value for $p(h/e)$.

This option avoids the unfortunate consequences of the "delta policy" —the Dutch-Book argument— and preserves the basic claims of IBE. The psychological role of explanation in inference seems vindicated. Attributions of prior probability —i.e., plausibility— are guided by explanatory value, and conditionalization on evidence e_1 takes place, then, after explanatory criteria are applied to set the value of $p(h)$.

Concerning the epistemic role of explanatory merit, Bayes's recipe asserts that the conditional probability, $p(h/e)$, depends both on the likelihood (the probability of the evidence given the hypothesis) and on the initial plausibility enjoyed by the hypothesis. Hence, high (low) prior probabilities do not ensure high (low) posterior probabilities. Turning to our example, h_2 —a bad explanation— eventually gets higher posterior probability than h_1. Then, if h_2 surpasses h_1 as predictor of the same body of evidence e to a great extent, that is, if $p(e/h_2) >> p(e/h_1)$, the conditional probability of h_2 on e could be higher than that of h_1. It is true, then, that explanatory value does not directly increase the conditional or posterior probability, but it may do indirectly. In fact, when the likelihood is the same for all competing hypotheses —suppose, for instance, that all them entail the evidence at issue—, priors are crucial for posterior probabilities. In that case, the factor $[p\ (e/h_j)/\sum p(e/h_i)\cdot p(h_i)]$ is the same for all h_i, that is: $[1/\sum p(e/h_i)\cdot p(h_i)]$ (see footnote 2). The factor is indeed equal to 1, when in addition to the deductive relation between rival hypotheses and the evidence, the set of hypotheses form a partition of the sample space. Consequently, the higher the prior, the higher the posterior probability. Therefore, in situations like these, prior probabilities decide the question: the best explanation, *ceteris paribus*, would also be the most probable hypothesis in respect of the available evidence e.

(3) *Increasing/decreasing likelihoods*. A hypothesis may be simple and elegant but provide a poor explanation of some particular piece of evidence. An example of this would be

[4] Incidentally, not all Bayesians accept that the Dutch-Book argument is sound (see Maher 1993). Concerning van Fraassen's particular version, Christensen 1991 and Douven 1999 reject it, while Day and Kincaid 1994 and Okasha 2000 endorse its conclusion, although they do not think that IBE is incompatible with Bayes's Theorem. In fact, Day and Kincaid suggest alternative 2 —see below—, and Okasha favours alternatives 2 and 3. Neither of them, on my view, addresses the crucial question: given that there is no inconsistency in including explanatory considerations in Bayes's Theorem, we may include them, but why should we do it? Why should probability and confirmation be affected by explanatory value?

Newton's law of gravity and the explanation it provided of the motion of the Moon.[5] However, a general feature of good explanations seems to be that they make the *explanandum* a natural consequence of the *explanans*. Since likelihoods express the probability of the evidence in case the hypothesis is true, it seems that explanatory merit and likelihood are somehow related.

Likelihoods are relatively easy to calculate when there are accumulated frequency data. Thus, it may be known that the symptom e is present in 60 per cent of patients in which a disease is present. It is much more complicated to calculate the likelihood of gravitational lenses, for instance, given that the General Theory of Relativity is true, even though it is precisely in such cases where explanatory considerations could be significantly relevant, according to IBE theorists. Rival hypotheses to h_1 (h_2, h_3, ..., h_n) may seem themselves very unlikely. The reason may well be that they have a lower prior probability than h_1. But a further aspect involved could also be that they do not suggest a convincing account of the underlying mechanism or process that relate the *explanans* to the particular *explanandum*. Perhaps it is very sketchy, or it is difficult to fit it into the background knowledge, while h_1 is, in contrast, very good as an explanation *of* e_1. And the values of their respective likelihoods $p(e/h_1)$, $p(e/h_2)$, ..., $p(e/h_n)$ should be affected by this difference.

But we should not exaggerate the similarities between explanatory merit and likelihood. Firstly, $p(e/h)$ may be high even though h does not explain e —think about standard examples against the deductive-nomological model of explanation. Besides, according to Bayes's formula, confirmation is a matter of mutual reinforcement: if e confirms h, then h confirms e. In contrast, very often h explains e, but e does not explain h. Causal explanations are paradigmatic examples here. Notwithstanding, the foregoing remarks show that explanatory value is not entirely encapsulated in $p(h)$. Explanatory value is not an intrinsic merit of h because it partly has to do with the probability of the *explanandum* in relation to the *explanans*. Explanatory value is, indeed, an umbrella term that covers a multitude of different properties. Some of them have to do with the initial plausibility of hypotheses —the qualities of the hypothesis *per se*, such as its simplicity or elegance. Some other are concerned with the link between the *explanans* and the *explanandum*, that is, with the appropriateness of the hypothesis *qua* explanation of the evidence at issue. Then, if the quality of an explanation is related both to the prior probabilities and to the likelihoods —and good explanations are intrinsically as well as "relationally" appealing—, those interested in including explanatory merit in the Bayesian algorithm should consider alternatives (2) and (3). My aim in this paper is

