The Plausibility-Informativeness Theory*

Franz Huber, California Institute of Technology
penultimate version: please cite the paper in *New Waves in Epistemology*

Contents

* This is a precursor of "Assessing Theories, Bayes Style" (to appear in *Synthese*).

Abstract

The problem adressed in this paper is "the main epistemic problem concerning science", viz. "the explication of how we compare and evaluate theories [...] in the light of the available evidence" (van Fraassen 1983, 27).

We first present the general plausibility-informativeness theory of theory evaluation. In a nutshell, the message is (1) that there are two epistemic values a theory should exhibit: truth and informativeness – measured respectively by a truth indicator and a strength indicator; (2) that these two values are conflicting in the sense that the former is a decreasing and the latter an increasing function of the logical strength of the theory to be evaluated; and (3) that in evaluating a given theory by the available data one should weigh between these two conflicting aspects in such a way that any surplus in informativeness succeeds, if the difference in plausibility is small enough.

Particular accounts of this general theory arise by inserting particular strength and truth indicators. The theory is spelt out for the Bayesian paradigm; it is then compared with incremental Bayesian confirmation theory. The first part closes by discussing a few epistemic problems in the philosophy of science in the light of the present approach. In particular, it is briefly indicated how the present account gives rise to a new analysis of Hempel's conditions of adequacy for any relation of confirmation (Hempel 1945), differing from the one Carnap gave in §87 of his (1962).

The second part discusses the question of justification any theory of theory evaluation has to face: Why should one stick to theories with high values rather than to any other theories? The answer given by the Bayesian version of the account presented in the first part is that one should accept theories given high values, because, in the medium run, theory evaluation almost surely takes one to the most informative among all true theories when presented separating data. The comparison between the present account and incremental Bayesian confirmation theory is continued.

1 The Problem

The problem adressed in this paper is this:

> the main epistemic problem concerning science ... is the explication of how we compare and evaluate theories ... in the light of the available evidence ... (van Fraassen 1983, 27)

In other and more mundane words, the question is: What is a good theory, and when is one theory better than another theory, given these data and those background assumptions. Let us call this the problem of a theory of theory evaluation. Its quantitative version can be put as follows:

- One is given a hypothesis or theory H, a set of data – the evidence – E, and some background information B.
- The question is: How good is H in light of E and B? I.e., what is the *value* of hypothesis H in view of evidence E and background information B?
- An answer to this quantitative question consists in the definition of a (set $\mathcal{A}$ of) function(s) a such that (for each a in $\mathcal{A}$:) $a\,(H, E, B)$ measures the value of H in view of E and B, i.e. how good H is in light of E and B.

Given this formulation of our problem, a theory of theory evaluation need not account for the way in which scientists arrive at their theories nor how they (are to) gather evidence nor what they may assume as background information. Furthermore, the purpose of this evaluation is that we *accept* those theories (among the ones we can or have to choose from) which score highest in the evaluation relative to the available data (as discussed in more detail below, the term 'accept' is not used in the sense of believe or hold to be true). This makes clear that a proper treatment of the whole problem not only *explicates* how we should evaluate theories in the light of the available evidence (sections 2-6); a proper treatment also *justifies* this normative theory of theory evaluation by answering the question why we should accept those theories that score highest (sections 7-8).

2 Theory, Evidence, and Background Information

In order for the above characterisation to be precise one has to make clear what is meant by theory, evidence, and background information. In what follows it is assumed that for every hypothesis or theory H, every piece of evidence E, and every body of background information B there exist finite axiomatizations (in a

first-order language including identity and function symbols) A_H, A_E, and A_B, respectively, which formulate H, E, and B, respectively.

In general, not all finite sets of statements are formulations of a piece of evidence or a scientific theory. Scientific theories, for instance, do not speak about particular objects of their domain of investigation, but express general regularities or patterns. Data, on the other hand, only speak about finitely many objects of the relevant domain – we are damned *qua* humans to be able to observe only finitely many objects.

However, for the general theory outlined below (and its Bayesian version) it suffices that these be finitely axiomatizable. As theory evaluation turns out to be closed under equivalence transformations, H, E, and B can and will be identified with one of their formulations A_H, A_E, and A_B, respectively.

3 Conflicting Concepts of Confirmation

Though some take theory evaluation to be *the* epistemic problem in philosophy of science, there is no established branch adressing exactly this problem. What comes closest is what is usually called confirmation theory. So let us briefly look at confirmation theory, and see what insights we can get from there concerning our problem.

Confirmation has been a hot topic in the philosophy of science for more than 60 years now, starting with such classics as Carl Gustav Hempel's "Studies in the Logic of Confirmation" (1945) and Rudolf Carnap's work on Inductive Logic and Probability (Carnap 1962). Roughly speaking, the last century has seen two main approaches to confirmation:

- On the one hand, there is the qualitative theory of *Hypothetico-Deductivism* HD (associated with Karl R. Popper), according to which a hypothesis H is confirmed by evidence E relative to background information B iff the conjunction of H and B logically implies E in some suitable way – the latter depending on the version of HD under consideration.

- On the other hand, there is the quantitative theory of probabilistic *Inductive Logic* IL (associated with Rudolf Carnap), according to which H is confirmed by E relative to B to degree r iff the probability of H given E and B is greater than or equal to r. The corresponding qualitative notion of confirmation is that E "absolutely" IL-confirms H relative to B iff the probabiliy of H given E and B is greater than some fixed value r in $[.5, 1)$.[1]

[1]This is *not* the way Carnap defined qualitative IL-confirmation in chapter VII of his (1962).

However, despite great efforts there is still no generally accepted definition of (degree of) confirmation. One reason for this is that there are at least two conflicting concepts of confirmation: A concept of confirmation aiming at *informative* theories; and a concept of confirmation aiming at *plausible* or *true* theories. These two concepts of confirmation are conflicting in the sense that the former is an increasing and the latter a decreasing function of the logical strength of the theory to be evaluated.

Definition 1 *A relation $|\sim \subseteq \mathcal{L}\times\mathcal{L}\times\mathcal{L}$ on a language (set of propositional or first-order sentences closed under negation and conjunction) $\mathcal{L}$ is an* informativeness relation on $\mathcal{L}$ *iff for all sentences H, H', E, B in $\mathcal{L}$:*

$$B, E \mid\sim H, \quad H' \vdash H \quad \Rightarrow \quad B, E \mid\sim H'.$$

$|\sim \subseteq \mathcal{L} \times \mathcal{L} \times \mathcal{L}$ *is a* plausibility relation *on $\mathcal{L}$ iff for all H, H', E, B in $\mathcal{L}$:*

$$B, E \mid\sim H, \quad H \vdash H' \quad \Rightarrow \quad B, E \mid\sim H',$$

where $\vdash \subseteq \wp(\mathcal{L}) \times \mathcal{L}$ is the classical deducibility relation (and singletons of formulae are identified with the formula they contain).

The idea is that a sentence or proposition is more informative, the more possibilities it excludes. Hence, the logically stronger a sentence, the more informative it is. On the other hand, a sentence is more plausible, the fewer possibilities it excludes, i.e. the more possibilities it includes. Hence, the logically weaker a sentence, the more plausible it is. The qualitative counterparts of these two comparative principles are the two defining clauses above: If H informs about E given B, then so does any logically stronger sentence H'. Similarly, if H is plausible given E and B, then so is any logically weaker sentence H'.

According to HD, E HD-confirms H relative to B iff the conjunction of H and B logically implies E. Hence, if E HD-confirms H relative to B, and if H' logically implies H, then E HD-confirms H' relative to B. So HD-confirmation is an informativeness relation. According to IL, E absolutely IL-confirms H relative to B iff $\Pr(H \mid E, B) > r$, for some value r in $[.5, 1)$. Hence, if E absolutely IL-confirms H relative to B and H logically implies H', then E absolutely IL-confirms H' relative to B. So absolute IL-confirmation is a plausibility relation.

There he required that the probability of H given E and B be greater than that of H given B in order for E to qualitatively IL-confirm H relative to B. Nevertheless, the above is the natural qualitative counterpart for the quantitative notion of the degree of *absolute* confirmation, i.e. $\Pr(H \mid E \wedge B)$. The reason is that later on the difference between $\Pr(H \mid E \wedge B)$ and $\Pr(H \mid B)$ – in whatever way it is measured (Fitelson 1999) – was taken as the degree of *incremental* confirmation, and Carnap's proposal is the natural qualitative counterpart of this notion of incremental confirmation.

The epistemic values behind these two concepts are, of course, *informativeness* on the one hand and *truth* or *plausibility* on the other. First, we want to know what is going on "out there", and hence we aim at true theories – more precisely, at theories that are true in the world we are in. Second, we aim at informative theories – more precisely, at theories that inform us about the world we are in. But usually we do not know which world we are in. All we have are some data. So we base our evaluation of the theory we are concerned with on the plausibility that theory is true in the actual world given that the actual world makes the data true; and on how much the theory informs us about the actual world given that the actual world makes the data true.

