WHY ARE GOOD THEORIES GOOD?
Reflections on epistemic values, confirmation, and formal epistemology.

Jesús Zamora-Bonilla
UNED, Madrid

ABSTRACT: Franz Huber's (2008) attempt to unify inductivist and hypothetico-deductivist intuitions on confirmation by means of a single measure are examined and compared with previous work on the theory of verisimilitude or truthlikeness. The idea of connecting 'the logic of confirmation' with 'the logic of acceptability' is also critically discussed, and it is argued that 'acceptability' takes necessarily into account some pragmatic criteria, and that at least two normative senses of 'acceptability' must be distinguished: 'acceptable' in the sense of 'being allowed to accept', and 'acceptable' in the sense of 'being obliged to accept'. Lastly, some connections of confirmation theory with naturalism, intertheoretic reduction, and explanation vs. understanding are explored.

KEYWORDS: Confirmation, Bayesianism, induction, hypothetico-deductive method, verisimilitude, truthlikeness, acceptance, coherence, pessimistic induction, naturalism, theoretical reduction, explanation, understanding.

1. Introduction.

One of the fundamental problems of epistemology (or 'theory of knowledge') is to understand what makes some cognitive states preferable to others. Philosophy of science, considered as a branch of epistemology, asks the same question but applied to the 'cognitive states' characteristic of scientific activity, as believing or accepting a theory, law, hypothesis, model, principle, or observational report, or, more often than what might seem from the reading of philosophy of science 'classics', as believing or accepting things like 'this procedure is rather useful here', 'this is a good journal to submit our paper', or 'only a fool would considered *that* a serious problem'. On the other hand, epistemology as a branch of *philosophy*, rather than as a branch of psychology or anthropology, usually tries to make us understand these questions at a very *abstract and general* level. Of course, there are differences between individual philosophers or philosophical schools regarding what level of abstractness and generality is the right one, and, as readers of *Synthèse* will immediately realise, there are many philosophers for which the obvious, or at least the default approach, is just to go directly to the *most* abstract and general level that can be conceived; one recent example of this is Huber (2008a), which tackles in an elegant way the problem of what is what makes good 'theories' good (at least, in the sense of what makes a theory to be 'better confirmed' than another), and which performs this as an exercise of (almost) pure mathematical logic. Though I agree with most of Huber's insights about confirmation, I think that the purely formalistic approach of which his paper is representative has serious problems to become acceptable as the most insightful account about the evaluation of scientific theories, as compared to a more 'naturalist' approach. The structure of my paper will be as follows: in section 2, a critical summary

Draft.
Forthcoming in *Synthèse*. Send comments to jpzb@fsof.uned.es

of Huber's theory of confirmation is presented; section 3 compares Huber's theory of confirmation with the theory of verisimilitude I have developed elsewhere; section 4 discusses the connection between 'the logic of confirmation' and 'the logic of acceptance'; and finally, section 5 sketches a formal naturalist approach to philosophy of science in which this type of discussions can be framed.

2. Huber's theory of confirmation.

One possible way to understand Huber (2008a) is as an attempt to unify two classical interpretations of the concept of 'confirmation'. It seems that one of the basic peculiarities of scientific knowledge, i.e., of what makes a certain proposition or claim *deserve* the label 'scientific knowledge', is that it has been 'empirically confirmed'. There is no need to summarise here the longstanding discussion about what are the appropriate or the logically possible ways (if there are any) of *confirming* scientific hypotheses, a discussion which started at least with Hume's analysis of inductive arguments, for Huber's main goal is the construction of a unified notion covering in a coherent and systematic way the insights coming from two different and apparently conflicting ways of understanding 'confirmation': the 'inductivist' and the 'hypothetico-deductivist'. I agree with Huber in that the conflict between these two notions is what plagued with deep problems Hempel's and Oppenheim's original work on the logic of qualitative confirmation (Hempel and Oppenheim, 1945), and forced researchers on the topic to choose either to work on something like the 'logic of induction' (e.g., Carnap, Hintikka, Levi), or to follow some other hypothetico-deductivist approach to understand scientific inference (like abduction theory, inference to the best explanation, etc.). Huber, instead, tries to save Hempel's intuition that there can be something like a unified 'logic of confirmation', one that respects *both* the idea that a hypothesis H is confirmed by empirical evidence E if H can be formally inferred from E (i.e., if *the theory* derives from the data) and also the idea that H is confirmed by E if we can formally infer E from H (i.e., if *the data* derive from the theory). According to Huber, the reason of this conflict is that there are (at least) two virtues a good theory must have, *truth* (or 'plausibility') and *strength* (or 'informativeness'), and there is a trade-off between them (i.e., the kind of things we can do to improve the 'truth' of our theories usually go against what we can do to improve their 'strength', and vice-versa). So, imagine we try to define some measures f of the epistemic value of a hypothesis H given the evidence E and background knowledge B; functions $f(H,E,B)$ may have many different mathematical properties, but the following ones are what would allow to call it, respectively, 'truth responsive' and 'strength responsive' *on the basis of evidence*[1] (see Huber (2008a), pp. 92-93):

(1) (a) If E entails $H \rightarrow H'$, then $f(H,E) \leq f(H',E)$
(b) If $\neg E$ entails $H' \rightarrow H$, then $f(H,E) \leq f(H',E)$

The condition in (1.a) means that, within the states allowed by E, all the states allowed by H are also allowed by H', i.e., H' covers a bigger portion than H of the states of the world consistent with our evidence; so, if f behaves in the way stated by (1.a), it will give a higher value to theories that would be true in a bigger proportion of the

[1] I will dispense of B in what follows, save when the discussion demands explicit reference to it.

states that *we know might be true*. The condition in (1.b) means, instead, that in the states that are *not* allowed by *E*, *H'* covers a smaller portion of them than *H*, and so, the more *content* that we know is false a hypothesis has, the better (recall that the content of a proposition is inversely related to the magnitude of the states it is consistent with). Fig. 1 allows to clearly understand these two conditions ((1.a) refers to what happens 'inside' *E*, whereas (1.b) refers to what happens 'out' of *E*; the square represents the set of states consistent with background knowledge *B*).

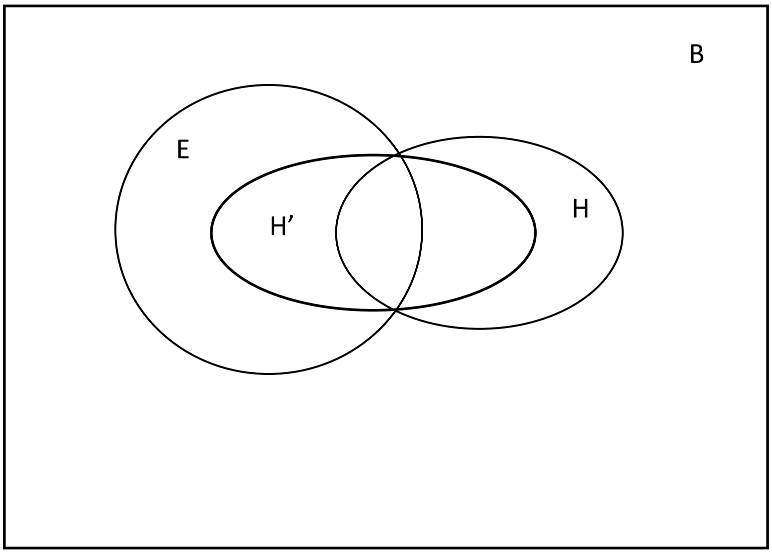

Figure 1

The main accomplishment of Huber's paper is that it shows that any measure of the epistemic value of theories that is sensitive both to plausibility (as formalized by 1a) and to informativeness (as formalized by 1b) would make justice to the two conflicting intuitions behind the notion of 'confirmation', and so, *f* might be taken as a model of a 'logic of confirmation'. Of course, *f* may be responsive to other values as well; for example, we can conceive of it as a functional $f(t,s,u_i)$ containing a 'truth function' *t* responsive to (1.a), a 'strength function' *s* responsive to (1.b), and one or more other functions u_i having to do with other concerns scientists may have (simplicity, heuristic power, precision, explanatoriness, , and so on). If we consider *s* and *t* alone, condition (1) summarises according to Huber the basic ideas about confirmation. Typically, a function satisfying (1.a) would be $p(H,E)$, whereas a function satisfying (1.b) would be $p(\neg H,\neg E)$. So, Huber proposed definition of epistemic value is (Huber (2008a, 97-98)):