[5] The former is simple and elegant, but the explanation it gives of the motion of the Moon (in Newton's hands) was notoriously poor unlike the explanation of the motion of the planets. Thus one might attempt, as some did, to improve the fit between theory and evidence in the Moon case by adding terms to Newton's law. This would make the hypothesis less elegant, but we would surely say that there are more facts —or empirical observations— explained by it. Certainly, it is debatable whether these modifications effectively increase the global amount of explanatory merit of the hypothesis. Anyway, the point is that the explanatory merit of a hypothesis has to do both with intrinsic features of it and with its relation to particular *explananda*.

more modest. Leaving aside the problem of translating likelihoods to explanation-related terms, in the sequel I will focus on prior probabilities.[6]

2. Prior probabilities

What are prior probabilities for Bayesians? So-called "subjective" Bayesianism (with B. De Finetti and F.P. Ramsey as the main historical figures) defends that the only epistemic constraints on belief are the axioms of probability and the iterated application of Bayes's rule (conditionalization) in light of new evidence (see above, p. 93). It must be said that subjectivism has been the orthodox position in the Bayesian camp for a long time (see Howson and Urbach 2006 for the standard account among philosophers of science). Concerning the priors, subjective Bayesians think that their values are constrained just by probabilistic coherence and may radically disagree for different agents. At this point subjective Bayesians use to invoke several "Convergence Theorems" in order to minimize the consequences of the initial disagreement. These theorems intend to prove that conditionalizing on evidence "washes out" the disparity about priors because, as the amount of evidence increases, posterior probabilities tend to a limit.

A common criticism against subjective Bayesianism is that Convergence Theorems do not work when applied in actual scientific practice because the process of gathering evidence is severely constrained and cannot proceed indefinitely.[7] Anyway, I do not think that those interested in combining IBE with Bayesianism are bound to endorse an overtly subjectivist standpoint on the priors. Consequently, they need neither full confidence on any theorems which set up conditions hardly fulfilled in "real science". Again, explanatory goodness is at root of scientific inference. Scientists prefer hypotheses that seem likely insofar as they have explanatory value —recall that this is the psychological import of IBE. We have noted that prior probabilities are intimately related to appraisals of virtues like theoretical elegance, simplicity and unification. Now it should be added that radical subjectivism about the priors hardly fits with the existence of a shared methodological core for the scientific community. Perfect agreement about prior probabilities is not a plausible scenario, not even in science. But current science is a cooperative project and some degree of coincidence is necessary. Although discrepancies about the priors are allowed concerning specific and controversial examples, there is a general agreement about which factors must be taken into account in the assessment of priors. Translated into IBE's terminology, the "perception" of explanatory goodness is rather homogeneous. The agreement is just a by-product of a highly institutionalized training process that reinforces some cognitive heuristics (on the cognitive mechanism involved here see Kuipers 2002). To acknowledge the

[6] Stathis Psillos, a partisan of IBE, has argued that "the base-rate fallacy shows that it is incorrect to equate the best explanation of the evidence with the hypothesis that has the highest likelihood" (Psillos 2004, p. 86). While I endorse Psillos's conclusion, I am not so sure that the base rate fallacy is the decisive rationale for it. I cannot pause to argue for that here.

[7] A critical discussion of Convergence Theorems can be found in Earman 1993, chapter 6. For a recent vindication of objective priors see Huber 2005.

collective dimension of science forces us to accept that social or pragmatic constraints operate on the priors. Consequently, although a purely subjectivist standpoint on the priors perhaps could be defended in some particular contexts, it does not seem appropriate to account for scientific judgment in general.

IBE-Bayesians, like Lipton, can accommodate this fact by arguing that priors are implicit assessments of explanatory value. Discrepancies about the priors should be imputed to differences in educational contexts, to scientists' personalities... Since, as a matter of fact, in scientific sub-communities there is a general agreement concerning the priors, it could be defended that, as a result of scientific education, discrepancies among scientists on the priors are not remarkable, except in particular circumstances. Nonetheless, this is not sufficient to sustain that priors —i.e., appraisals of explanatory merit— should be considered as a reliable guide to posterior probability. Recall that, according to the alleged epistemic role of IBE, as the initial estimates are subsequently modified by the impact of evidence, those hypotheses initially considered as good explanations by scientists are more probable than bad ones. Scientists' agreement on the priors discards radical subjectivism, certainly, but it does not make priors objective in the sense required by a genuine epistemic role. Consensus may well be a necessary condition for research in current science. But if it were a brute sociological fact, the tendency to focus on hypotheses that score better than its rivals concerning prior probability —that is, the tendency to prefer some particular features that make hypothesis good qua explanations— would be on the same footing as the tendency to wear white coats when working at laboratories, for instance.