Turning back to the question we started from – What is a good theory? – we can say the following: According to HD, a good theory is informative, whereas IL says good theories are true. Putting together the insights of last century's confirmation theory, the answer of the new millenium is this: *A good theory is both informative and true*. Consequently, we should make these aims explicit in the evaluation of theories.

4 Indicating Strength and Indicating Truth

Given evidence E and background information B, a hypothesis H should be both as informative and as plausible as possible. A *strength indicator* s measures how much H informs us about E given B; a *truth indicator* t measures how plausible it is that H is true in view of E and B. Of course, not any function will do.

Definition 2 *A possibly partial function $f : \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is a* truth indicator *on the language $\mathcal{L}$ iff for all sentences H, H', E, B in $\mathcal{L}$ for which $f(H, E, B)$ and $f(H', E, B)$ are defined:*

$$B, E \vdash H \to H' \quad \Rightarrow \quad f(H, E, B) \leq f\left(H', E, B\right).$$

Observation 1 *Let f be a truth indicator on $\mathcal{L}$. Then we have for all H, H', E, B in $\mathcal{L}$ for which $f(H, E, B)$, $f(\neg H, E, B)$, and $f(H', E, B)$ are defined:*

$$B, E \vdash H \quad \Rightarrow \quad f(\neg H, E, B) \leq f\left(H', E, B\right) \leq f(H, E, B).$$

The range of f is taken to be the set of real numbers $\Re$. The defining clause takes care of the fact that the set of possibilities (possible worlds, models) falsifying a hypothesis H is a subset of the set of possibilities falsifying any hypothesis that logically implies H, where the set of possibilities is restricted to those not already ruled out by the data and the background information. It follows that logically

equivalent hypotheses always have the same plausibility (f-value), provided the relevant expressions are defined.

The observation states that we cannot demand more – as far as only our aim of arriving at true theories is concerned – than that the evidence and the background information our evaluation is based on *guarantee* (in the sense of logical implication) that the theory to be evaluated is true. Similarly, a theory cannot do worse – as far as only our aim at arriving true theories is concerned – than that the conjunction of the data and the background information guarantees that the theory is false.

Definition 3 *A possibly partial function $f : \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is an* evidence based strength indicator on the language $\mathcal{L}$ *iff for all sentences H, H', E, B in $\mathcal{L}$ for which $f(H, E, B)$ and $f(H', E, B)$ are defined:*

$$B, \neg E \vdash H \to H' \quad \Rightarrow \quad f\left(H', E, B\right) \leq f(H, E, B).$$

f is an evidence neglecting strength indicator on the language $\mathcal{L}$ *iff for all sentences H, H', E, B in $\mathcal{L}$ for which $f(H, E, B)$ and $f(H', E, B)$ are defined:*

$$B \vdash H \to H' \quad \Rightarrow \quad f\left(H', E, B\right) \leq f(H, E, B).$$

f is a strength indicator on the language $\mathcal{L}$ *iff there is an evidence based strength indicator f_1, an evidence neglecting strength indicator f_2, and a possibly partial function $g : \Re \times \Re \to \Re$ such that (i) $f(H, E, B)$ is defined and $f(H, E, B) = g(f_1(H, E, B), f_2(H, E, B))$ for all H, E, B in $\mathcal{L}$ for which $f_1(H, E, B)$ and $f_2(H, E, B)$ are defined, and (ii) g is non-decreasing in both and increasing in at least one of its arguments f_1 and f_2.*

Observation 2 *Let f be an evidence based strength indicator on $\mathcal{L}$. Then we have for all H, H', E, B in $\mathcal{L}$ for which $f(H, E, B)$, $f(\neg H, E, B)$, and $f(H', E, B)$ are defined:*

$$B, \neg E \vdash H \quad \Rightarrow \quad f(H, E, B) \leq f\left(H', E, B\right) \leq f(\neg H, E, B).$$

Let f be an evidence neglecting strength indicator on $\mathcal{L}$. Then we have for all H, H', E, B in $\mathcal{L}$ for which $f(H, E, B)$, $f(\neg H, E, B)$, and $f(H', E, B)$ are defined:

$$B \vdash H \quad \Rightarrow \quad f(H, E, B) \leq f\left(H', E, B\right) \leq f(\neg H, E, B).$$

Every evidence based strength indicator is a strength indicator, and every strength indicator is an evidence neglecting strength indicator.

The requirements take into account that the set of possibilities falsified by a theory H is a subset of the set of possibilities ruled out by any theory logically implying H, where the set of possibilities is restricted to those (ruled out by the data but) allowed for by the background assumptions. It follows that logically equivalent theories are always equally informative (about the data) (have the same f-value), provided the relevant expressions are defined.

The first part of the observation says that a theory cannot do better in terms of informing *about the data* than logically implying them. Although this is not questionable, one might take this as a reason to reject the notion of informing about the data (because it is inappropriate to ascribe maximal informativeness to any theory logically implying the evidence). Two sentences, one might say, both logically implying all of the data can still differ in their informativeness: consider, for instance, a complete theory consistent with the data and a collection of all the data gathered so far. This argument is perfectly reasonable. Hence the distinction between evidence based and evidence neglecting strength indicators. The notion of a strength indicator is introduced in order to avoid that one *has to* take sides, though one *can* do so (g need not be increasing in both arguments).

In all three cases, the defining clauses express that strength indicators and truth indicators increase and decrease, respectively, with the logical strength of the hypothesis to be evaluated. These quantitative requirements correspond to the defining clauses of the qualitative relations of informativeness and plausibility, respectively.

Obviously, an evaluation function a should *not* be both a strength and a truth indicator, for any strength indicating truth indicator is a constant function. Let us call this observation the *singularity* of simultaneously indicating strength and truth. Instead, an evaluation function a should *weigh* between these two conflicting aspects: a has to be *sensitive to both informativeness and truth*.

Definition 4 *Let s be a strength indicator on $\mathcal{L}$, and let t be a truth indicator on $\mathcal{L}$. A possibly partial function $f : \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is* sensitive to informativeness and plausibility in the sense of s and t *– or for short: an s,t-*function *– iff there is a possibly partial function $g : \Re \times \Re \times X \to \Re$ such that (i) $f(H,E,B)$ is defined and $f(H,E,B) = g(s(H,E,B), t(H,E,B), x)$ for all H, E, B in $\mathcal{L}$ for which $s(H,E,B)$ and $t(H,E,B)$ are defined, and (ii)*

1. *Continuity: Any surplus in informativeness succeeds, if the difference in plausibility is small enough.*

   $$\forall \varepsilon > 0 \quad \exists \delta_\varepsilon > 0 \quad \forall s_1, s_2 \in R_s \quad \forall t_1, t_2 \in R_t \quad \forall x \in X : \\ s_1 > s_2 + \varepsilon \quad \& \quad t_1 > t_2 - \delta_\varepsilon \quad \Rightarrow \quad g(s_1, t_1, x) > g(s_2, t_2, x).$$

2. *Demarcation:* $\quad \forall x \in X : \quad g(s_{\max}, t_{\min}, x) = g(s_{\min}, t_{\max}, x) = 0.$

If $s(\bot, E, B)$ and $s(\top, E, B)$ are defined, they are the maximal and minimal values of s, $s_{\max}$ and $s_{\min}$, respectively. If $t(\top, E, B)$ and $t(\bot, E, B)$ are defined, they are the maximal and minimal values of t, $t_{\max}$ and $t_{\min}$, respectively. 'R_s' and 'R_t' denote the range of s and the range of t, respectively. $f(H, E, B)$ is a function of, among others, $s(H, E, B)$ and $t(H, E, B)$. I will sometimes write '$f(H, E, B)$', and other times '$g(s_1, t_1)$', dropping the additional argument place, and other times '$f(s_1, t_1)$', treating f as $g(s, t)$.

Continuity implies

3. Weak Continuity

$$\begin{array}{l}\forall s_1, s_2 \in R_s : s_1 > s_2 \quad \exists \delta_{s_1,s_2} > 0 \quad \forall t_1, t_2 \in R_t \quad \forall x \in X : \\ t_1 > t_2 - \delta_{s_1,s_2} \quad \Rightarrow \quad g(s_1, t_1, x) > g(s_2, t_2, x).\end{array}$$

The difference is that, in its stronger formulation, Continuity requires δ just to depend on the lower bound ε of the difference between s_1 and s_2, and not on the numbers s_1 and s_2 themselves. The difference between Continuity and Weak Continuity is related to the difference between evidence based and evidence neglecting strength indicators. When one is concerned with two hypotheses H_1 and H_2 and considers the incoming evidence once at a time, the plausibility of H_1 and H_2 in general changes with each new piece of evidence. In case of evidence based strength indicators, the informativeness of H_1 and H_2 also changes with each new piece of evidence, whereas it remains the same for evidence neglecting strength indicators. The idea behind Continuity now is that the more informative of two hypotheses eventually comes out as the better theory, if the plausibility of both hypotheses converges to certainty in the same truth value. If the informativeness of H_1 and H_2 itself changes with each new piece of evidence, though the informativeness of H_1 is always greater than that of H_2, one cannot refer to *the* informativeness values of H_1 and H_2, respectively. One can, however, refer to a minimal difference between the two informativeness values – unless this difference converges to 0 itself, in which case H_1 need not come out as the better theory anyway.