(2) $f(H,E) = p(H,E) + p(\neg H,\neg E)$

The most important insight in Huber's theory is, in my opinion, not merely his thesis that both informativeness and plausibility are values that an appropriate notion of confirmation must take into account, and that no explication based on only one of them will do the work, but the way in which he suggests these (and probably other) values must be taken into account by a theory of confirmation: traditionally, philosophers of science have understood the acceptance of a scientific hypotheses under the model of *belief* (or degrees of belief); studies on confirmation have often been elaborated as aiming at a set of properties a theory should have in order to 'deserve to be believed'. Huber proposes, instead, that scientific acceptance is to be understood under the model of the *purchase of a good* (Huber, 2008a, 106): the aim of theory assessment should be to define the set (or 'pool') of *acceptable* hypotheses, those that have a combination of virtues that make them to be 'good enough' (or that surpass what Huber calls elsewhere a 'demarcation' level, though little is offered by Huber that allows to illuminate whether there is some sense in which the choice of one level instead of another can be considered 'rational', i.e., 'optimal'; more on this below).[2]

3. Confirmation and truthlikeness.

Huber's claims about the logic of confirmation are consequential and accurate; I think, however, that they suffer from an important limitation due to the fact that he seems to have reached his conclusions by thinking 'from within' the tradition of inductive logic or 'studies on confirmation'. Basically, though it is true that both the inductivist and the hypothetico-deductivist insights must be taken into account in the analysis of the epistemic values of scientific theories, Huber disregards much of the work that has been done in the latter field, from Popper's and others' attempts to define the value of a theory on the basis of a notion of distance to the full truth (i.e., verisimilitude or truthlikeness, which is defined by some authors precisely on the basis of a combination of truth and information; cf. Maher (1993, 210 ff.)), to the abundant literature on predictive and explanatory capacity as the main reasons that justify the acceptance of theories. As a matter of fact, hypothetico-deductivists (championed by Popper) tend to be little satisfied when 'confirmation' is used to refer to 'the' virtue that makes theories acceptable, for they would rather reserve that term *only* for Huber's inductivist intuition, and they would say, instead, that acceptable theories must have (if at all) a high degree of confirmation, and also *other* values, like the capacity of explaining or predicting the empirical evidence.[3] I shall not dispute in this paper Huber's employment of 'confirmation' to refer to that combination of 'inductivist' and 'hypothetico-deductivist' virtues that makes a hypothesis 'acceptable', but I think that it is *de rigueur* to point to the fact that *informativeness* is by no means the only epistemic value that hypothetico-deductivist philosophers identify as

[2] Huber also discusses (ibid., p. 104; cf. also Huber (2008b)) Carnap's thesis that the cause of the 'confusion' behind Hempel's qualitative explication of confirmation might have been his implicit commitment to both a notion of 'incremental confirmation' (according to which a theory *H* is confirmed if $p(H,E) > p(H)$), and to a notion of 'absolute confirmation' (i.e., if $p(H,E)$ surpasses some arbitrarily determined 'high' threshold). Huber rightly suggests, instead, that the two notions in conflict might have originally been 'absolute confirmation' and '*informativeness*', and that it was the apparent contradiction between these two goals what finally led Hempel to replace informativeness by 'plausibility' (i.e., something related to 'incremental confirmation'), though keeping, in his list of desiderata for a notion of confirmation what he called 'converse consequence condition' (i.e., the condition asserting that, if *E* confirms *H*, and *H*' entails *H*, then *H'* , then *E* also confirms *H'*; this condition seems reasonable if we understand 'confirms' in the sense that *E* derives –is a 'prediction'– from *H*).

[3] This is particularly valid when we take into account the Popperian claim that all 'interesting' theories are probably *false*, and will be probably refuted by some empirical data sooner or later.

relevant; i.e., perhaps the idea of *p(H,E)* being 'high' (or being higher than an arbitrary threshold, or tending to 1) is what most inductivists consider 'the' main value a theory must have in order to be accepted (and perhaps not), but 'being informative' certainly does not exhaust what people like Popper, Lakatos, Niiniluoto, Kuipers, or Lipton would think should be added to (or taken into account *instead* of) 'plausibility' for having a complete explication of epistemic value; for example, truthlikeness may be one of these other possible values.[4] The main intuition under the notion of truthlikeness is that of the value of two *false* theories can be different, and even that the value of *some false* (or falsified) theories can be higher than the value of *some true* (or not yet falsified) theories. If truth and strength were the only relevant virtues of a hypothesis, then, if *H* and *H'* have been falsified, then either both will have the same value, or the most informative of it will be better, but the latter entails that increasing the epistemic value of a falsified theory is a child's play: just add to it by conjunction any statement not entailed by it (i.e., *H&G* will be better than *H*, if *H* is refuted). So, if we are to take seriously the idea that *H* can be epistemically better than *H'* in spite of both being known to be false, then their value must depend on something else than informativeness:[5] in a nutshell, this value will not only depend on *how much* they say about the world, but also on something that has to do with *what* they say, in particular, on what is the *relation* between what they say and the truth about the matter. And, obviously, if the value of falsified theories depends on that kind of things, the same may happen with not-yet-falsified ones.

So, conditions (1.a-b), even if reasonable, should be taken at most as a *partial* account of what values a theory must have in order to become 'confirmed' (i.e, 'acceptable'). The example of the theory of empirical verisimilitude I have developed in a number of papers nicely illustrates why an account of confirmation like Huber's is, hence, right but partial.[6] Let's start by defining a *naive* notion of verisimilitude (i.e., one that does not take into account the possibility of valuing falsified theories), and

[4] Some readers will have noticed the similarity between fig. 1 and Kuipers' explication of his notion of 'naive truthlikeness'. The difference is that, in the case of Kuipers' definition, the set *E* is replaced by the set of all models or worlds that are physically possible according to the strongest true theory. See Kuipers (1992).

[5] Actually, Huber (2008, 114-5) explicitly considers the possibility of taking into account other values, though he does not discuss what these other values can be, nor if there would be some difference between how the introduction of epistemic vs. non-epistemic values would affect the global acceptability of theories.

[6] See Zamora-Bonilla (1996), (2000) and (2002a). Though I called *'empirical verisimilitude'* my own definition of epistemic value, it was a rather atypical one within the verisimilitude programme, for it is based on a direct comparison between the theories and the empirical evidence, not assuming that there is something like an '*objective* degree of truthlikeness'; the intuition behind this is that, if we take seriously the notion of an 'epistemic *utility'*, we have to respect the fact that *utility functions are subjective*, i.e., one subject's utility can only be affected *by what the agent can notice*. This point is relevant for almost all other approaches in which the assumption is made that the 'value' of having a true theory is higher than that of having a false theory: as long as we consider subjective valuations (and I cannot conceive of what we could mean by a valuation which were not 'subjective', in the sense of made by one particular subject in function on the data she contingently has), only things that the subject can actually perceive (in *some* relevant sense) should be taken into account. In this sense, 'truth' (as something conceptually independent on empirical information) is perhaps not the right kind of thing to be considered a 'value', not even an 'epistemic value'. This sceptical comment, however, does not play any role in my subsequent arguments, which must either stand or fall by their own merits, but is rather a mere personal motivation for the choice of my approach.

then proceed to offer some sophisticated definitions applicable to more realistic cases. The naive definition asserts that the epistemic value of a theory depends on two factors:

a) how *similar or coherent* are the view of the world offered by the theory and the view of the world that derives from our empirical evidence; and

b) how *informative* our *empirical evidence* is (for being coherent with a very shallow empirical knowledge is not as indicative of 'deep truth' as being coherent with a much richer corpus of empirical information).