A natural suggestion at this point is that scientists' judgments about the initial plausibility of hypotheses convey the scientific lore about those kinds of hypotheses which have been successful in the past. In other words, prior probabilities are "our best estimates of the frequencies with which certain kind of hypotheses succeed" (Salmon 1996, p. 270). I will label this point of view as Frequentist-Bayesianism as an alternative to IBE-Bayesianism (Freq-Bay and IBE-Bay, shortly).

It is important to realize that Salmon's proposal rejects an essential claim of IBE. A Frequentist-Bayesian may accept the aforementioned explanation on the existence of some degree of consensus about the priors. But now it is not explanatory virtue by itself that increases the prior probability, since more or less intuitive judgments about it are grounded on frequencies of success. In addition to this, she can add that disagreements about priors may be right or wrong in a sense that goes beyond group conformity. An attribution of prior probability is wrong if it does not reflect the rate of past successes. The relevant technical term in Bayesian jargon is *calibration*. An agent is perfectly calibrated if there is a perfect match between her probability attribution of property L to individual a and the actual frequency of L in the appropriate class of reference to which a belongs (for a formal account, see van Fraassen 1983).

Someone could point out here that, even though a perfectly calibrated agent could know exactly the frequency of success in an idealized situation, scientists are not perfectly calibrated agents. Moreover, attributing an initial plausibility to a hypothesis is not commonly explicated in probabilistic terms.

Salmon somehow anticipated such objections, but his remarks were really brief. He acknowledged that frequentism on the priors "does not imply that all scientists will agree on the numerical values or prefer the same theory" (Salmon 1996, p. 281) and also claimed that "the evaluation of prior probabilities clearly demands the kind of scientific judgment whose importance Kuhn has rightly insisted upon" (*id.*). I interpret these quotations as follows. The success rate obtained in the past is an objective fact, but it may be really complicated to ascertain it. Anyway, even though in actual situations of theory choice "kuhnian qualitative judgments" may be inevitable, assessments of initial plausibility have a normative import insofar as they are disguised judgments about frequencies. Besides, in order to attribute an initial probability to a new hypothesis it must be established what are the relevant respects of similarity between it and the old ones. From this perspective, prior probability is just an encoded measure of how much similar is the new hypothesis to the previous ones which were successful —presumably the similarity involved here is formal as well as in content. Prior probabilities, in sum, are right or wrong insofar as they approximate the relevant frequencies of past success. Although the information required to assess those frequencies is not accessible to us by now, the correctness of priors depends on what happens in the world.

The rationale for the frequentist interpretation is that after several centuries of doing science we have learned the distinctive features of good scientific hypotheses. Simplicity, compatibility with other accepted theories, unifying power, are now preferred because hypotheses that possessed them in the past yielded better results than hypotheses that did not enjoyed them. It should be added that these methodological criteria evolve as the rest of our scientific knowledge about the world. But, even though they are revisable, they seem a very good, the best indeed, point of departure.

Now we have two interpretations of Bayes's formula: IBE-Bay and Freq-Bay. Both include explanatory merit through prior probabilities, but they disagree on the interpretation of the priors. Which one should be preferred? Before making the comparison, the frequentist interpretation deserves a closer look.

3. A frequentist interpretation of the priors

Salmon claimed that the frequentist interpretation of the priors is itself an empirical hypothesis. In the early seventies he pointed out that there were no reliable statistics on such matters, but he affirmed that "it is enough to have very very rough estimates" (Salmon 1970, pp. 85-86). Thirty years later he still maintained these opinions, but as far as I know, he never went deeper into details (see his papers included in Hon and Rakover 2001). However, if our estimates are so rough and there is no reliable evidence to test them, we are at risk of turning Freq-Bay into a mere hint with no empirical advantage over an explanationist approach to the priors. Therefore, it is important to be more precise about *how the frequentist hypothesis could be tested and how it could be subsequently used to specify the priors of novel hypotheses.*

Think again of h_1 and h_2 —the good and the "ugly" hypotheses, respectively. Let us assume that both are novel hypotheses and that we are concerned with how to set their priors. Suppose now that L contains all those hypotheses previously proposed in the history of science that can be considered good ones from an explanatory point of view (the *L*-hypotheses, for "lovely" hypotheses). Stated in very general terms, Freq-Bay's line of reasoning is:

Most of *L*-hypotheses have been successful.

h_1 is similar to *L*-hypotheses and h_2 is not similar to *L*-hypotheses.

Therefore, h_1 merits higher prior probability than h_2.