As mentioned, the idea behind Continuity is that the more informative of two hypotheses eventually comes out as the better one, when the plausibility of the two hypotheses converges to certainty in the same truth value. That is,

4. Continuity in Certainty: Any surplus in informativeness succeeds, if plausibility becomes certainty in the same truth value.

$$\begin{array}{l}\forall \varepsilon > 0 \forall t_i, t_i' \in R_t : t_i, t_i' \rightarrow_i \left\{ \begin{array}{l} t_{\max} \\ t_{\min} \end{array} \right. \exists n \forall m \geq n \forall s_m, s_m' \in R_s \forall x \in X : \\ s_m > s_m' + \varepsilon \quad \Rightarrow \quad g(s_m, t_m, x) > g(s_m', t_m', x).\end{array}$$

Continuity generalizes this idea from $t_{\max}$ and $t_{\min}$ to any value of t.

Weak Continuity implies that g increases in s, i.e.

5. Informativeness: $\quad s_0 > s_1 \quad \Rightarrow \quad g(s_0, t_0, x) > g(s_1, t_0, x)$.

If we additionally assume that g is a function of s and t *only*, we get

Lovelinesss: $\quad g(s_0, t_0, x) \geq g(s_1, t_0, x) \quad \Leftrightarrow \quad s_0 \geq s_1$.

Continuity does not imply that g increases in t, i.e.

0. Plausibility: $\quad t_0 > t_1 \quad \Rightarrow \quad g(s_0, t_0, x) > g(s_0, t_1, x)$.

s_0, s_1 are any values in the domain of s; t_0, t_1 are any values in the domain of t; and x is any value in X.

This asymmetry is due to the fact that truth is a qualitative yes-or-no affair: a sentence either is or is not true in some world, whereas informativeness (about some data) is a matter of degree. In case of truth, degrees enter the scence only because we do not know in general, given only the data, whether or not a theory is true in any world the data could be taken from. In case of informativeness, however, degrees are present even if we have a complete and correct assessment of the informational value of the theory under consideration (or more cautiously, there is at least a partial order that is induced by the consequence relation).

Weak Continuity in Certainty (which you get from reformulating 4 along the lines of 3) implies

6. Maximality: $\quad g(s_0, t_0, x) = g_{\max} \quad \Rightarrow \quad s_0 = s_{\max}$

7. Minimality: $\quad g(s_0, t_0, x) = g_{\min} \quad \Rightarrow \quad s_0 = s_{\min}$.

If we additionally assume Plausibility, we get

8. Maximality II: $\quad g(s_0, t_0, x) = g_{\max} \quad \Rightarrow \quad s_0 = s_{\max} \quad \& \quad t_0 = t_{\max}$

9. Minimality II: $\quad g(s_0, t_0, x) = g_{\min} \quad \Rightarrow \quad s_0 = s_{\min} \quad \& \quad t_0 = t_{\min}$.

If we add that g is a function of s and t *only*, we get the converse of 8 and of 9.

The conjunction of Continuity, Demarcation, and Plausibility does not imply

Symmetry: $\quad g(s_1, t_1, x) = g(t_1, s_1, x)$.

Evaluation functions may consider one aspect, say plausibility, more important than the other. The only thing that is ruled out is to totally neglect one of the two aspects, as do, for instance,

$$r = \frac{t}{1-s} \quad \text{and} \quad l = \frac{t \cdot s}{(1-t) \cdot (1-s)},$$

when $t = 0$ and $R_s = R_t = [0, 1]$. These functions have the following properties

$$s_0 > s_{\min} \quad \Rightarrow \quad g(s_0, t_{\min}) = g_{\min},$$

$$s_{\max} > s_0 > s_{\min} \quad \Rightarrow \quad g(s_0, t_{\min}) = g_{\min} \quad \& \quad g(s_0, t_{\max}) = g_{\max},$$

respectively. The first says that in the special case of plausibility being minimal, informativeness does not count anymore. But clearly, a theory which is refuted by the data – in which case its plausibility is minimal – can still be better than another theory which is also refuted by the data. After all, (almost) every interesting theory from, say, physics, has turned out to be false – and we nevertheless think there has been progress! The second property additionally says that in the special case of plausibility being maximal, informativeness does not count anymore either. So not only is any falsified theory as bad as any other falsified theory; we also have that every verified theory is as good as any other verified theory. In contrast,

$$d = t + s - 1, \quad R_t = R_s = [0, 1],$$

is sensitive to informativeness and plausibility, and thus does not exhibit the discontinuity of r and l. If f is a positive function not depending on H,

$$d_f = [t + s - 1] \cdot f(E, B)$$

also satisfies Plausibility, Continuity, and Demarcation, though it is not a function of s and t only. Finally, note that any s, t-function is invariant with respect to (or closed under) equivalence transformations of H, if it is a function of s and t only.

5 Evaluating Theories

5.1 The General Theory

What has been seen so far is the general plausibility-informativeness theory of theory evaluation. In a nutshell, its message is (1) that there are two epistemic values a theory should exhibit: truth and informativeness – measured by a truth indicator t and a strength indicator s, respectively; (2) that these two values are conflicting in the sense that the former is a decreasing, and the latter an increasing function of the logical strength of the hypothesis to be evaluated; and (3) that in evaluating a given hypothesis one should weigh between these two conflicting aspects in such a way that any surplus in informativeness succeeds, if the difference in plausibility is small enough. Particular accounts arise by inserting particular strength indicators and truth indicators.

5.2 Evaluating Theories, Bayes Style

The theory can be spelt out in terms of Spohn's (1988; 1990) ranking theory (Huber 2007), and in a syntactical paradigm that goes back to Hempel (1943; 1945) (Huber 2004). Here, however, I will focus on the Bayesian version, where I take Bayesianism to be the threefold thesis that (i) scientific reasoning is probabilistic; (ii) probabilities are adequately interpreted as an agent's actual subjective degrees of belief; and (iii) they can be measured by the agent's betting behaviour.

Spelling out the general theory in terms of subjective probabilities simply means that we specify a (set of) probabilistic strength indicator(s) and a (set of) probabilistic truth indicator(s). Everything else is accounted for by the general theory. The nice thing about the Bayesian paradigm is that once one is given hypothesis H, evidence E, and background information B, one is automatically given the relevant numbers $\mathrm{Pr}\left(H \mid E \wedge B\right), \ldots$, and the whole problem reduces to the definition of a suitable function of Pr.

In this paradigm it is natural to take

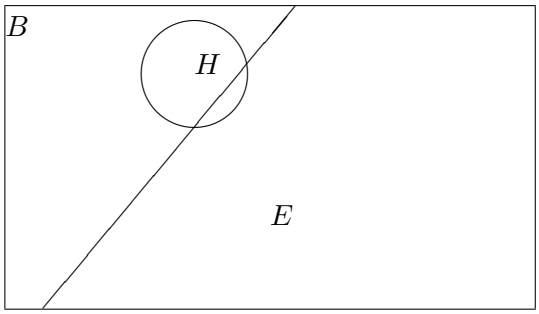
$$t_{\mathrm{Pr}}\left(H, E, B\right) = \mathrm{Pr}\left(H \mid E \wedge B\right) = p$$

as truth indicator, and

$$s_{\mathrm{Pr}}\left(H, E, B\right) = \mathrm{Pr}\left(\neg H \mid \neg E \wedge B\right) = i, \quad s'_{\mathrm{Pr}}\left(H, B\right) = \mathrm{Pr}\left(\neg H \mid B\right) = i'$$

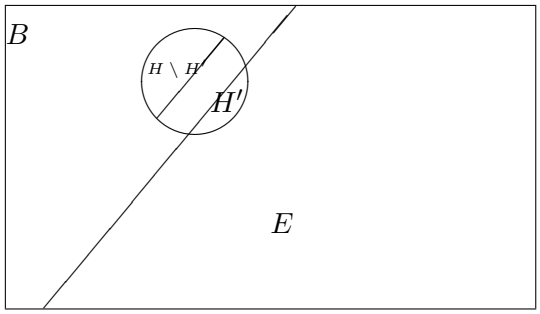
as evidence based and evidence neglecting strength indicators, respectively, where Pr is a *regular* probability.[2] The choice of p hardly needs any discussion, and for the choice of i consider the following figure with hypothesis H, evidence E, and

[2]Regularity is often paraphrased as open-mindedness (Earman 1992), because it demands that no consistent statement is assigned probability 0. Given a subjective interpretation of probability, this sounds like a restriction on what one is allowed to believe (to some degree). Regularity can also be formulated as saying that any statement H_1 which logically implies but is not logically implied by some other statement H_2 must be assigned a strictly lower degree of belief than H_2. (In case of probabilities conditional on B, logical implication is also conditional on B.) Seen this way, regularity requires degrees of belief which are sufficiently fine-grained. For this reason I prefer to think of regularity not as a restriction on *what* (which propositions) to believe (to some degree), but as a restriction on *how* to believe (propositions), namely, sufficiently fine-grained so that differences so big as to be expressable purely in terms of the logical consequence relation are not swept under the carpet.



background information B (conceived of as propositions).