The coherence or similarity between H and E can be defined as $p(H\&E)/p(HvE)$,[7] whereas the informativeness of a proposition A can be measured by $1/p(A)$. Hence, the naive definition of empirical verisimilitude would be as follows:

(3) $Vs(H,E) = [p(H\&E)/p(HvE)][1/p(E)] = p(H,E)/p(HvE)$

It is easy to check that, if the two conditions in (1.a-b) are satisfied, then $Vs(H,E) \leq Vs(H',E)$. Of course, if only one of them is satisfied (which, as an anonymous referee indicates, is the normal case), it is still possible that the reverse comparison of verisimilitude is true. Fig. 2 illustrates this situation: the condition in (1.a) is fulfilled in fig. 2.a, and the condition in (1.b) is fulfilled in fig. 2.b, but in both cases H would be more verisimilar than H'. On the other hand, since $Vs(H,E)$ is a function of $p(H,E)$, it can be taken as a member of the Bayesian family. It has many other nice properties, like the following ones, many of them it shares with Huber's f:[8]

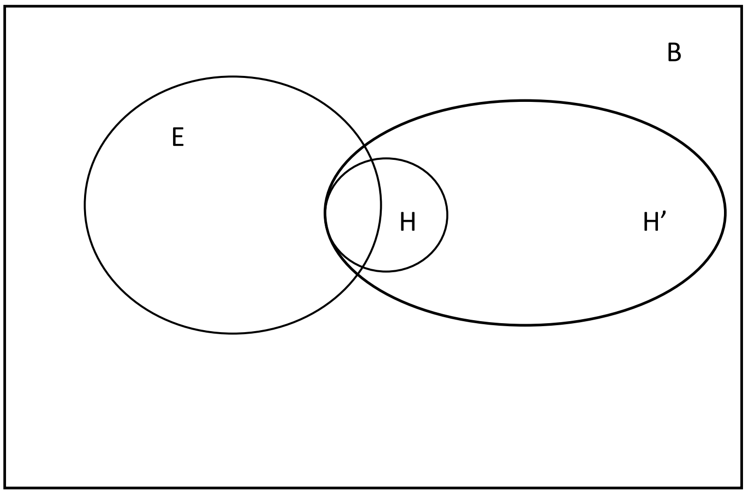

Figure 2.a

[7] This is analogous to the definition of coherence in Olsson (2002). The problem I'm discussing here is not directly related, however, to the recent discussion on coherence and confirmation, which has basically to do with the question of whether the coherence between several *data* is indicative or not of the truth *of the data*; what is relevant here is, instead, the relation between data and theories (and the truth of the data is assumed). Note that confirmation relations are usually asymmetric (we want to assess whether the data confirm the theory, not whether the theory 'confirms' the data). Furthermore, in that discussion 'confirmation' is usually understood as a high conditional probability (with which some results prove that no logical connection exists with 'coherence'; see, e.g., Bovens and Hartmann (2005)), and not as a combination of inductivist and other types of virtues, as here is assumed.

[8] For the proofs, see Zamora Bonilla (1996), section 2, except for (4.j); for this, see Zamora Bonilla (2002a, 357).

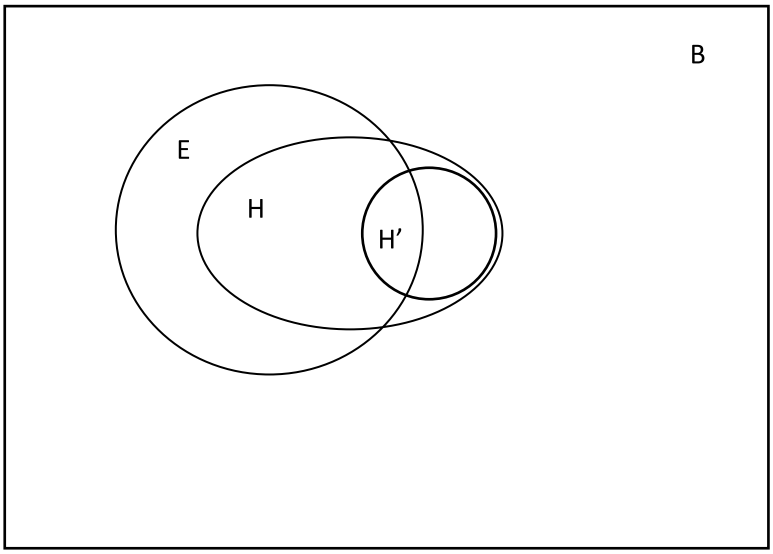

Figure 2.b

(4) (a) Respect for plausibility and for strength:

If $p(H) = p(H')$ and $p(H,E) \leq p(H',E)$, or if $p(H,E) = p(H',E)$ and $p(H') \leq p(H)$, then $Vs(H,E) \leq Vs(H',E)$.

(b) For theories explaining the data, the more probable, the better:

If $H \vdash E$, then $Vs(H,E) = p(H)/p(E)^2$

(hence, if $H \vdash H'$, and $H' \vdash E$, then $Vs(H,E) \leq Vs(H',E)$).

(c) For theories verified by the data, the more informative, the better:

If $E \vdash H$, then $Vs(H,E) = 1/p(H)$

(hence, if $E \vdash H$, and $H \vdash H'$, then $Vs(H',E) \leq Vs(H,E)$; also as a corollary, $Vs(Taut,E) = 1$).

(d) Confirming a prediction increases verisimilitude (deterministic version):

If $H \vdash F$, then $Vs(H,E) \leq Vs(H,E\&F)$

(e) Confirming a prediction increases verisimilitude (probabilistic version):

If $p(F,E) \leq p(F,H\&E)$, then $Vs(H,E) \leq Vs(H,E\&F)$

(f) Confirming a prediction increases verisimilitude (conditional version):

If $H \vdash (E \rightarrow F)$, then $Vs(H,E) \leq Vs(H,E\&F)$

(g) Partial verification (joining to a theory a hypothesis verified by the data, increases the theory's verisimilitude):

If $E \vdash G$, then $Vs(H,E) \leq Vs(H\&G,E)$

(h) Partial disconfirmation (the empirical evidence can help to disconfirm one part of a theory):

If $p(G,H\&E) \leq p(G,H)$, then $Vs(H\&G,E) \leq Vs(H,E)$

(i) Upper verisimilitude limits ('unended quest'):

$1/p(H) \geq Vs(H,E) \leq 1/p(E)$

(j) Expected verisimilitude equals confirmation ratio:[9]

$$E[Vs(H,E)] = p(H,E)/p(H) = p(E,H)/p(E)$$

Of course, this definition is 'naive' in the sense that it gives the value *0* to all falsified theories (i.e., theories for which $p(H,E) = 0$), and so, it does not respect the intuition that falsified theories can have different degrees of epistemic values, and even be more valuable than some non falsified theories. There are several possible strategies to solve this problem. One would be to take into account margins of approximation; let S^u be the statement that asserts that 'the actual state of the world is within a margin of approximation *u* from some state in which *S* is true"; then, a plausible definition of *approximate verisimilitude* would be:

(5) $AVs(H,E) = max_u Vs(H,E^u)$

A different strategy would consist in taking into account the fact that the empirical evidence usually consists in a finite set of mutually independent statements E_i (e.g., empirical laws); let $\mathfrak{F}$ the set of all possible conjunctions of the E_i's; then a definition of *sophisticated verisimilitude* would be:

(6) $SVs(H,E) = max_{F\in\mathfrak{F}} Vs(H,F)$

A theory will have a high degree of approximate verisimilitude, even if it has been literally refuted, if it explains well the empirical evidence within a small margin of approximation. It will have a high degree of sophisticated verisimilitude, if it explains most of the known empirical laws. Of course, other similar definitions, or combinations of them, are possible.