Admittedly, the frequentist interpretation of the priors is a second order hypothesis —a meta-hypothesis— about past scientific hypotheses and testing it is not an easy task.

First of all, success is an empty word. What kind of past successes could be invoked in favour of —or against— the prior probability of a new hypothesis. Which are the relevant successful accomplishments of *L*-hypotheses? Since what is at issue is precisely the alleged link between explanatory merit and probability, an epistemic justification for good explanations —both current and past ones— cannot be provided by explanatory success, obviously.

Truth is other candidate for filling the gap. The aforementioned argument could be rephrased by saying that there is a higher proportion of *true* theories in those included in *L* than in the remaining. Therefore, we should confer higher prior probabilities to the former. Anyway, truth is particularly elusive for Bayesians. For a tautological claim, $p = 1$. For a true theory, presumably, $p = 1$. But an empirical theory is not a tautology devoid of content. Consequently, from a Bayesian perspective, truth must be understood just as a limit for empirical hypotheses.[8] Besides, theoretical truth is a disputable notion for many philosophers of science. As an IBE-Bayesian puts it,

> ... whatever the explanatory features happen to be, we can observe no correlation between those features and theoretical truth, for *we do not know which of our theories are true*. At most we could hope to observe a correlation between the features in question and continued empirical success, which is no help, for *we know nothing about the proportion of empirically successful theories that are true*. (Okasha 2000, p. 699; my emphasis).

Given all this, I think that it is desirable that Freq-Bay does not take any position on either side of the realism/antirealism debate. Identifying success with truth cannot be done on pain of giving a strong realist flavour to Freq-Bay. It is preferable then, not to take for granted realism and depart from observational success.

Freq-Bay goes, then, as follows. By and large, as evidence increased, *L*-hypotheses enjoyed a relatively long period of acceptance in contrast with non-*L* hypotheses. *L*-hypotheses coped better with empirical testing than their counterparts. Then, if h_1 is an *L*-hypothesis and h_2 does not belong to *L*, it is more probable that h_1 enjoy observational success than h_2. Given that observational success has confirmational import, and confirmational import is understood in probabilistic terms in the Bayesian framework, it is also likely that h_1 enjoys higher posterior probability than h_2. Now, let us grant that we get conclusive evidence from the historical record, that is, the percentage of *L*-theories which enjoyed a long-standing period of observational success is actually

[8] However, verisimilitude and distance from truth are legitimate scientific values for some Bayesians. See, for instance, Maher 1993.

much higher than the percentage of non-L ones. Freq-Bay would conclude that "h_1 merits greater prior probability than h_2."

Someone could object that this is a *non sequitur*. Prior probabilities are based on evidence about the historical record of observational success. This evidence allows us to infer at most that prior probabilities refer to the probability of getting continued observational success, that is, to the probability of verifying the observational consequences of h_1 and h_2, but not to the probability of the explanatory hypotheses —h_1 and h_2— themselves in case that they are theoretical hypotheses. Therefore what we are entitled to infer from the evidence is a different conclusion, namely, that h_1 is likelier to have *observational success* than h_2.[9]

This objection may be pressing for those who think that there is a cleavage between the observational consequences and the theoretical hypotheses. But it is not a proper objection in this context. Given that we are trying to discern the respective advantages of two views about the priors, we must focus on the differences between them. For Bayesians, observational success is understood in probabilistic terms and it has specific effects on posterior probability. Needless to say, Freq-Bay and IBE-Bay *agree* that observational evidence may increase (or decrease) the probability *of the theoretical hypotheses*. Prior probability attributions for novel hypotheses can be seen as estimations about the future empirical prospects of competing hypotheses and, also, of their eventual posterior probabilities. No further qualifications related to the observational-theoretical distinction are required here.[10]

Nevertheless, the frequentist meta-hypothesis is still rather ambiguous. To begin with, it should be distinguished from the rule —if there is any rule— to set the priors. Notice firstly that calculating prior probabilities by equating $p(h)$ —the probability of the new hypothesis under discussion— to $p(h/e)$ —where e is the global historical record of past success—, does not work, since the algorithm demands a value for $p(h)$ again. So no progress has been made. Besides, we should ascertain $p(e/h)$, that is, the likelihood of past scientific hypotheses' record of success given that h, the novel hypothesis, is true, and this conditional probability hardly makes any sense. We must conclude, then, that the information conveyed by the historical record of success as a whole cannot be considered as the evidence on which the new hypotheses should be conditionalized. That information, however, is indeed relevant to the frequentist hypothesis, since the latter is a general second-order hypothesis about the historical record of success of past first-order hypotheses. In the following paragraphs will try to confer a more detailed content to the frequentist second-order hypothesis. The pros-

---

[9] Radical sceptics about induction would not even accept such conclusion, of course. But we are not concerned here with the refutation of overall scepticism about induction.