Suppose you are asked to strengthen H by deleting possibilities verifying it, that is, by shrinking the area representing H.[3] Would you not delete possibilities outside E? After all, given E, those are exactly the possibilities known not to be the actual one, whereas the possibilities inside E are still alive options. Indeed, the probabilistic evidence based strength indicator i increases when H shrinks to H'



as depicted above.

For the probabilistic evidence neglecting strength indicator i' it does not matter which possibilities one deletes in strengthening H (provided all possibilities have equal weight on Pr). i' neglects whether they are inside or outside E. The strength indicator i^*_α with parameter α in $[0, 1]$ is given by

$$i^*_\alpha = \alpha \cdot \Pr(\neg H \mid \neg E \wedge B) + (1 - \alpha) \cdot \Pr(\neg H \mid B) = \alpha \cdot i + (1 - \alpha) \cdot i'.$$

For i^*_α, it depends on α how much it matters whether the deleted possibilities lie inside or outside of E.

Other candidates for measuring informativeness that are (suggested by measures) discussed in the literature (Carnap & Bar-Hillel 1952, Bar-Hillel & Carnap 1953, Hintikka & Pietarinen 1966) are

$$i'' \quad = \quad \Pr(\neg H \mid E \wedge B)$$

[3] I owe this graphical illustration to Luc Bovens.

$$\begin{aligned} \text{cont} &= \Pr(E) \cdot \Pr(\neg H \mid E \wedge B) \\ \text{inf} &= -\log_2 \Pr(H \mid E \wedge B). \end{aligned}$$

(In Levi 1967, i'' is proposed as, roughly, a measure for the relief from agnosticism afforded by accepting H as strongest relative to total evidence $E \wedge B$.) These measures, all of which assign minimal informativeness to any theory entailed by the data and the background information, do even worse on this count by requiring the deletion of the possibilities inside E. They measure how much the information in H goes beyond the information provided by E, which is not the appropriate notion of informativeness for the present context.

Note that the background information B plays a role different from that of the data E for i and i', but not for i'', cont, or inf. Clearly, if there is a difference between data on the one hand and background information on the other, then this difference should show up somewhere. Background assumptions determine the set of possibilities in the inquiry, and thus are nothing but restrictions on the set of possible worlds over which inquiry has to succeed (Hendricks 2006). Furthermore, evidence based strength indicators measure how much a theory informs about the data, but not how much they inform about the background assumptions. However, if one holds there should be no difference between E and B as far as measuring informativeness is concerned, then one can nevertheless adopt the above measures by substituting $E' = E \wedge B$ and $B' = \top$ for E and B, respectively.

5.3 Incremental Confirmation

Let us see how the plausibility-informativeness approach compares to Bayesian confirmation theory. The following notion is central in this literature (Fitelson 2001): A possibly partial function $f = f_{\Pr} : \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is a *relevance measure based on* $\Pr$ iff it holds for all H, E, B in $\mathcal{L}$ with $\Pr(E \wedge B) > 0$:

$$f(H,E,B) \begin{array}{c} > \\ = 0 \\ < \end{array} \quad \Leftrightarrow \quad \Pr(H \mid E \wedge B) \begin{array}{c} > \\ = \\ < \end{array} \Pr(H \mid B).$$

As

$$\Pr(H \mid E \wedge B) > \Pr(H \mid B) \quad \Leftrightarrow \quad \Pr(\neg H \mid \neg E \wedge B) > \Pr(\neg H \mid B)$$

for $0 < \Pr(E \mid B) < 1$ and $\Pr(B) > 0$, every i,p-function $a = p + i - 1$ is a relevance measure in the Bayesian sense (where p and i depend on $\Pr$). Similarly, every i',p-function $a' = p + i' - 1$ is a relevance measure. Hence, every i^*,p-function

$$a^* = p + i^* - 1$$

is a relevance measure, where i^* is a strength indicator based on i and i'. For $i^* = i'$ one gets the distance measure d,

$$d_{\Pr}(H, E, B) = \Pr(H \mid E \wedge B) - \Pr(H \mid B)$$

(Earman 1992), and for $i^* = i$ one gets the Joyce-Christensen measure s,

$$s_{\Pr}(H, E, B) = \Pr(H \mid E \wedge B) - \Pr(H \mid \neg E \wedge B)$$

(Joyce 1999, Christensen 1999). As noted earlier at the end of section 4, for positive f not depending on H, the functions

$$d_f = [i + p - 1] \cdot f(E, B)$$

are i, p-functions. For $f = \Pr(\neg E \mid B)$ we get (again) the distance measure d, and for $f = \Pr(\neg E \mid B) \cdot \Pr(B) \cdot (E \wedge B)$ we get the Carnap measure c,

$$c_{\Pr}(H, E, B) = \Pr(H \wedge E \wedge B) \cdot \Pr(B) - \Pr(H \wedge B) \cdot \Pr(E \wedge B)$$

(Carnap 1962). Hence the Carnap measure c, the difference measure d, and Joyce-Christensen measure s are three different ways of weighing between the two functions i and p (or between i' and p, for $s = d/\Pr(\neg E \mid B)$ and $c = d \cdot \Pr(B) \cdot \Pr(E \wedge B)$). Alternatively, the difference between d and s can be seen not as one between the way of weighing, but as one between *what* is weighed – namely two different pairs of functions, viz. i and p for the difference measure d, and i' and p for the Joyce-Christensen measure s. This is clearly seen by rewriting d and s as

$$\begin{aligned} d_{\Pr} &= \Pr(H \mid E \wedge B) + \Pr(\neg H \mid B) - 1, \\ s_{\Pr} &= \Pr(H \mid E \wedge B) + \Pr(\neg H \mid \neg E \wedge B) - 1. \end{aligned}$$

In this sense, part of the discussion about the right measure of incremental confirmation turns out to be a discussion about the right measure of informativeness of a hypothesis relative to a body of evidence. This view is endorsed by the observation that d and s actually employ the same decision-theoretic considerations (cf. Hempel 1960, Hintikka & Pietarinen 1966, Levi 1961; 1963):

$$\begin{aligned} d_{\Pr} &= \Pr(H \mid E \wedge B) - \Pr(H \mid B) \\ &= \Pr(H \mid E \wedge B) - \Pr(H \mid B) \cdot \Pr(H \mid E \wedge B) - \\ &\quad - \Pr(H \mid B) + \Pr(H \mid B) \cdot \Pr(H \mid E \wedge B) \\ &= \Pr(\neg H \mid B) \cdot \Pr(H \mid E \wedge B) - \Pr(H \mid B) \cdot \Pr(\neg H \mid E \wedge B) \\ &= i'(H, B) \cdot \Pr(H \mid E \wedge B) - i'(\neg H, B) \cdot \Pr(\neg H \mid E \wedge B) \end{aligned}$$

$$
\begin{aligned}
s_{\Pr} &= \Pr(H \mid E \wedge B) - \Pr(H \mid \neg E \wedge B) \\
&= \Pr(H \mid E \wedge B) - \Pr(H \mid \neg E \wedge B) \cdot \Pr(H \mid E \wedge B) - \\
&\quad - \Pr(H \mid \neg E \wedge B) + \Pr(H \mid \neg E \wedge B) \cdot \Pr(H \mid E \wedge B) \\
&= \Pr(\neg H \mid \neg E \wedge B) \cdot \Pr(H \mid E \wedge B) - \Pr(H \mid \neg E \wedge B) \cdot \Pr(\neg H \mid E \wedge B) \\
&= i(H, E, B) \cdot \Pr(H \mid E \wedge B) - i(\neg H, E, B) \cdot \Pr(\neg H \mid E \wedge B).
\end{aligned}
$$

So d and s are exactly alike in the way they combine or weigh between informativeness and plausibility – which is to form the expected informativeness of the hypothesis (about the data and relative to the background assumptions); their difference lies in the way they measure informativeness.