What can we infer from the comparison of functions like those defined by (3), (5) and (6) with Huber's account of confirmation? Here is a list of some relevant aspects, some more technical, some more philosophical:

a) 'Respect of strength' and 'respect for truth' need not be modelled as two separate arguments of a confirmation function; this is something that will only depend on analytical convenience.

b) Huber's theory implies that, amongst falsified hypotheses, epistemic value covaries with strength, and so, the best one amongst falsified theories will be the contradiction; actually, Huber's confirmation function entails that the value of a contradiction equals β (the arbitrary 'demarcation' threshold between acceptable and non-acceptable hypotheses, used in Huber (2008a, 94)). Definition (2), even if naive, does not lead to the paradoxical conclusion that a contradiction is epistemically better

[9] Instead, the expected value of $f(H,E)$ is $p(H\&E) + p(\neg H)p(E)$, assuming, like is also supposed in (4.j), that *E* can be decomposed into points of negligible probability. $E[G(H,E)]$ is calculated with the formula $\sum_{x\vdash E\&H} p(x)G(H,E) + \sum_{x\vdash E\&\neg H} p(x)G(H,E)$. I admit, however, that the use of the expectation operator would mean different things within my essentially subjectivist approach (in which it fits as a mere case of expected utility) and within other approaches (in which the values on which we average are taken as 'objective' –e.g., truth– values); so, for example, it can be the case that in Huber's approach the calculation of the expected value of function *f* would not have necessarily a relevant meaning, and so, comparisons of our approaches on the basis of this result is not illuminating. But, as long as one *can* interpret *f* as a kind of subjective utility (though this is not necessarily Huber's preferred interpretation, of course), then these comparisons could make some sense.

than any other falsified theory, but strategies like those behind definitions (4) and (5) allow us to be more realistic about the valuation scientists implicitly do of theories-known-to-be-false. Of course, nothing precludes Bayesians of also using such strategies to represent scientific valuations, though I don't know of any example of doing this.

c) Huber's condition on β (i.e, that β equals the value of a hypothesis with maximal strength and minimal plausibility, or with minimal strength and maximal plausibility) entails that, for his Bayesian confirmation function, $f(H,E) \geq \beta$ iff $p(H,E) \geq p(H,\neg E)$, which seems a too arbitrary requisite.[10] For *Vs*, instead, the value of a theory with maximal strength and minimal plausibility (i.e., one entailing *E*, but of null prior probability) is *0*, whereas the value of a theory with minimal strength and maximal plausibility (i.e., a tautology) is *1* (for 4.c). It seems more natural to choose the latter as a lower limit for the choice of a 'confirmation threshold' (the other option amounts to 'anything goes'): we demand that, for a theory to be considered as 'confirmed', it must have *at least* the epistemic value of a trivial truth (scientists may decide to choose a higher threshold, of course). So, for *H* to be acceptable, $Vs(H,E)$ must be higher than *1*; in particular, for a theory *H* explaining the evidence, *H* will be acceptable *only* if $p(H) \geq p(E)^2$ (for 4.b). This has some nice consequences: it not only discards 'too weird' or 'ad hoc' theories (those that explain the evidence by adding innumerable unlikely hypotheses), but also entails that, the richer is the evidence, the more tolerant can we be with 'weird' hypotheses if they happen to explain it; i.e., the more surprising empirical facts we happen to find, the more willing we will be to consider as 'acceptable' theories with a low prior probability (for the threshold varies inversely with the *square* of the prior probability of the evidence), something which is consistent with the actual history of science. Furthermore, that the epistemic value of a 'tautological' proposition (e.g., mathematical theorems?) is higher than that of a contradiction, as function *Vs* entails, seems also to respect the idea that logical or mathematical 'facts' are scientifically valuable.

d) In Huber's, as in most other inductivist approaches, the only reason there seems to be to search for new empirical evidence is that perhaps the existing data have not yet happened to confirm or refute our hypotheses with certainty. They also seem to presuppose a scenario in which the hypotheses in competition are given, and one just acts by sheer curiosity looking for evidence that may lead one to know with certainty if those hypotheses are true or false, *but nothing demands that new hypotheses are devised*. Notice as well that the maximum epistemic value that Huber's function $f(H,E) = p(H,E) + p(\neg H,\neg E)$ can reach is 2, which is equal to $f(A,A)$ (i.e., the value of a hypothesis that were logically equivalent to the evidence). My definitions of empirical verisimilitude, instead, allow to see why science is an 'unended quest' (4.i): a given amount of evidence *E* can provide to any theory *H* a verisimilitude equal to $1/p(E)$ *at most*; hence, there will be situations in which the only way of increasing the epistemic value of our knowledge will be by *looking for more empirical data*. Analogously, the maximal epistemic value *H* can reach under any possible corpus of empirical evidence is $1/p(H)$, which means that there will be situations in which the only way of increasing epistemic value will be by *devising still stronger theories*.

e) Lastly, it seems reasonable to use the expected epistemic value function (4.j) when we consider that evidence *E* is provisional, not in the sense of being fallible, but

[10] This derives from the definition of $f(H,E)$ as $p(H,E) + p(\neg H,\neg E)$, and from the choice of β as the value of $f(Taut,E) = f(Contrad,E) = 1$.

in the sense of being *incomplete*, i.e., when we expect that new evidence is going to be added in the short run. In this case, let's compare the expected value of *f* and *Vs* under two different scenarios: one in which *H* perfectly explains the data (i.e., $H \vdash E$), and one in which *H* has been already verified by the data (i.e., when $E \vdash H$). In the first case, $E[Vs(H,E)] = 1/p(E)$, whereas $E[f(H,E)] = 1+ p(H)(p\neg E)$; we see that $E[Vs]$ gives the same value to all the theories explaining the evidence, compensating the loss of plausibility of very unlikely theories with their probable gain of verisimilitude if they happened to explain the still undiscovered evidence; $E[f]$, instead, still values more those theories with a higher prior probability. In the second case, $E[Vs(H,E)] = 1/p(H) = Vs(H,E)$, i.e., we don't expect that new evidence is going to change the value of *H*, for it is *already* confirmed. However, $E[f(H,E)] = p(E)(2- p(H))$, which means that the addition of new evidence (e.g., when we pass from *E* to *E&F*) will make *decrease* the expected value of *H*, which does not seem reasonable.

I don't want to conclude this section without insisting on the fact that, from the point of view defended here, any function of epistemic value must be seen as an *empirical hypothesis*, i.e., the hypotheses that real scientists actually evaluate their theories according to an epistemic utility function which is similar to the one proposed by the epistemologist's hypothesis. The theorems collected in (4) and in the five points developed above should be taken, hence, as partial attempts to empirically test Huber's and my definitions of epistemic value against some 'idealised facts' about how scientists really compare the values of their theories.