[10] Bas van Fraassen combines a particular brand of Bayesianism with agnosticism about theoretical hypotheses. According to him, theoretical agnosticism could be properly represented in a Bayesian format as complete vagueness of opinion, i.e., for theoretical hypotheses the prior probability equates to the interval [0, 1] (see van Fraassen 1989, pp. 189 ff., and van Fraassen 1998). It should be remarked, however, that van Fraassen's agnosticism is grounded on a controversial notion of observability that is alien to the Bayesian framework (for a "Fraassenian" reply to a sophisticated version of the standard abductive defence of scientific realism, see Iranzo 2008a).

pects for testing it will be consequently improved. Then we will see how this proposal also suggests a simple rule for setting the priors.

Let H be the set of all hypotheses h_i ever formulated by scientists. L is a subset of H which contains all the L-hypotheses and L' is the complementary of L, that is, the subset of hypotheses which are non-L-hypotheses. Furthermore, E is the available evidence now, so $p(h_i/E)$ is the conditional probability of h_i from our current epistemic perspective. Some second-order hypotheses about the success record of all scientific hypotheses ever formulated are:

(a) Let κ be the probability for the most probable hypothesis in L'. Then, for every $h_i \in L$, $p(h_i/E) > \kappa$.

(b) Let λ be a determinate threshold of probability ($0 < \lambda < 1$). Let f_L be the relative frequency of theories in L —and $f_{L'}$ the relative frequency of theories in L'— whose probability is higher than λ. That is,

$$f_L = \frac{\left|\{h_i \in L : p(h_i / E) > \lambda\}\right|}{|L|} \text{ and } f_{L'} = \frac{\left|\{h_i \in L' : p(h_i / E) > \lambda\}\right|}{|L'|}.$$

Then, $f_L > f_{L'}$.

(c) Let m_L be the average probability in L and $m_{L'}$ the average probability in L'. That is,

$$m_L = \frac{\sum_{h_i \in L} p(h_i / E)}{|L|} \text{ and } m_{L'} = \frac{\sum_{h_i \in L'} p(h_i / E)}{|L'|}.$$

Then, $m_L > m_{L'}$.

Since each of these meta-hypotheses stipulate an alleged *sufficient* condition to attribute higher priors to L-hypotheses than to L' ones, all they consist in a strong inequality. Now let us compare them.

On my view, (a) is not a good option. It excludes the possibility that any $h_i \in L$ has been conclusively refuted, that is, $p(h_i/E) = 0$ is ruled out for every h (since κ should be below zero and L'-hypotheses would enjoy negative probability!). In addition to this, (a) is a very strong requirement. Maybe L-hypotheses are more probable than L'-hypotheses in general, but not always. According to (a), in contrast, we are bound to accept that L-hypotheses long time ago discarded are more probable than all L'-hypotheses, even those that are still in the arena. But many theories in the past, after enjoying observational success during some period of time, were abandoned. Some of them were also regarded as brilliant explanations of a particular observational domain. However, as the amount of evidence increased, their probability decreased and now is very low on the available evidence. On the other side, we could find good examples of L'-hypotheses whose observational success is impressive (in Quantum Physics, for instance). The moral is that we should not compare the probability of the hypotheses taken one by one. The relevant comparison should be between the overall results of two different sets of hypotheses, L and L'.

In contrast to (a), alternatives (b) and (c) allow that some L'-hypotheses may be more probable than some L-hypotheses. In particular, (b) assumes that good explanations are *highly* probable. That is the point for setting a threshold. At first sight it sounds reasonable to focus on theories which are more probable than its negation. But even though the threshold is 0.5, for instance,[11] and $f_L > f_{L'}$, in some particular circumstances it could be unjustified to attribute a higher prior to a novel L-hypothesis than to its L' rivals. The reason is that the dispersion of probability values in L and L' seems crucial since it may well occur that $f_L > f_{L'}$, but $m_{L'} > m_L$. Even though the proportion of L-hypotheses above the threshold is higher that the corresponding proportion of L'-hypotheses, if the average probability is higher for L' than for L, should we attribute higher priors to L-hypotheses?