Given that the plausibility-informativeness theory has a nice justification in terms of conflicting epistemic virtues, given that it can be motivated historically[4], and given that – due its generality – it is free from being committed to the credo of any paradigm, Bayesian confirmation theory should welcome the connection between i^*, p-functions and relevance measures afforded by s and d (and c). This should be so the more, since in the light of the present approach part of the discussion about the right measure of incremental confirmation is one about the right measure of informativeness. Finally, as will be seen in sections 7-8, the present theory provides the only reasonable (and not yet occupied) answer to the question why one should stick to well confirmed theories rather than to any other theories.[5]

6 Selected Success Stories

This section briefly indicates how the plausibility-informativeness theory is successfully applied to some epistemological problems in the philosophy of science. Being notoriously short, the discussion often does not do justice to these problems. The following – Hempel's conditions of adequacy and the question of a logic of confirmation or theory assessment – is treated in more detail in Huber (2007).

[4] As to historical motivation, the ideas behind the strength indicator LO and the truth indicator LI in the Hempel paradigm (Huber 2004) go back to Hempel's prediction and satisfaction criteria (which is why that paradigm is called Hempel paradigm). It is interesting to see that these two criteria are both present in Hempel's seminal paper on confirmation (Hempel 1945), who thus seems to have felt the need for two concepts of confirmation, one aiming at true and another aiming at informative theories (see also Hempel 1960). This is particularly revealing as his triviality result that every observation report confirms every theory is basically due to the fact that informativeness is an increasing and plausibility a decreasing function of the logical strength of the theory to be assessed – and thus amounts to the singularity observation of section 4.

[5] For more on the relation between incremental confirmation and the plausibility-informativeness theory see Huber (2005b).

6.1 Hempel's Logic of Confirmation

6.1.1 Hempel's Conditions of Adequacy

In his "Studies in the Logic of Confirmation" (1945) Carl G. Hempel presented the following conditions of adequacy for any relation of confirmation $|\sim \subseteq \mathcal{L} \times \mathcal{L}$[6] on some language $\mathcal{L}$ (the name of 3.1 is not used by Hempel):

1. Entailment Condition: $\quad E \vdash H \quad \Rightarrow \quad E \mid\sim H$
2. Consequence Condition: $\quad \{H : E \mid\sim H\} \vdash H' \quad \Rightarrow \quad E \mid\sim H'$
   - 2.1 Special Consequence Cond.: $\quad E \mid\sim H, \quad H \vdash H' \quad \Rightarrow \quad E \mid\sim H'$
3. Consistency Condition: $\quad \{E\} \cup \{H : E \mid\sim H\} \nvdash \bot$
   - 3.1 Special C. C.: $\quad E \nvdash \bot, \quad E \mid\sim H, \quad H \vdash \neg H' \quad \Rightarrow \quad E \not\mid\sim H'$
4. Converse Consequence Condition: $\quad E \mid\sim H, \quad H' \vdash H \quad \Rightarrow \quad E \mid\sim H'$

Hempel then showed that 1, 2, and 4 entail that every sentence (observation report) E confirms every sentence (hypothesis) H, i.e. for all E, H in $\mathcal{L}$: $E \mid\sim H$. This is clear, since 1 and 4 already entail this result: By 1, $E \mid\sim E \vee H$, whence $E \mid\sim H$ by 4. Since Hempel's negative result, there has hardly been any progress in constructing a logic of confirmation.[7] One reason seems to be that up to now the predominant view on Hempel's conditions is the analysis Carnap gave in his *Logical Foundations of Probability* (1962), § 87.

6.1.2 Carnap's Analyis of Hempel's Conditions

In analyzing the consequence condition, Carnap argues that

> ... Hempel has in mind as explicandum the following relation: 'the degree of confirmation of H by E is greater than r', where r is a fixed value, perhaps 0 or $1/2$. (Carnap 1962, 475; notation adapted)

In discussing the consistency condition, Carnap mentions that

[6] Following Hempel, the background information B is neglected in section 6.1.

[7] The exceptions I know of are Flach (2000), Milne (2000), and Zwirn & Zwirn (1996). Roughly, Zwirn & Zwirn (1996) argue that there is no unified logic of confirmation (taking into account all of the partly conflicting aspects of confirmation); Flach (2000) argues that there are two logics of "induction", as he calls it, viz. confirmatory and explicatory induction (corresponding to Hempel's conditions 1-3 and 4, respectively); and Milne (2000) argues that there is a logic of confirmation (namely the logic of positive probabilistic relevance), but that it does not deserve to be called a logic.

> Hempel himself shows that a set of physical measurements may confirm several quantitative hypotheses which are incompatible with each other (p. 106). This seems to me a clear refutation of [3.1]. [...] What may be the reasons that have led Hempel to the consistency conditions [3.1] and [3]? He regards it as a great advantage of any explicatum satisfying [3] "that is sets a limit, so to speak, to the strength of the hypotheses which can be confirmed by given evidence" [...] This argument does not seem to have any plausibility for *our* explicandum, (Carnap 1962, 476-7; emphasis in the original)

which is the concept of positive probabilistic relevance,

> [b]ut it is plausible for the second explicandum mentioned earlier: the degree of confirmation exceeding a fixed value r. Therefore we may perhaps assume that Hempel's acceptance of the consistency condition is due again to an inadvertant shift to the second explicandum. (Carnap 1962, 477-8.)

Carnap's analysis can be summarized as follows: In presenting his first three conditions Hempel was mixing up two distinct concepts of confirmation, two distinct explicanda in Carnap's terminology, viz. the concept of incremental confirmation (positive probabilistic relevance) according to which E confirms H iff $\Pr(H \mid E) > \Pr(H)$; and the concept of absolute confirmation according to which E confirms H iff $\Pr(H \mid E) > r$, for some r $[.5, 1)$. The special versions of Hempel's second and third conditions hold true for the second explicandum, but they do not hold true for the first explicandum. On the other hand, Hempel's first condition holds true for the first explicandum, but it does so only in a *qualified* form (cf. Carnap 1962, 473) – namely only if E does not have probability 0, and H does not already have probability 1.

This, however, means that Hempel first had in mind the explicandum of positive probabilistic relevance for the Entailment Condition; then he had in mind the explicandum of absolute confirmation for the Special Consequence and the Special Consistency Conditions; and then, when Hempel presented the Converse Consequence Condition, he got completely confused, so to speak, and had in mind still another explicandum or concept of confirmation (neither absolute nor incremental confirmation satisfy 4.). Apart from not being very charitable, Carnap's reading of Hempel also leaves open the question what the third explicandum might have been.

6.1.3 Hempel Vindicated

As to Hempel's Entailment Condition, note that it is satisfied by the concept of absolute confirmation *without* the second qualification: If E logically implies H,

then $\Pr(H \mid E) = 1 > r$, for any r in $[0, 1)$, provided E does not have probability 0. So the following more charitable reading of Hempel seems plausible: When presenting his first three conditions, Hempel had in mind Carnap's second explicandum, the concept of absolute confirmation, or more generally: a plausibility relation. But then, when discussing the Converse Consequence Condition, Hempel also felt the need for a second concept of confirmation aiming at informative theories.

Given that it was the Converse Consequence Condition which Hempel gave up in his *Studies*, the present analysis makes perfect sense of his argumentation: Though he felt the need for two concepts of confirmation, Hempel also realized that these two concepts are *conflicting* – that is the content of his triviality result, corresponding to the singularity observation of section 4 – and so he abandoned informativeness in favour of plausibility.

Let us check this by going through Hempel's conditions. Absolute confirmation satisfies the Entailment Condition, as shown above. As to the Special Consequence and the Special Consistency Condition (where the present analysis agrees with Carnap's), it is clear that $\Pr(H' \mid E) > r$ whenever $\Pr(H \mid E) > r$ and $H \vdash H'$, and $\Pr(H' \mid E) < r$ whenever $\Pr(H \mid E) > r$, $H \vdash \neg H'$, and r in $[.5, 1)$. (Non-empty informativeness relations do not satisfy 3.1. Informativeness relations satisfying 2.1 or 1 are trivial in the sense that E confirms at least one H iff E confirms all H.) The culprit, according to Hempel (pp. 103-7, esp. pp. 104-5 of his 1945), is the Converse Consequence Condition. The latter condition coincides with the defining clause of informativeness relations by expressing the requirement that informativeness increases with the logical strength of the theory to be assessed. It is, for instance, satisfied by HD-confirmation.

6.1.4 The Logic of Theory Assessment

However, in a sense one can have Hempel's cake and eat it, too: There is a logic of confirmation – or rather: theory assessment – that takes into account both of these two conflicting concepts. Roughly speaking, HD says that a good theory is informative, whereas IL says that a good theory is plausible or true. The driving force behind Hempel's conditions, so the proposed analysis, is the insight that *a good theory is both true and informative*. Hence, in evaluating a given theory by the available data, one should account for both of these two conflicting aspects.