4. Back to the many senses of confirmation.

I come back to Huber's idea that the logic of confirmation is similar to the logic of purchasing a good: the question is not whether a scientific hypothesis, theory, model, law, etc., is 'known to be true with enough certainty', but whether it shows a panoply of virtues that justifies to take it into the pool of acceptable claims. 'Acceptance' has, then, an unavoidable and essentially pragmatic side. Having developed in the last decade a number of works that try to understand the epistemic decisions of science on the basis of *economic* models,[11] I cannot be but extremely sympathetic to this suggestion of Huber, but I must also point to some difficulties this move has for any attempt to connect the notion of 'confirmation' with that of 'acceptance' (at least in the sense that a hypothesis being 'confirmed' should be one primary reason in order to decide to it).

In the first place, 'acceptable' is ambiguous because it can refer to (at least) two different attitudes an individual scientist, but more significantly a scientific community, may have towards a theory: *H* can be 'acceptable' in the sense that the community *allows* that individual scientists accept *H* (for example, *H* can be amongst the theories proposed to solve a certain problem, and also belongs into the subset of those theories that, according to the available evidence, have a high chance of being 'right', and it is not considered illegitimate to defend that theory or any rival one that also might be right 'in the end'), or *H* can be 'acceptable' in the sense that the community *commands* its members to accept it (i.e., you will be sanctioned in some way if you do not respect in your own work the validity of that proposition, like, e.g., the facts

[11] See, esp., Zamora Bonilla (2002b), (2006a) and (2006b). See particularly Zamora Bonilla (2006a, 183) for a diagram which models the choice of a theory amongst the acceptable ones, in a way analogous to the microeconomic explanation of the consumers' choice.

organised in the periodic table, or the laws of thermodynamics, etc.). The idea of a 'pool' of acceptable theories seems to reflect better the first sense than the second, but usually, it is to the *second* sense to what scientist and philosophers seem to refer when they say that a theory or hypothesis has been 'confirmed'. Actually, something very similar happens in the example employed by Huber (the purchase of a bottle of wine): you do not actually buy *all* 'acceptable' bottles, but only those your budget constraint (and your desire to buy other things) allows. So, 'confirmation' seems to indicate the property a theory has to have in order for the acceptance of that theory being *compulsory*. Of course, this property can simply consist in being *the best one* amongst the 'acceptable' (in the first sense) theories, but this would seem to require a *double* demarcation: one demarcation mark that separates 'acceptable' from 'not acceptable' theories *in the first sense* (so that it is legitimate for two scientists to accept different theories, if both are acceptable in this sense, but no theory exists whose acceptance is compulsory), and one demarcation level that separates 'acceptable theories in the first sense' from 'acceptable theories in the second sense' (i.e., theories that, if it happens to exist at least one of it, then no 'plurality' is admitted, and the best one *must* be accepted). Furthermore, it could not be the case that the only criterion is 'accept the best available theory', for it might happen that this theory were very bad indeed: Huber is right, then, in that some demarcation line in the classical inductivist sense of 'absolute confirmation' seems to be needed; hence, in some cases 'suspension of belief' should be considered epistemically better than the adoption of one amongst the available theories. This reinforces the choice made in point *c* above of $Vs = 1$ (i.e., the verisimilitude of a tautology, which can be interpreted as an empty answer) as a lowest limit of the demarcation criterion for acceptable-theories-in-the-first-sense.

The point made in the previous paragraph is consequential enough to summarise it: when asking 'why are good theories good', we are asking several different questions: one is 'what makes a theory *better* than another'; functions like Huber's Bayesian confirmation or my family of verisimilitude functions provide a partial answer to that question. Another two different questions are, 'what makes a theory so good that it is *legitimate* to accept it?' (though one could still legitimately accept other alternative theories) and 'what makes a theory so good that it is *compulsory* to accept it?' (so that it is taken as 'certified knowledge'). Formal studies of confirmation have hardly touched until now the analysis of what is what makes these three questions different, and there is a good reason for that, as we shall see immediately.

In the second place, purely formalist definitions of epistemic values, like those of Huber's or those presented in the past section, are unable to illuminate by themselves the question of *which* demarcation criteria are 'best'. From the point of view of those functions, any point above 1 serves equally as any other. It is *the rest of the preferences of the scientists (epistemic or non epistemic)* what must be taken into account in order to determine what choice of demarcation levels is optimal. From a rational choice perspective, if we call *X* the set of all possible mutually exclusive sets of *consequences* that the choice of a demarcation level β will have for scientist *i*, then, the optimal choice for that scientist will correspond to:

(7) $\beta^*_i = argmax \sum_{x \in X}(p_i(x,\beta)\ u_i(x)),$

where $p_i(x,\beta)$ is obviously the probability with which *i* judges that the choice of β will lead to consequences *x*, and $u_i(x)$ is the utility that *i* would experiment under *x*. This depends on *all* the things the scientist values, as long as these things are affected by the choice of a demarcation level; for example, since this choice affects the probabilities she has of finding a theory that passes the mark, and hence of becoming the discoverer of a theory that her colleagues can (or, still better, *must*) accept, the pursuit of recognition will undoubtedly be a strong factor influencing the preference of scientists for one or another threshold.[12] Furthermore, we must take into account the reasonable possibility of different scientists having different preferred demarcation thresholds, what would make its *collective* choice considerably more complicated, and depending on the institutional ways of making collective decisions that each scientific community may have established. All this means that *pragmatics is unavoidably intermingled within the choice of a standard of confirmation*, so that this choice cannot be a matter of 'logic' alone (no more than what the 'logic of business' or –to quote Popper's famous example– the 'logic of chess' are, though I suspect that the 'logic of confirmation' is more similar to the first example than to the second, because of its strong dependence on casual pragmatic considerations and of its not being reducible to a fixed and universal set of norms).

In third place, one reason why the two 'conflicting intuitions' behind the confirmation notion are in conflict is probably because they are related to two different types of scientific claims, which are tested through relatively different types of processes. Basically it is the difference between, so to say, more 'observational facts' and more 'speculative theories'. Personally, I don't think that an absolutely clear distinction can be drawn between both types of items; after all, what makes of something a 'datum' is that it is empirical in some sense, and that it can be used to test some theory, and what makes of something a 'theory' is that it can be used to explain or understand some 'facts'; so, something can be a datum in relation to something, and a theory in relation to something else. But, in spite of that, some claims are more apt to be tested by means of inductive methods (including eliminative induction), and some claims by hypothetico-deductive, or abductive methods. *The main difference regarding the confirmation question is whether we think that new evidence can 'disconfirm' the claim*. Usually, we do not reasonably expect that propositions that have been 'established' by means of inductive reasoning can be falsified by new observations (or new types of observations), though we may be open to reconsidering the range of validity of those propositions. Other claims, however, are still considerably hypothetical *even if they have been successful* in explaining and predicting the relevant data, and we acknowledge that new empirical discoveries may completely refute those hypotheses (though these may persist as 'good approximations' in some cases). My impression is that we hardly would say of most of these second type of theories that they have been 'confirmed', even if they are currently accepted. The limits, I insist, are vague, and perhaps in some cases the accumulation of different abductive arguments is as demonstrative of the truth of a claim as inductive arguments can be, but in general, 'theories' that try to explain many different things will tend to have a relatively low posterior probability, even if 'corroborated' by *E* in Popper's sense (see fig. 3.a), whereas 'laws' will have a high

[12] Cf. Zamora Bonilla (2002b).

degree of inductive support (fig. 3.b.). As the figure allows to see, the 'theory' in the first case can even have a much higher verisimilitude than the 'law' in the second one (even if *p(H',E)* is much bigger than *p(H,E)*, so that it is *H'* the one we would tend to say that has been 'confirmed' by *E*).