Alternative (c) gives more credence to L-hypotheses and is not affected by the dispersion of probability values in L and L'. Couldn't we combine conditions (b) and (c)? Yes, but we shouldn't, because the necessary connection between good explanations and high probability assumed by alternative (b) is not required. Freq-Bay claims that the basis of explanatory assessments is frequency. It gives a justification in probabilistic terms for defending that good explanations deserve higher priors. But in order to justify such higher attributions of prior probabilities to L-hypotheses, it suffices if L-hypotheses *generally make the best* in comparison to L'-hypotheses, even though m_L is below 0.5. Besides, our particular version of the frequentist meta-hypothesis suggests an easy rule to set the priors: attribute to the novel hypothesis the average value found. That is,

$$\forall h_i \mid h_i \in L, \text{ then } p(h_i) = p(h_i/(m_L = x)) = x$$

$$\forall h_i \mid h_i \in L', \text{ then } p(h_i) = p(h_i/(m_{L'} = y)) = y$$

This rule is formally similar to some other rules proposed —like Lewis's Principal Principle, van Fraassen's Reflection Principle, Gaifman's "expert functions"—, although its rationale is different.

The idea that past success may be useful to foresee future success is a reasonable claim, save for those who are sceptics about induction, so I take it for granted in my argument. The problem for us is not the justification of induction but, rather, the justification of scientists' preferences for explanatory rewarding theories within the Bayesian framework. From a general perspective my answer is the same as Salmon's one, namely, that the initial plausibility assigned to extant scientific hypotheses somehow reflects the success obtained by other hypotheses in the past. However, I have tried to go beyond his sketchy suggestion. To sum up my proposal:

- the justification of scientists' preferences for explanatory merit depends on the correctness of a conjecture —the "frequentist meta-hypothesis", as I call it—

[11] Peter Achinstein has suggested a threshold of 0.5 for defining evidence. A necessary condition for e to be evidence for h —a good reason to believe that h— is that $p(h/e) > 0.5$ (Achinstein 2001, chapters 6 and 7). I endorse this condition in Iranzo, 2008b.

about the history of science: good explanations are more successful than bad ones;

- the meta-hypothesis relates all scientific hypotheses devised in the history of science to the available evidence right now;
- success should be understood in probabilistic terms: probability conditioned to the evidence available now $p(h/E)$;
- a detailed interpretation of the meta-hypothesis asserts that sounds initially plausible is that the average value of $p(h/E)$ for explanatory valuable hypotheses is higher than the average value for explanatory defective ones;
- insofar as this conjecture is sound, novel hypotheses that are explanatory valuable deserve higher priors than their defective rivals;
- m_L and $m_{L'}$, that is, the average values of $p(h/E)$ for explanatory lovely and non-explanatory hypotheses, respectively, are the recommended values for the initial probability of novel hypotheses —provided that they can be somehow estimated.

After giving specific content to the frequentist meta-hypothesis we are in a better position to compare IBE-Bay to Freq-Bay.

4. The redundancy of empirical testing?

The subjectivist view about the priors was mentioned above. In the end, prior probabilities are nearly irrelevant for subjective Bayesians, because they rely on the evidence's power to wash them out through a convergence process. However, for Freq-Bay and IBE-Bay, as they were characterized in section 2, priors are not just a device to make the required calculations. Non-subjectivist Bayesians accept that conditionalization on evidence is a good policy to learn from experience, of course. But, given that the priors are a normative/objective constraint for them, if a hypothesis had a very high prior probability —or if it were a very good explanation—, perhaps we should not worry about how to test it. High priors should be considered as highly reliable indicators of future empirical success and, consequently, of future increments in posterior probabilities. Needless to say, a theory of confirmation with such consequence —the redundancy of empirical testing— would be severely impaired, at least as a theory about science as we know it.[12]

Notice that this question does not arise in the subjectivist framework. Conditionalization to experience does the whole work leading to the consensus. It is the non-subjective status of (high) priors that suggests the potential redundancy of empirical tests. In the remainder of the paper I will discuss whether "the redundancy of empirical testing objection" is a serious risk for any of both versions of Bayesianism, Freq-Bay and IBE-Bay.

---

[12] On the reasons for relying on experience, not only in science but in general, see my paper "On the Epistemic Authority of Experience" (in press).

The redundancy objection against IBE-Bay goes as follows: granted that explanatory merit has confirmational value, if the explanatory merit of h_1 —the lovely hypothesis— is very high, why do we need empirical testing for it? Couldn't its prior probability be so high that it would render useless to test it? A direct reply is that the explanatory merit of a hypothesis cannot be so high as to turn empirical testing into a dispensable constituent of scientific practice. But this reply clashes with our assessments of explanatory goodness. In everyday contexts there are explanations so good that we are sure they are true. We do not need to compare them with some bizarre alternatives. Admittedly, things are more complicated in science. We remarked earlier that, to some extent, the explanatory merit of a scientific hypothesis depends on its relation to a particular explanandum. The situation then is similar to what occurs in everyday contexts. It is not difficult to find historical examples of hypotheses which were considered really good *qua* explanations (at least from the perspective of the scientific community at time t). In consequence, there are highly probable hypotheses for the IBE-Bayesian.