According to the logic of theory assessment as presented in Huber (2007), H is an *acceptable theory for* E iff H is at least as plausible as and more informative than its negation relative to E, or H is more plausible than and at least as plausible

as its negation relative to E. In terms of probabilities, this means:

$$\begin{aligned} E \mid\sim H \quad \Leftrightarrow \quad & \Pr(H \mid E) \geq \Pr(\neg H \mid E) \quad \& \quad i^*(H, E) > i^*(\neg H, E) \quad \text{or} \\ & \Pr(H \mid E) > \Pr(\neg H \mid E) \quad \& \quad i^*(H, E) \geq i^*(\neg H, E), \end{aligned}$$

where i^* is any function of $i = \Pr(\neg H \mid \neg E)$ and $i' = \Pr(\neg H)$ that is non-decreasing in both arguments, and increasing in at least one. $\mid\sim$ is the i^*-assessment relation induced by $\Pr$ on $\mathcal{L}$.

Before going on, let me note that the term 'acceptable' is meant to be the qualitative counterpart of the quantitative assessment value. 'accept' is not used in the sense of believe or hold to be true. Indeed, the account suggests that there is not even a specific propositional attitude towards theories given high assessment values. Rather, the proposed attitude towards scientific theories is like the attitude one has towards bottles of wine. One has a certain amount of money and one would like to buy a good bottle of wine. On the one hand, one wants to spend as little money as possible (one's theory should be as informative as possible). On the other hand, one wants to drink reasonably good wine (one's theory should be sufficiently plausible). Sometimes one need not care much about money, and the main focus is on the quality of wine – like when one is concerned with several alternative theories all sufficiently informative to answer one's questions, and one wants to choose the most plausible one. At other times money does matter, for one cannot spend more than one has. Likewise, in many situations very plausible theories just won't do, because they are too uninformative to be of any use.

This picture of the trade-off between price and quality does not tell one when a bottle of wine is worth its price, and when one should buy which bottle of wine (except when one gets a good bottle of wine for free). In the same way the plausibility-informativeness theory does not tell one when one should adopt or stick to a scientific theory (except when a theory is sufficiently informative to answer one's questions and known to be true). Instead, a theory which is acceptable given the data is a *possible* candidate to stick with.

Another, perhaps more natural way of defining the qualitative counterpart to the quantitative assessment value is to say that H is acceptable relative to E iff the overall value of H relative to E is greater than that of its negation. The reason I prefer the stronger notion of above is that the weaker notion is heavily dependent on the way one weighs between plausibility and informativeness. On the other hand, according to the stronger notion there may be hypotheses H_1, H_2, data E, and evaluation functions a such that H_1 is an acceptable theory for E, but H_2 is not, even though H_2 has a greater a-value relative to E than does H_1.[8]

[8]This was pointed out to me by Alexander Moffett.

Let us turn back. How do assessment relations compare to Carnap's favourite concept of confirmation? Positive probabilistic relevance between E and H is necessary in order for H to be an acceptable theory for E:

$$E \mid\sim H \quad \Rightarrow \quad \Pr(H \mid E) > \Pr(H)$$

However, the converse is not true, for positive probabilistic relevance is symmetric, whereas assessment relations are not – which, as noted by Christensen (1999, 437f), is as it should be.

6.2 Problems in Confirmation Theory

6.2.1 The Problem of Old Evidence

As Christensen (1999) further notices, a second problem of subjective[9] Bayesian confirmation theory is the problem of old evidence: If E is known in the sense of being assigned a degree of belief of 1, then E incrementally confirms no H relative to whatever B.

By Jeffrey conditionalisation, $i = \Pr(\neg H \mid \neg E \wedge B)$ and $p = \Pr(H \mid E \wedge B)$, and thus all i, p-functions which are functions of i and p only, are invariant with respect to changes in $\Pr(E \mid B)$. This means that no such function faces the more general version of the problem of old evidence. This general version says that H is more confirmed by E relative to B in the sense of $\Pr_2$ than in the sense of $\Pr_1$ just in case $\Pr_2(E \mid B) < \Pr_1(E \mid B)$, where $\Pr_2$ results from $\Pr_1$ by Jeffrey conditioning on E, and E is positively relevant for H given B in the sense of $\Pr_1$.[10] In other words, the problem is that the less reliable the source of information, the higher the degree of confirmation. (The traditional problem of old evidence, i.e. the special case where $\Pr(E \mid B) = 1$, does not arise, because $\Pr$ is regular, and it is assumed that $\{\neg E, B\} \nvdash \bot$ and $\{E, B\} \nvdash \bot$ – otherwise i, p-functions are not defined.) The more general version is faced by the distance measure d, the log-likelihood ratio l, and the ratio measure r (Huber 2005a).

6.2.2 Tacking by Conjunction

If evidence E confirms hypothesis H relative to background information B, then E generally does not confirm (relative to B) the conjunction of H and an arbitrary

[9] Though the importance of interpreting $\Pr$ is sometimes dismissed in Bayesian confirmation theory, some problems – e.g. the problem of old evidence – arise only under particular interpretations of probability. In this case it is the subjective interpretation that takes $\Pr$ to be an agent's actual degree of belief function.

[10] In case E is negatively relevant for H given B in the sense of $\Pr_1$, this holds just in case $\Pr_2(E \mid B) > \Pr_1(E \mid B)$. Negative evidence provides more disconfirmation and positive evidence provides more confirmation, the lower the degree of belief in it.

hypothesis H' (Fitelson 2002, Hawthorne & Fitelson 2004). This is in accordance with the present approach. Although adding H' does not decrease the informativeness of H relative to E and B, it generally does lead to a decrease in the plausibility of H relative to E and B. In fact, the present account can specify when an added conjunct should not be tacked on: Given E and B, H' should not be tacked on H relative to an evaluation function a precisely if the a-value of $H \wedge H'$ is smaller than that of H.

6.2.3 Theory Hostility

It is sometimes claimed that confirmation is inappropriate for theory evaluation, because confirmation does not take into account the fact that theories should possess several other virtues besides being true or having a high probability. This exclusive focus on truth or probability is referred to as theory hostility. An adequate theory of theory evaluation should yield that good theories are not only true or probable; they should also be informative. Obviously this holds of any theory with a high value in the combined sense of informativeness and plausibility.

7 What Is the Point?

The crucial question any theory of theory evaluation has to face is this: What is the point of having theories that are given high assessment values? That is, why are theories with high values better than any other theories? In terms of confirmation the question is: What is the point of having theories that are well confirmed by the available data relative to some background information? That is, why should we stick to well confirmed theories rather than to any other theories?

The traditional answer to this question is that the goal is truth, and that one should stick to well confirmed theories because, in the long run, confirmation takes one to the truth.[11] But as we have seen, truth is only one side of the coin – the other is informativeness. Thus, the answer of the new millenium is that the goal is informative truth, and that one should stick to theories with high values because, in the medium run, theory evaluation takes one to the most informative among all true theories.

Indeed, if being taken to the most informative among all true theories is not the goal of confirmation, it seems there is no point to incremental confirmation at all. The traditional approach to confirmation the early Carnap had, i.e. absolute confirmation setting confirmation equal to (logical) probability, has long been aban-

[11] This is the line of reasoning the early Carnap (could have) had. He held that confirmation is equal to (logical) probability – absolute confirmation, in contrast to incremental confirmation.

doned in favour of incremental confirmation setting confirmation equal to increase in probability. Though I do not want to opt for a revival of absolute confirmation, it is fair to say that absolute confirmation at least could be justified by arguing that, in the long run, absolute confirmation takes one to the truth (which is the content of the convergence theorem; Gaifman & Snir 1982). So if the goal were truth and only truth, there would have been no need for abandoning absolute confirmation. Hence there must be another goal for incremental confirmation. But then, if arriving at (the most) informative (among all) true theories is not the goal, what else could it be? Yet, as will be seen in section 8, incremental confirmation does not take one to informative true theories nor to the most informative among all true theories.

What is an informative true theory? Given a possible world (possibility, model) ω, contingent theory H_1 is to be preferred over contingent theory H_2 in ω if

1. H_1 is true in ω, but H_2 is false in ω; or

2. both H_1 and H_2 have the same truth value in ω, but H_1 logically implies H_2, whereas H_2 does not logically imply H_1.

In case H_1 is logically false, it is worse in ω than every contingent theory H_2 that is true in ω (because they are all true in ω, whereas H_1 is false in ω), but better than every contingent theory H_2 that is false in ω (because H_1 is more informative than each of them). Similarly, if H_1 is logically true, it is worse in ω than every contingent theory H_2 that is true in ω (because they all are more informative than H_1), but better than every contingent theory H_2 that is false in ω (because they all are false in ω, whereas H_1 is true in ω).