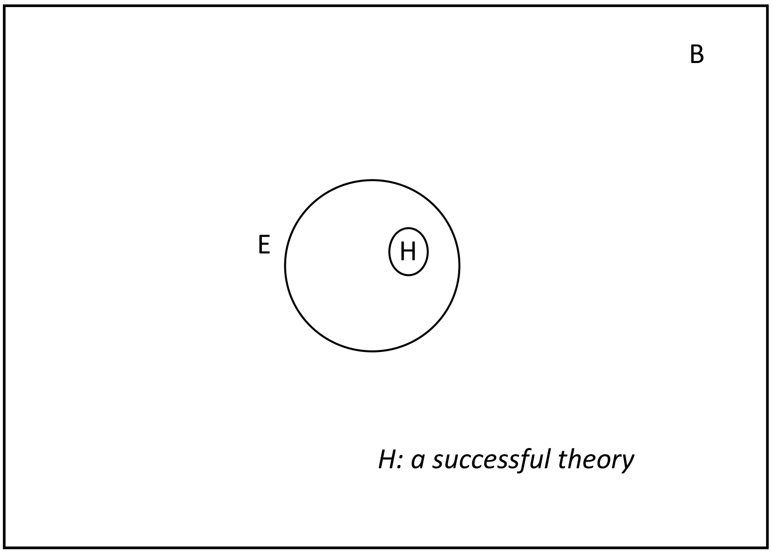

Figure 3.a

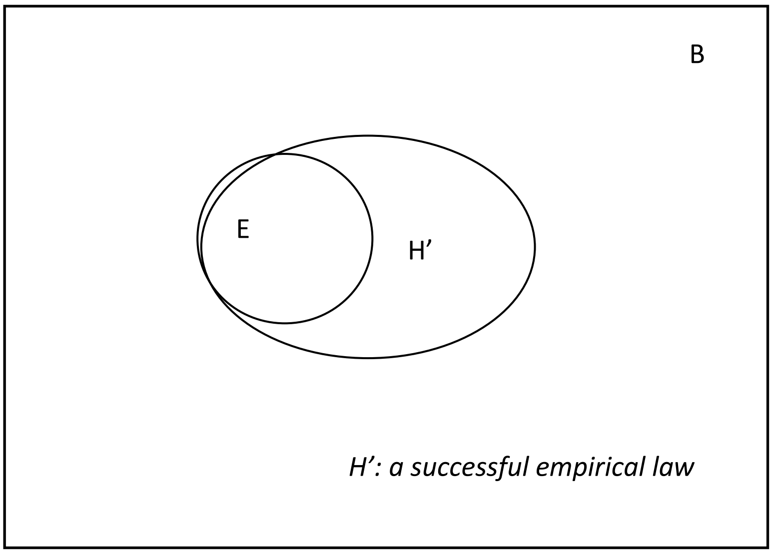

Figure 3.b

This question can connected with a recent discussion by Kyle Stanford (2006) about whether the 'pessimistic induction' argument undermines scientific realism.[13] Stanford modifies this famous argument by pointing to the fact that it is not only that many past successful theories posited non-existent things, properties or mechanisms, but that the proponents of those theories were *unable to imagine or conceive* the theories that, years or centuries later, have replaced them. This means that, if those scientists thought that their own theories had been 'confirmed' in the sense that their data eliminate the possibility of other rival theories being true, or made extremely unlikely any other rival theories, they were utterly wrong! Stanford is critical, hence, of the 'inference to the best explanation' argument: the best explanation we have may actually be very unlikely (for, example, in fig. 3.a, $p(\neg H,E)$ is much smaller than $p(H,E)$). Still more precisely, his reasoning is that the scientists' (or the realist philosophers') *subjective* judgments of probability may fool them in making them believe that, because H has managed to explain and predict so many surprising findings, $p(H,E)$ is very large, whereas the truth is that there can be a lot of unconceived theories that also explain E equally well, so that the *objective* probability of H is really very low. It is important for our discussion the fact that Stanford explicitly denies that this puts a difficulty for the acceptability of the kinds of propositions whose truth is established by more inductive procedures (Stanford (2006, 32ff.)), and hence, Stanford's discussion would be coherent with my previous point illustrated by fig. 3: many successful theories need not to be taken as 'confirmed' in spite of their success.

5. Confirmation, naturalism, and Bayesian tinkering.

5.1. For a formal naturalist epistemology.

So, why are good theories good? In contrast to the common image of the goals of analytical epistemology I offered in the introduction, I want to defend in this last section a *naturalist* approach. According to this view, the task of the philosopher of science is not to answer that fundamental question as if it were part of a logico-mathematical science, but as an *empirical scientific problem*. The empirical fact in question is that there is a lot of populations of real scientists out there, working with many things, some of them we can call (just as a convenient label) 'theories', and those scientists claim (or behave as if they thought) that some theories are 'better' than others, accept some of them when they are 'good enough', and very often, they even punish their colleagues when these fail to accept some theories that are 'extremely good'. In principle, *the 'problem of theory evaluation' is not a 'philosophical' problem, but a problem for the communities of flesh-and-bone scientists*.[14] The philosopher enters here in an analogous way as the scientist does when she approaches her objects of study: trying to understand, by means of theories and models, what is what is going on within that parcel of the world, and testing his theories against the data attained by the observation of that parcel. This 'naturalism' does not entail (as some may have concluded) that there is no role for a *formal epistemology*, no more than physics being a natural science precludes its being strongly mathematized, but it simply reminds us

[13] For the 'pessimistic induction' argument (according to which most successful theories in the past failed to make reference to real entities, and so we can infer that the same will happen with current theories), see Laudan (1981).

[14] Of course, we might reason in the same way *à propos* of 'everyday knowledge', but I'm circumscribing here to the case of science.

that the abstract objects we construct and study within the field of formal epistemology are not to be understood as systems whose validity is to be determined just by logico-mathematical arguments, but are essentially *tentative models of empirical phenomena*, phenomena that consist in the actual practices of theory evaluation real scientists carry out. This does also not entail that formal epistemology must be 'descriptive' and not 'normative': it can be normative, but not because the philosopher has a privileged access to the right essence of the epistemic values, and constructs his formal models by deriving them from those essences, but simply because the empirical subject matter that formal philosophy of science examines is normative in itself, i.e., it is constituted by the normative *attitudes* of scientific communities (what they consider 'good', 'legitimate', or 'compulsory'), and because it is conceivable that the philosopher's analysis leads to conclusions of the type 'this community is not behaving according to the norms it preaches', or 'the cognitive practice of this community is detrimental to such and such other values'. In this sense, formal epistemology can be exactly as empirical and as normative as economic theory is (and suffer from the same type of methodological problems because of that).[15]

Very often we read thinks like that "formal models of philosophically interesting concepts are tested largely by seeing how well they match intuitive judgments in particular cases".[16] The suggestion made in this paper is, instead, that the most appropriate way of assessing a theory of confirmation is by showing its coherence with *real* practices of theory evaluation, which, obviously, are based themselves on scientists' 'intuitive judgments', *but which have also been modulated by a long history of scrupulous trial and error*. Before the formal analytical epistemologists that have kept reading this paper till this point stand up and go, I shall reassure them by pointing to the fact that this does not necessarily entail that epistemologists become just another species in the fauna of 'historico-sociological case studies': in the same way as there is room in physics or in economics for the 'empirical' (or 'experimental') and for the 'theoretical', the same may happen in epistemology: you can earn your life as a *theoretical* formal epistemologist. But of course, theoretical formal epistemology is not less empirical than stuck-to-the-ground history and sociology of science: it simply accesses the empirical phenomena in a more indirect, 'stylised' way. The idea is just to complement (if not to replace) methodological attitudes like the one exemplified by the quotation at the beginning of this paragraph, with the attitude of supposing that our analysis is a toy model of an imaginary scientific community, and use our general knowledge of how scientific communities work in order to see if our model depicts accurately enough the *relevant aspects* of those real practices. For example, is the way scientists evaluate their theories more similar to what might be derived from the assumption that they are trying to maximise something like classical Bayesian functions (like $p(H,E) - p(H)$), or to Huber's f, or to the functions of verisimilitude described above? Just by way of example, besides the properties indicated in (4), Zamora Bonilla (2002b) shows that the combination of sophisticated verisimilitude and the assumption that when new empirical discoveries are awaited scientists evaluate their theories according to expected verisimilitude,

[15] See, e.g., Yuengert (2004) for the problems about the connection between the positive and the normative in economics, and Kahn *et al*. (1996) for one early example of 'economics-oriented' methodology.
[16] Christensen (1999, 460).

leads to a ‘rational reconstruction’ of the ‘stylized facts’ on the evolution of scientific research programmes explicated by Lakatos.