The foregoing suggests that, in principle, there is no reason not to think that prior probabilities could be high —assuming that priors' values should exactly parallel comparative assessments of explanatory merit. Nevertheless, IBE-Bayesians could try a different answer to meet the challenge. Prior probability is based on generic considerations which are independent of the specific observational consequences of h_1. But, as we noticed in section 1, some aspects of explanatory value are intrinsic properties of the hypothesis while some others have to do with the relation between the hypothesis and the evidence. And if explanatory value should be included not only in $p(h_1)$ but in $p(e/h_1)$ as well, then, the redundancy objection is not well-founded. It concludes that $p(h_1)$ is high because h_1 is very good qua explanatory hypothesis, but it may well be that h_1 is very good from an explanatory point of view just because $p(e/h_1)$ is high, even though $p(h_1)$ is notably low. In that case h_1 would still be a good explanation, but since its prior probability would be nearly irrelevant for that, empirical testing still seems mandatory for h_1 (at least it is non-redundant).

Unfortunately, I do not think that IBE-Bay could maintain that h is a good explanation even though $p(e/h)$ is high and $p(h)$ is low. The numerator of Bayes's formula —$[p(e/h) \cdot p(h)]$— is the Bayesian translation, so to say, of explanatory merit, so a good explanatory hypothesis is an explanation that enjoys a high value for the numerator. And we cannot get such high value with a high likelihood unless $p(h)$ is high too —recall that probability function ranges from 0 to 1, and a low initial probability can dramatically decrease the value for the numerator. Therefore, if $p(e/h)$ were high and $p(h)$ low, h would not be a good explanation. Again, if $p(h)$ is high, why should we concern about empirical tests for h?

Apparently, the prospects for Freq-Bay are not better. According to it, prior probabilities are based on objective and *empirical* frequencies. Insofar as priors are based on that, there is no reason not to take them as seriously as posterior probabilities. The difference is, of course, that posterior probabilities take into account specific observational —empirical— consequences of the hypothesis. But a high value for h_1's prior probability means that h_1 is well supported by the evidence even though it is the sort of "indirect" evidence provided by the frequentist meta-hypothesis.

Nevertheless, it is not very complicated for the frequentist to defend the claim that priors cannot be high, even for hypotheses that enjoy several explanatory virtues. Recall that the version of the frequentist hypothesis I favoured in section 3 demands that $m_L > m_{L'}$, and it is worth noticing here that this condition may be fulfilled even though both average values are not high. We are still waiting for a complete probabilistic ranking of all scientific hypotheses ever postulated. It is doubtful whether that task could eventually be accomplished. Anyway, we should expect that the values of m_L and $m_{L'}$ are low, since many of the past theories have been discarded because their probabilities are low, even zero, according to the available evidence.

Just a final remark. We have accepted that scientists' assessments of explanatory merit occasionally may be high. But the frequentist-Bayesian says that prior probabilities are disguised assessments of relative frequencies of success and that they, in general, should be low. Is there any contradiction here? I do not think so. Scientists' plausibility assessments operate through rules of thumb that try to discern those features which have been associated to success in the past. There are psychological biases among scientists in favour of some methodological preferences. These preferences are those reinforced in the learning process accomplished for becoming a scientist and very often scientists themselves can hardly make them explicit. Now, according to IBE-Bay the value of the new hypothesis' prior probability is exclusively grounded on the qualitative scientific appreciation of its explanatory merit. But for Freq-Bay it is not necessary that scientific judgment exactly matches the value found in the historical record. Rather, the important thing is that, when confronted with a set of rival hypotheses, scientists' preferences agree with the ordered list obtained by the rule proposed in section 3.

5. Conclusions

The prospects for bayesianizing IBE crucially depend on the interpretation of the priors involved in scientific judgment. I have distinguished two interpretations, IBE-Bay and Freq-Bay, that disagree on the ultimate rationale for assessments of explanatory merit. While IBE-Bay demands an epistemic import for explanatory value, it is not clear how it can substantiate this claim. I have argued that Freq-Bay, in contrast, allows a genuine epistemic role for explanatory merit via its historical link to success and probability. Besides, Freq-Bay fits better with two basic facts about science than IBE-Bayesianism. The first one is that a large proportion of the past scientific hypotheses have a low probability. Certainly, this is a conjecture by now, but I do not think it is pure guesswork. The second one is that scientists make great efforts to test hypotheses. The fact is that they look for confirming or refuting evidence even for those hypotheses which are really good as explanations.

The frequentist approach has an additional slight advantage over IBE-Bay. Ironically, it is better than IBE-Bay from an explanatory point of view. In section 2 we alluded to a basic consensus among scientists about the values which should be taken into account to assess the priors. This fact could be explained by IBE-Bay as a result of scientific education. Frequentists would not deny such explanation. But they could also explain why those explanatory values are precisely the ones favoured by scientists: because they are linked to observational success and maximization of probability. Scientific education is well suited for obtaining agents who are good trackers of those epistemic values. After all, this is one of the goals scientific education has been designed for.