Consequently, a possibly partial function $f : \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is said to *reveal the true assessment* (or *confirmational*) *structure in world* ω iff for any hypotheses H, H' in $\mathcal{L}$, every background information B in $\mathcal{L}$ which is true in ω, and any data stream $e_0, \ldots, e_n, \ldots$ from ω (i.e. a sequence of sentences e_i in $\mathcal{L}$ expressing distinct propositions all of which are true in ω):

1. If H is contingently true in ω and H' is contingently false in ω, then there is n such that for all $m \geq n$: $f\,(H, E_m, B) > 0 > f\,(H', E_m, B)$;

2. if H and H' are contingently true in ω, but H is logically stronger than H', then there is n such that for all $m \geq n$: $f\,(H, E_m, B) > f\,(H', E_m, B) > 0$;

3. if H and H' are contingently false in ω, but H is logically stronger than H', then there is n such that for all $m \geq n$: $0 > f\,(H, E_m, B) > f\,(H', E_m, B)$;

4. if H is logically determined, then it holds for all m: $f\,(H, E_m, B) = 0$;

where $E_m = e_0 \wedge \ldots \wedge e_{m-1}$ is the conjunction of the first m data sentences. So f must *stabilize* to the correct answer, i.e. get it right after finitely many steps, and continue to do so forever without necessarily halting (or giving any other sign that it has arrived at the correct answer).[12] The smallest n for which the above holds is called the point of stabilisation.

The central question is whether evaluation functions do in fact further the goal they are supposed to further; that is, whether they in fact reveal the true assessment structure and thus lead to true and informative theories. The answer is affirmative – which is the promised justification for the Bayesian version of the plausibility-informativeness theory.[13]

More precisely, let $e_0, \ldots, e_n, \ldots$ be a sequence of sentences of $\mathcal{L}$ which separates the set of all models for $\mathcal{L}$, $Mod_{\mathcal{L}}$. This means that for any two distinct ω_1, ω_2 in $Mod_{\mathcal{L}}$ there is a sentence e_i which is true in ω_1 and false in ω_2. Let e_i^{ω} be e_i, if e_i is true in ω, and $\neg e_i$ otherwise. Let $\Pr$ be a regular probability on $\mathcal{L}$, and let a be any function satisfying Continuity in Certainty and Demarcation for i^* and p, where i^* is any strength indicator based on $i = \Pr(\neg H \mid \neg E \wedge B)$ and $i' = \Pr(\neg H \mid B)$ and $p = \Pr(H \mid E)$. Finally, let $\Pr^*$ be the unique probability measure on the smallest σ-field $\mathcal{A}$ containing the field $\{Mod(A) : A \in \mathcal{L}\}$ such that for all H in $\mathcal{L}$: $\Pr(H) = \Pr^*(Mod(H))$, where $Mod(A)$ is the set of models of A. Then there exists X in $\mathcal{A}$ with $\Pr^*(X) = 1$ such that the following holds for every ω in X and any two H, H' in $\mathcal{L}$:

1. If H is contingently true in ω and H' is contingently false in ω, then there is n such that for all $m \geq n$: $f(H, E_m^{\omega}, \top) > 0 > f(H', E_m^{\omega}, \top)$;

2. if H and H' are contingently true in ω, but H is logically stronger than H', then there is n such that for all $m \geq n$: $f(H, E_m^{\omega}, \top) > f(H', E_m^{\omega}, \top) > 0$;

[12] Stabilisation to the correct answer is a stronger requirement than convergence to the correct answer (Kelly 1996). The latter is a bit odd to formulate for revealing the true assessment structure, but in general says that for any $\varepsilon > 0$ (as small as you like) there exists a point n (depending on ε) such that for all later points $m > n$, f's conjecture differs form "the truth" only by an amount smaller than ε. The difference between stabilisation and convergence was the reason for appealing to the medium run (stabilisation) as compared to the long run (convergence). Note, however, that the Gaifman and Snir convergence theorem can be used to obtain an almost-sure stabilisation result by assigning 1 to H, if the probability of H is above .5, and 0 otherwise.

[13] As always, there are problems. For instance, the result stated below holds only for almost every world and is restricted to data sequences that separate $Mod_{\mathcal{L}}$. However, this is not due to anything being wrong with the plausibility-informativeness theory. Instead, this flaw is inherited from the present paradigm. The flaw is serious (Kelly 1996, ch. 13), but not inevitable, because there are other paradigms one might adopt such as ranking theory, where "pointwise reliability" is possible (Kelly 1999). However, the price of pointwise reliability is that the set of possible worlds be countable, and it is fair to say that measure 1 results are not problematic in this case.

3. if H and H' are contingently false in ω, but H is logically stronger than H', then there is n such that for all $m \geq n$: $0 > f\left(H, E_m^\omega, \top\right) > f\left(H', E_m^\omega, \top\right)$;

4. if H is logically determined, then it holds for all m: $f\left(H, E_m^\omega, \top\right) = 0$;

where $E_m^\omega = e_0^\omega \wedge \ldots \wedge e_{m-1}^\omega$ is the conjunction of the first m sentences from the data stream from ω. In other words, every function satisfying Continuity in Certainty and Demarcation in the sense of i^* and p reveals the true assessment structure in almost every world when presented data separating the set of all possible worlds. The background information has been assumed to be empty. This is justified because the above entails that there exists X in $\mathcal{A}$ with $\Pr^*\left(X \mid Mod\left(B\right)\right) = 1$ for every B in $\mathcal{L}$ with $\Pr\left(B\right) > 0$, such that 1-4 hold for every ω in $Mod\left(B\right) \cap X$.

8 Relevance Measures and Their Exclusive Focus on Truth

All one needs to do to reveal the true assessment structure in almost every world when presented separating data is to stick to any function satisfying Continuity in Certainty and Demarcation for i^* and p, where i^* is any function of i and i' that is non-decreasing in both and increasing in at least one of its arguments. What about the central notion in Bayesian confirmation theory – that of a relevance measure?

The connection to the i, p-function $s = i + p - 1$ and the function d_f for $f = \Pr\left(\neg E \mid B\right)$ respectively $f = \Pr\left(\neg E \mid B\right) \cdot \Pr\left(B\right) \cdot \Pr\left(E \wedge B\right)$ has already been pointed out. So for any regular probability $\Pr$, $s_{\Pr}$ and $d_{\Pr}$ and $c_{\Pr}$ reveal the true assessment structure in almost every world when presented separating data. But there are many other relevance measures. Do they all further that goal?

If H_1 is contingently true in ω, and H_2 is contingently false in ω, then, after finitely many steps, H_1 has to get a positive value in ω and H_2 has to get a negative value in ω. Any relevance measure r reveals this part of almost any ω's assessment structure. By the Gaifman and Snir convergence theorem,

$$\Pr\left(H_1 \mid E_n^\omega\right) \to_n 1 \quad \text{and} \quad \Pr\left(H_2 \mid E_n^\omega\right) \to_n 0,$$

whence there exists n such that for all $m \geq n$:

$$\Pr\left(H_1 \mid E_m^\omega\right) > \Pr\left(H_1\right) \quad \text{and} \quad \Pr\left(H_2 \mid E_m^\omega\right) < \Pr\left(H_2\right),$$

provided $\Pr$ is regular. Thus, by the definition of a relevance measure, it holds for all $m \geq n$:

$$r\left(H_1, E_m^\omega, \top\right) > 0 > r\left(H_2, E_m^\omega, \top\right).$$

Moreover, if defined, the value of any logically determined hypothesis is always 0.

So far, so good. But the definition of a relevance measure by itself does not imply anything about the relative positions of two hypotheses, if they have the same truth value in some world ω. This exclusive focus on truth – in contrast to the weighing between the conflicting goals of truth and informativeness of an s,t-function – is what prevents relevance measures from revealing the true assessment structure in general. As we have seen, relevance measures sometimes do weigh between i^* and p. Yet, they are not required to take into account both aspects. This is illustrated by briefly looking at the most popular relevance measures (Fitelson 2001). It is assumed throughout that $\Pr$ is regular.

As already mentioned, the Joyce-Christensen measure s, the distance measure d, and the Carnap measure c get it right in all four cases. The ratio measure r,

$$r_{\Pr}(H, E, B) = \log\left[\frac{\Pr(H \mid E \wedge B)}{\Pr(H \mid B)}\right],$$

gets it right in case both H_1 and H_2 are contingently true in ω and H_1 is logically stronger than H_2. In this case

$$r_{\Pr}(H_1, E_n^\omega) \to_n \log\left[1/\Pr(H_1)\right] \quad \text{and} \quad r_{\Pr}(H_2, E_n^\omega) \to_n \log\left[1/\Pr(H_2)\right],$$

whence there exists n such that for all $m \geq n$:

$$r_{\Pr}(H_1, E_m^\omega, \top) > r_{\Pr}(H_2, E_m^\omega, \top) > 0.$$

However, r does not get it right when both H_1 and H_2 are contingently false in ω and H_1 is logically stronger than H_2. In this case,

$$\frac{\Pr(H_1 \mid E_m^\omega)}{\Pr(H_1)} > \frac{\Pr(H_2 \mid E_m^\omega)}{\Pr(H_2)} \quad \Leftrightarrow \quad \frac{\Pr(H_2)}{\Pr(H_1)} > \frac{\Pr(H_2 \mid E_m^\omega)}{\Pr(H_1 \mid E_m^\omega)}.$$

For $\varepsilon = \Pr(H_2) - \Pr(H_1)$ and $\varepsilon_m = \Pr(H_2 \mid E_m^\omega) - \Pr(H_1 \mid E_m^\omega)$, this can be written as

$$1 + \frac{\varepsilon}{\Pr(H_1)} > 1 + \frac{\varepsilon_m}{\Pr(H_1 \mid E_m^\omega)}.$$

So even if both $\Pr(H_1 \mid E_m^\omega)$ and $\Pr(H_2 \mid E_m^\omega)$ converge to 0, the logically weaker H_2 may always have a greater r-value than H_1, as is the case when $\Pr(H_1 \mid E_m^\omega) = 1/2^m$ and $\Pr(H_2 \mid E_m^\omega) = 1/m$. The failure of r is even clearer when both H_1 and H_2 are eventually falsified, in which case the only thing that matters is the minimal plausibility value, and they both get same r-value $\log 0 = -\infty$. So all falsified theories are equally, viz. maximally bad. (For logically determined H, r takes on the value $\log 1 = 0$, if it is stipulated that $0/0 = 1$.)