5.2. Nagelian reduction and epistemic values.

I shall end by illustrating this naturalist attitude by means of two further examples, in which the formal and the empirical mix together. Both are based on the fact that the probabilities employed in our models are *subjective*, and hence, subjected to the possibility of what we might call ‘Bayesian tinkering’: the strategic handling of prior probabilities. The first example refers to the evaluation of theories in those cases where one theory is reduced to another; the second example refers to the explanatoriness of theories as a cognitive virtue. Once upon a time, positivist philosophers saw intertheoretical reduction as the fundamental road to scientific progress; the anti-positivist movement of the second part of the 20th century led a majority of philosophers to abandon, not only the idea that science progresses through reduction to some theories to others, but in many cases even the belief that intertheoretical reduction is logically possible at all. For my argument’s purposes, there is no need now to enter any deep discussion on ‘reductionism’, but just to point out that, as a matter of fact, scientists try sometimes to reduce a theory to another, and often they succeed; in the minimalist theory of reduction of Nagel (1961), ‘reduction’ consists simply in somehow being able of showing that the laws of the reduced theory (T) can be derived from the laws of the reducing theory (T'), together with some auxiliary assumptions and bridge laws (see Dizadji-Bahmani *et al.* (2010 and forthcoming) for a defence of the logical validity and historical reality of ‘Nagelian’ reduction). The question I want to concentrate on is, why do scientists *bother* in trying to show that such a mathematical relation exists between two theories? The answer will be, of course, that this strategy can be fruitful in the struggle for increasing the epistemic value of the theories. Dizadji-Bahmani *et al*. (2010) show that this can be so in the case of simple Bayesian confirmation, i.e., the conditional probability of both the reduced and the reducing theories can be higher after the reduction relation has been proved; I will employ a similar argument to theirs’ for showing that the same is true in the case of verisimilitude.

If T can be reduced to T', thanks to auxiliary assumptions A, this means that scientists have been able to establish the truth of the proposition $(T' \& A) \rightarrow T$ (in what follows, I will simplify by considering A as part of the axioms of T', so that the relevant relation is just $T' \rightarrow T$). The existing empirical evidence is E; assume for simplicity that H successfully explained E (it is more realistic to assume that it didn’t, and that it even has some ‘anomalies’, so that sophisticated verisimilitude functions need to be used; but I will limit my example to the simplest case). *Before* a mathematical proof existed of the proposition $T' \rightarrow T$, the situation from the point of view of scientists is as depicted in fig. 4.a: no logical connection exists between T and T' nor between T and E. The relative sizes of T and T' are also significant: T may have a relatively low prior probability (it consists in a set of assumptions accepted just because they successfully explain the evidence), whereas T' may have a higher prior probability, because its axioms are more coherent with the background knowledge, for example (I do not claim that this is a necessary condition of reduction, just an illustrative possibility). In this case, the verisimilitude of T' is low in spite of its being perhaps plausible *a priori*, and the verisimilitude of T is higher, but probably not too much, because of its low prior probability. The situation changes when it is proved that *T'* entails *T* (fig. 4.b): now, scientists know that *T'* also entails *E*, and the verisimilitude of *T'* will grow accordingly (not only because the increase in *p(T,E)*, but because of the decrease in *p(TvE)*); furthermore, since *T* logically follows from *T'*, the prior probability of *T* must be at least as big as that of *T'*, and if this one was based on its coherence with background assumptions, it is more reasonable that the adjustment leads to the increase of *p(T)*. than to a decrease of *p(T')*.

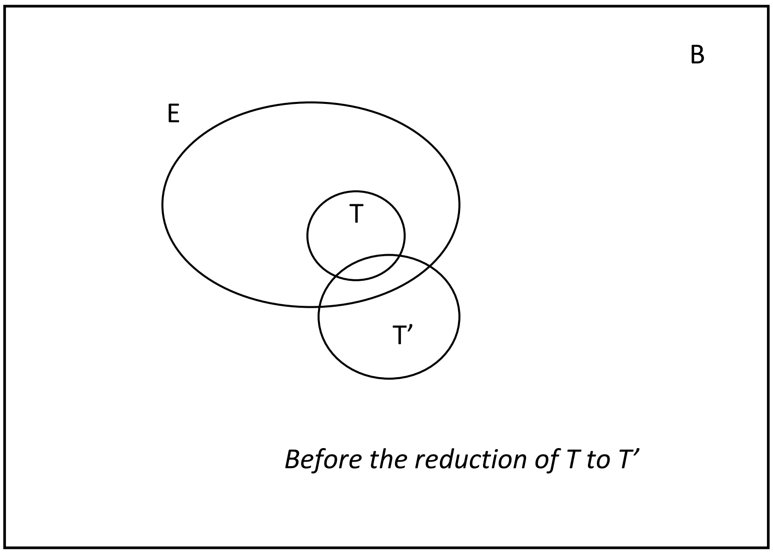

Figure 4.a

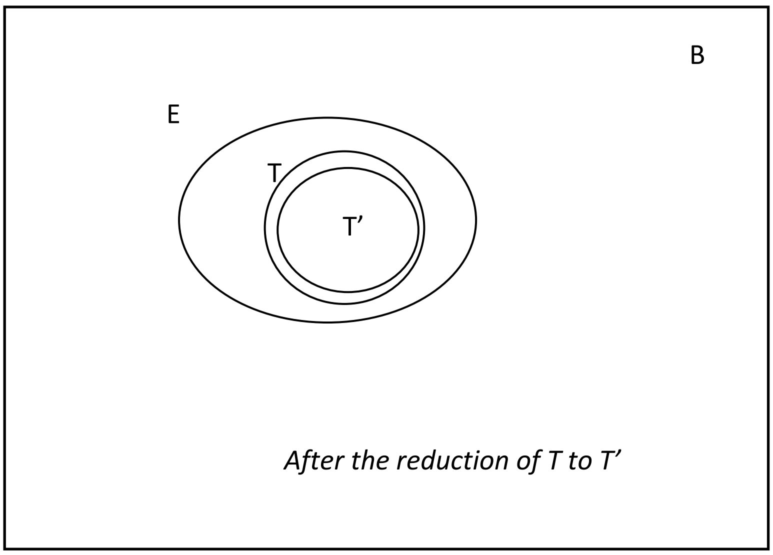

Figure 4.b

So, showing that a theory is reducible to another one can increase the verisimilitude of both theories, and certainly the verisimilitude of the reducing theory. Reduction is a way of 'importing' the empirical success of the reduced theory to the reducing theory, and sometimes also a way of improving the plausibility of the reduced theory's assumptions.