On account of all this my conclusion is that the explanationist approach to Bayes's Theorem afforded by IBE-Bay is seriously flawed.*

REFERENCES

Achinstein, P. (2001). *The Book of Evidence*. New York: Oxford University Press.

Christensen, D. (1991). "Clever Bookies and Coherent Beliefs", *Philosophical Review* 100, 229-247.

Day, T., and H. Kincaid (1994). "Putting Inference to the Best Explanation in Its Place", *Synthese* 98, 271-295.

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

Earman, J. (1992). *Bayes or Bust. A Critical Examination of Bayesian Confirmation Theory*. Cambridge (Ma): MIT Press.

Gillies, D. (2000). *Philosophical Theories of Probability*. London: Routledge.

Hon, G., and S. Rakover (eds.) (2001). *Explanation. Theoretical Approaches and Applications*. Dordrecht: Kluwer.

Howson, C. (2003). *Hume's Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.

——— and P. Urbach (2006). *Scientific Reasoning. The Bayesian Approach*. 3rd ed. Chicago: Open Court.

Huber, F. (2005). "Subjective Probabilities as Basis for Scientific Reasoning?", *British Journal for the Philosophy of Science* 56, 101-116.

Iranzo, V. (2008a). "Reliabilism and the Abductive Defence of Scientific Realism", *Journal for General Philosophy of Science*.

——— (2008b). "Abduction and Inference to the Best Explanation", *Theoria* 22/3 (60), 339-346.

——— (in press). "On the Epistemic Authority of Experience", *International Journal of Philosophical Studies*.

Joyce, J. (2003). "Bayes's Theorem", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2003 Edition), URL: http://plato.stanford.edu/archives/win2003/entries/bayes-theorem/.

Kuipers, Th. (2002). "Beauty, A Road to the Truth?", *Synthese* 131, 291-328.

Lipton, P. (2001). "What Good is An Explanation?", and "Is Explanation a Guide to Inference? A Reply to W. Salmon", in Hon and Rakover (eds.), 43-59, 93-120.

——— (2004). *Inference to the Best Explanation*. London: Routledge.

Maher, P. (1993). *Betting on Theories*. Cambridge: Cambridge University Press.

Niiniluoto, I. (2004). "Truth-Seeking by Abduction", in F. Stadler (ed.), 57-82.

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

Psillos, S. (2002). "Simply the Best: A Case for Abduction", in A.C. Kakas and F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond (Essays in Honour of Robert A. Kowalski, Part I)*. Berlin: Springer.

——— (2004). "Inference to the Best Explanation and Bayesianism", in F. Stadler (ed.), 83-91.

Salmon, W. (1967). *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

——— (1970). "Bayes's Theorem and the History of Science", in R. Stuewer (ed.), *Historical and Philosophical Perspectives on Science*. Minneapolis: University of Minneapolis Press.

——— (1996). "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes", in D. Papineau (ed.), *The Philosophy of Science*. London: Oxford University Press.

* I am grateful to the audiences at the XI Italian-Spanish Meeting for Analytic Philosophy (especially to Roberto Festa), the Fifth European Congress for Analytic Philosophy, and the V Coloquio Compostelano de Lógica y Filosofía de la Ciencia. The paper also owes a good deal to several helpful referees.

Salmon, W. (2001). "Explanation and Confirmation: A Bayesian Critique of Inference to the Best Explanation" y "Reflections of a Bashful Bayesian: A Reply to P. Lipton", in Hon and Rakover (eds.), 61-91 and 121-136.

Stadler, F. (ed.) (2004). *Induction and Deduction in the Sciences*. Dordrecht: Kluwer Academic Press.

van Fraassen, B. (1983). "Calibration: A Frequency Justification for Personal Probability", in R.S. Cohen and L. Laudan (eds.), *Physics, Philosophy and Psychoanalysis*. Dordrecht, Reidel, 295-319.

——— (1989). *Laws and Symmetry*. New York: Oxford University Press.

——— (1998). "The agnostic subtly probabilified", *Analysis* 58, 212-220.

Valeriano IRANZO is Senior Research Fellow (Programa Ramón y Cajal) at the department of Lógica y Filosofía de la Ciencia (Universitat de València, España). He has published on epistemology and philosophy of science, especially on theory of justification and scientific realism. His current philosophical interests are confirmation theory, foundations of probability and statistics, and the "theory-choice" problem.

ADDRESS: Departamento de Lógica y Filosofía de la Ciencia, Facultad de Filosofía y CC. Educación, Blasco Ibáñez 30, Valencia – 46010, España. E-mail: Valeriano.Iranzo@uv.es.