The situation is even worse for the log-likelihood ratio l,

$$l_{\Pr}(H, E, B) = \log\left[\frac{\Pr(E \mid H \wedge B)}{\Pr(E \mid \neg H \wedge B)}\right] = \log\left[\frac{\Pr(H \mid E \wedge B) \cdot \Pr(\neg H \mid B)}{\Pr(\neg H \mid E \wedge B) \cdot \Pr(H \mid B)}\right].$$

When H_1 and H_2 are contingently true or contingently false in ω, and H_1 is logically stronger than H_2, it need not be the case that there is n such that for all $m \geq n$:

$$\frac{\Pr(H_1 \mid E_m^\omega) \cdot \Pr(\neg H_1)}{\Pr(\neg H_1 \mid E_m^\omega) \cdot \Pr(H_1)} > \frac{\Pr(H_2 \mid E_m^\omega) \cdot \Pr(\neg H_2)}{\Pr(\neg H_2 \mid E_m^\omega) \cdot \Pr(H_2)}.$$

For $\varepsilon = \Pr(H_2) - \Pr(H_1)$ and $\varepsilon_m = \Pr(H_2 \mid E_m^\omega) - \Pr(H_1 \mid E_m^\omega)$ the latter holds iff

$$1 + \frac{\varepsilon}{\Pr(H_1) \cdot (1 - \Pr(H_1) - \varepsilon)} > 1 + \frac{\varepsilon_m}{\Pr(H_1 \mid E_m^\omega) \cdot (1 - \Pr(H_1 \mid E_m^\omega) - \varepsilon_m)}.$$

So even if both $\Pr(H_1 \mid E_m^\omega)$ and $\Pr(H_2 \mid E_m^\omega)$ converge to 1 or to 0, the logically weaker H_2 may always have a greater l-value than H_1, as is the case when $\Pr(H_1 \mid E_m^\omega) = 1 - 1/m$ and $\Pr(H_2 \mid E_m^\omega) = 1 - 1/2^m$, or when $\Pr(H_1 \mid E_m^\omega) = 1/2^m$ and $\Pr(H_2 \mid E_m^\omega) = 1/m$. The failure of l is even clearer when both H_1 and H_2 are eventually verified or falsified, in which case the only thing that matters is the maximal or minimal plausibility value, and they both get the maximal or minimal l-value, respectively. So all verified theories are equally, viz. maximally good; and all falsified theories are equally, viz. maximally bad. (If H is logically determined, one would have to stipulate that $0 \cdot 1/1 \cdot 0 = 1 \cdot 0/0 \cdot 1 = 1$.)

It is interesting to see that the log-likelihood ratio l seems to come out on top when subjectively plausible desiderata are at issue (Fitelson 2001), but to do much more poorly when it comes to the matter-of-fact question whether an evaluation function (or measure of confirmation) furthers the goal it is supposed to further – whether it gets it right in the sense that it reveals the true assessment (or confirmational) structure and thus leads to true and informative theories. Due to their focus on truth, relevance measures – like s, t-functions – separate true from false theories, but due to the exclusiveness of this focus, they do not – in contrast to s, t-functions – distinguish between informative and uninformative true or false theories.

Acknowledgements

My research was in part supported by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government through a Sofja Kovalevskaja Award to Luc Bovens, while I was a member of the *Philosophy, Probability, and Modeling* group at the Center for Junior Research Fellows at the University of Konstanz.

References

[1] Bar-Hillel, Yehoshua & Carnap, Rudolf (1953), Semantic Information. *British Journal for the Philosophy of Science* **4**, 147-157.

[2] Carnap, Rudolf (1962), *Logical Foundations of Probability*. 2^{nd} ed. Chicago: University of Chicago Press.

[3] Carnap, Rudolf & Bar-Hillel, Yehoshua (1952), *An Outline of a Theory of Semantic Information*. Technical Report No. 247 of the Research Laboratory of Electronics, MIT. Reprinted in Bar-Hillel, Y. (1964), *Language and Information. Selected Essays on Their Theory and Application*. Reading, MA: Addison-Wesley, 221-274.

[4] Christensen, David (1999), Measuring Confirmation. *Journal of Philosophy* **96**, 437-461.

[5] Earman, John (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

[6] Fitelson, Branden (1999), The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science* **66** (Proceedings), S362-S378.

[7] —— (2001), *Studies in Bayesian Confirmation Theory*. PhD Dissertation. Madison, WI: University of Wisconsin-Madison.

[8] —— (2002), Putting the Irrelevance Back Into the Problem of Irrelevant Conjunction. *Philosophy of Science* **69**, 611-622.

[9] Flach, Peter A. (2000), Logical Characterisations of Inductive Learning. In D.M. Gabbay & R. Kruse (eds.), *Abductive Reasoning and Learning*. Dordrecht: Kluwer Academic Publishers, 155-196.

[10] Gaifman, Haim & Snir, Marc (1982), Probabilities over Rich Languages, Testing, and Randomness. *Journal of Symbolic Logic* **47**, 495-548.

[11] Hawthorne, James & Fitelson, Branden (2004), Re-Solving *Irrelevant Conjunction* with Probabilistic Independence. *Philosophy of Science* **71**, 505-514.

[12] Hempel, Carl Gustav (1943), A Purely Syntactical Definition of Confirmation. *Journal of Symbolic Logic* **8**, 122-143.

[13] —— (1945), Studies in the Logic of Confirmation. *Mind* **54**, 1-26, 97-121. Reprinted in Hempel, C.G. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, 3-51.

[14] —— (1960), Inductive Inconsistencies. *Synthese* **12**, 439-469. Reprinted in Hempel, C.G. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, 53-79.

[15] Hendricks, Vincent F. (2006), Mainstream and Formal Epistemology. Cambridge: Cambridge University Press.

[16] Hintikka, Jaakko & Pietarinen, Juhani (1966), Semantic Information and Inductive Logic. In J. Hintikka & P. Suppes (eds.), *Aspects of Inductive Logic*. Amsterdam: North-Holland, 96-112.

[17] Huber, Franz (2004), *Assessing Theories. The Problem of a Quantitative Theory of Confirmation*. PhD Dissertation. Erfurt: University of Erfurt.

[18] —— (2005a), Subjective Probabilities as Basis for Scientific Reasoning? *The British Journal for the Philosophy of Science* **56**, 101-116.

[19] —— (2005b), What Is the Point of Confirmation? To appear in *Philosophy of Science* (Proceedings).

[20] —— (2007), The Logic of Theory Assessment. To appear in the *Journal of Philosophical Logic*.

[21] Joyce, James M. (1999), *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

[22] Kelly, Kevin T. (1996), *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.

[23] —— (1999), Iterated Belief Revision, Reliability, and Inductive Amnesia. *Erkenntnis* **50**, 11-58.

[24] Levi, Isaac (1961), Decision Theory and Confirmation. *Journal of Philosophy* **58**, 614-625.

[25] —— (1963), Corroboration and Rules of Acceptance. *The British Journal for the Philosophy of Science* **13**, 307-313.

[26] —— (1967), *Gambling With Truth. An Essay on Induction and the Aims of Science*. London: Routledge.

[27] Milne, Peter (2000), Is There a Logic of Confirmation Transfer? *Erkenntnis* **53**, 309-335.

[28] Spohn, Wolfgang (1988), Ordinal Conditional Functions: A Dynamic Theory of Epistemic States. In W.L. Harper & B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics II*. Dordrecht: Kluwer, 105-134.

[29] —— (1990), A General Non-Probabilistic Theory of Inductive Reasoning. In R.D. Shachter et. al. (eds.), *Uncertainty in Artificial Intelligence 4*. Amsterdam: North-Holland, 149-158.

[30] van Fraassen, Bas C. (1983), Theory Comparison and Relevant Evidence. In J. Earman (ed.), *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, 27-42.

[31] Zwirn, Denis & Zwirn, Hervé P. (1996), Metaconfirmation. *Theory and Decision* **41**, 195-228.