5.3. Explanatoriness and confirmation.

Let me come back to a couple of important points of my discussion of Huber's theory of confirmation:

a) as Huber insisted, the 'real' logic of confirmation includes not only 'inductivist' values, but also 'hypothetico-deductivist' virtues; however, many hypothetico-deductivist philosophers usually consider that a good theory is not only one that 'explains' the available evidence in the sense that the evidence can be *logically derived* from the theory, but also in the sense that it is a *good* explanation; i.e., the explanatoriness of a theory is not only a matter of *how much* it explains, but also of *how well* it does it, and this has to do not only with the logical properties of the theory, but also with its other cognitive and pragmatic features;[17]

b) as a conclusion of the previous point, if we identify the 'logic of confirmation' with the 'logic of acceptability', and admit that acceptability depends on a number of virtues amongst which there may be some *trade-offs*, we must be open to the possibility that scientists prefer to *accept* a theory that they think explains *better*, even if a rival theory explains *more*.

Point *b* is, in a sense, the old Kuhnian story of senior scientists refusing to accept a new theory in spite of recognising that it 'solves' the anomalies of the old 'paradigm'. One reason of this reluctance may be that, for the defenders of the old theory, the new one does not provide 'good' explanations, because these are based on principles and techniques that seem 'absurd' to them. Be that as it may be, the fact is that scientists take into account something more than the relations of logical entailment between theories and data in order to decide what theories to accept. This is in conflict, of course, with the Hempelian, positivist account of nomological explanation, but is more coherent with other approaches that assume that cognitive and pragmatic virtues like 'understanding' are necessary in an explication of explanation. One recent example of these other approaches is de Regt (2009), according to whom a theory *H* explains the facts *F* for the scientific community *C* if and only if *F* can be derived from *H* by *C*, *and* the members of *C understand H*, i.e., if *H* is 'intelligible' for them.[18] An obvious way to take into account this 'intelligibility' requirement in the (quasi)-Bayesian functions we are examining in this paper is by connecting it to the measures of prior probability of the relevant propositions, and to the relation these have with the 'background assumptions', by means of a conjectural principle like the following:

[17] Cf. the following recent quotation by a biochemist: "We could encourage a more realistic (…) understanding of evolution by admitting that what we believe in is not a unified meta-theory but a versatile and well-stocked *explanatory toolkit*" (Doolittle, 2010, 455, original's italics).

[18] Donato and Zamora Bonilla (2009) refer to 'enlightening' as a similar cognitive virtue of models or theories, identified with the *easiness* (or the increment thereof) with which these models or theories allow to navigate the network of inferential links to which they belong.

(8) *Coeteris paribus*, if *X* is easier to understand than *Y*, then $p(Y) < p(X)$.

Of course, this principle is extremely vacuous, since the types of things that enter into the *coeteris paribus* clause are so numerous and varied that it is unlikely that a couple of real theories happen to be exactly equal in everything else besides how easy to understand for somebody. This, however, is not a disproof of the claim that the intelligibility of a proposition (that may depend, amongst other things, from its coherence with the 'background knowledge') affects the *subjective* prior probability with which it is assessed by scientists. An argument in favour of this claim is the fact that scientists tend to employ an array of *rhetorical* arguments (i.e., arguments not directly consisting in confronting the theories with the data) to persuade their colleagues about the plausibility of their theories. My hypothesis that scientists try to maximise a function like (3) (or its more sophisticated versions) 'predicts' that this type of arguments, as attempts to increase $p(H)$, will be used at least when *H* is empirically successful (i.e., when it entails *E*, or a big part of the laws contained in *E*), for this increases $p(H,E)$ without enlarging $p(HvE)$. It also 'predicts' that similar rhetorical arguments, but in the opposite direction, will be employed to show that $p(E)$ is *low* (for this, when *H* is successful, increases $p(E,H)$ and reduces $p(HvE)$). I.e., scientists will try to show that their hypotheses are very plausible *a priori*, and that the facts explained by those theories are very implausible.[19]

If some readers reacted to the 'naturalism' defended at the beginning of this section with the feeling that I was trying to remove the 'philosophy' from the philosophy of science (what was obviously not my aim), my last argument shows that philosophy inescapably enters science by the back door, for in the end, it is 'philosophical' (as opposed to 'empirical') arguments what scientists need to use to tinker with the plausibility judgments of their colleagues (and the general public). It is not strange, then, that plenty of philosophical argumentation takes place during scientific revolutions.

ACKNOWLEDGEMENTS

Financial support from Spanish Government's research projects FFI2008-03607/FISO and FFI2008-01580/FISO is acknowledged.

REFERENCES

Bovens, L., & S. Hartmann. (2005) "Why there cannot be a single probabilistic measure of coherence?", *Synthèse*, 63:361-374.

Christensen, D. (1999), "Measuring confirmation", *The Journal of Philosophy*, 96:437-461.

Dizadji-Bahmani, F., R. Frigg, & S. Hartmann (2010), "Confirmation and Reduction: A Bayesian Account", *Synthèse*, online.

Dizadji-Bahmani, F., R. Frigg, & S. Hartmann (forthcoming), "Who's afraid of Nagelian reduction?", *Erkenntnis*, forthcoming.

Donato, X., and J.P. Zamora Bonilla (2009). "Credibility, idealisation, and model building", *Erkenntnis*, 70:101-118.

Doolittle, W.F. (2010). "The attempt on the life of the Tree of Life: science, philosophy and politics", *Biology and Philosophy*, 25:455-473.

Hempel, C. G., & Oppenheim, P. (1945). A definition of "degree of confirmation". *Philosophy of Science*, 12:98–115.

[19] By the way, this can be add as another pragmatical criterion to distinguish 'theories' from 'laws', to the one I suggested in section 4.

Huber, F. (2008a). “Assessing theories, Bayes style”, *Synthèse*, 161:89–118.

Huber, F. (2008b). “Hempel’s logic of confirmation”, *Philosophical Studies*, 139:181-189.

Kahn, J., S.E. Landsburg & A.C.Stockman (1996), “The positive economics of methodology”, *The Journal of Economic Theory*, 68:64-76.

Kuipers, T.A.F. (1992), “Naive and refined truth approximation”, *Synthése*, 93:299-341.

Laudan, L. (1981). “A Confutation of Convergent Realism”. *Philosophy of Science* 48:19-49.

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

Nagel, E. (1961). *The Structure of Science*. London: Routledge and Keagan Paul.

Olsson, E.J. (2002). “What is the problem of coherence and truth?”, *The Journal of Philosophy*, 99:246-278.

de Regt, H. (2009). “The epistemic value of understanding”, *Philosophy of Science*, 76:585-597.

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

Yuengert, A. (2004). *The Boundaries of Technique. Ordering Positive and Normative Concerns in Economic Research*. Lanham. Lexington Books.

Zamora Bonilla, J. P. (1996). “Verisimilitude, Structuralism and Scientific Progress”, *Erkenntnis*, 44:25-47.

Zamora Bonilla, J. P. (2000). “Truthlikeness, Rationality and Scientific Method”, *Synthèse*, 122:321-35.

Zamora Bonilla, J. P. (2002a). “Scientific Inference and the Pursuit of Fame: A Contractarian Approach”, *Philosophy of Science*, 69:300-23.

Zamora Bonilla, J. P. (2002b). “Verisimilitude and the Dynamics of Scientific Research Programmes”, *Journal for General Philosophy of Science*, 33:349-68.

Zamora Bonilla, J. P. (2006a). “Science Studies and the Theory of Games”, *Perspectives on Science*, 14:639-71.

Zamora Bonilla, J. P. (2006b). “Rhetoric, Induction, and the Free Speech Dilemma”, *Philosophy of Science*, 73:175-93.