Synthese (2008) 161:89–118
DOI 10.1007/s11229-006-9141-x

ORIGINAL PAPER

# Assessing theories, Bayes style

**Franz Huber**

Received: 17 July 2004 / Accepted: 18 December 2006 / Published online: 30 March 2007


**Abstract** The problem addressed 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, BC, 1983, Theory comparison and relevant Evidence. In J. Earman (Ed.), *Testing scientific theories* (pp. 27–42). Minneapolis: University of Minnesota Press). Sections 1–3 contain the general plausibility-informativeness theory of theory assessment. In a nutshell, the message is (1) that there are two 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 assessed; and (3) that in assessing 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 shortfall in plausibility is small enough. Particular accounts of this general theory arise by inserting particular strength indicators and truth indicators. In Section 4 the theory is spelt out for the Bayesian paradigm of subjective probabilities. It is then compared to incremental Bayesian confirmation theory. Section 4 closes by asking whether it is likely to be lovely. Section 5 discusses a few problems of confirmation theory 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, CG, 1945, Studies in the logic of comfirmation. *Mind, 54*, 1–26, 97–121.), differing from the one Carnap gave in §87 of his *Logical foundations of probability* (1962, Chicago: University of Chicago Press). Section 6 adresses the question of justification any theory of theory assessment has to face: why should one stick to theories given high assessment values rather than to any

A precursor of this paper appears as "The Plausibility-Informativeness Theory" in V. F. Hendricks & D. Pritchard (eds.), *New Waves in Epistemology*. Aldershot: Ashgate, 2007.

F. Huber (✉)
California Institute of Technology, Pasadena, CA, USA
e-mail: franz@caltech.edu

other theories? The answer given by the Bayesian version of the account presented in section 4 is that one should accept theories given high assessment values, because, in the medium run, theory assessment almost surely takes one to the most informative among all true theories when presented separating data. The concluding section 7 continues the comparison between the present account and incremental Bayesian confirmation theory.

**Keywords** Theory evaluation · Confirmation · Probability

## 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, p. 27)

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

- We are given a hypothesis or theory $H$, a set of data—the evidence—$E$, and some background information $B$.
- The question is: how "good" is $H$ given $E$ and $B$? That is, what is the "*value*" of $H$ in view of $E$ and $B$?
- An answer to this 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 given $E$ and $B$.

Given this formulation of our problem, a theory of theory assessment 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, one 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 assessment 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 indicates that a proper treatment of the whole problem not only *explicates* how we evaluate theories in the light of the available evidence (sections 2–5). A proper treatment also *justifies* this evaluation by answering the question why we should accept those theories that score highest (sections 6 and 7).

In order for the above characterization 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 $H$, 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. As theory assessment 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.

## 2 Conflicting concepts of confirmation

Although some take theory assessment to be the main epistemic problem concerning science, there is no established branch addressing 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)[1] and Rudolf Carnap's work on inductive logic and probability (Carnap, 1952, 1962). The first decades have been dominated by the following two approaches.

- The qualitative theory of *Hypothetico-Deductivism* HD (sometimes associated with Karl R. Popper) says that 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.
- The quantitative theory of probabilistic *Inductive Logic* IL (associated with Rudolf Carnap) says that $H$ 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 probability of $H$ given $E$ and $B$ is greater than some fixed value $r \in [.5, 1)$.[2]

So there are at least two concepts of confirmation. There is a concept of confirmation that aims at *informative* theories, and there is a concept of confirmation that aims 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 assessed.

Let us turn this into a definition.

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

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

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

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

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 the more informative, the more possibilities it excludes. Hence, the logically stronger a sentence, the more informative it is.

[1] Cf. also Hempel (1943) and Hempel and Oppenheim (1945).

[2] This is *not* the way Carnap defines qualitative IL-confirmation in chapter VII of his (1962). There he requires 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 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)—is taken as the degree of *incremental* confirmation, and Carnap's proposal is the natural qualitative counterpart of this notion of incremental confirmation. See section 5.

On the other hand, a sentence is the 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 $H$ is informative relative to $E$, then so is any logically stronger sentence. Similarly, if $H$ is plausible given $E$, then so is any logically weaker sentence.

According to HD, $E$ confirms $H$ relative to $B$ iff the conjunction of $H$ and $B$ logically implies $E$ (in some suitable way). Hence, if $E$ HD-confirms $H$ relative to $B$, and if $H'$ logically implies $H$, then $E$ HD-confirms $H'$ relative to $B$ (provided the suitable way does not render logical implication non-monotonic). So HD-confirmation is an informativeness relation. According to qualitative IL, $E$ confirms $H$ relative to $B$ iff $\Pr(H \mid E \wedge B) > r$, for some $r \in [.5, 1)$. Hence, if $E$ absolutely IL-confirms $H$ relative to $B$, and if $H$ logically implies $H'$, then $E$ absolutely IL-confirms $H'$ relative to $B$. So absolute IL-confirmation is a plausibility relation.

The epistemic values behind these two concepts are *informativeness* on the one hand and *truth* or *plausibility* on the other. We aim at theories that are true in the world we are in. And we aim at theories that inform us about the world we are in. Usually we do not know which world we are in, though. All we have are some data (and background assumptions). So we base our evaluation of the theory we are concerned with on the plausibility that the 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 now say that, according to HD, a good theory is informative, whereas IL says good theories are probable or true. Putting together these two claims gives us the plausibility-informativeness theory of theory assessment:

*a good theory is true and informative*.

## 3 Assessing theories

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 informative $H$ is relative to $E$ and $B$. 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 an *evidence based truth indicator on* $\mathcal{L}$ iff for all $\langle H, E, B\rangle, \langle H', E, B\rangle \in Dom_f$:

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

$f$ is an *evidence neglecting truth indicator on* $\mathcal{L}$ iff for all $\langle H, E, B\rangle, \langle H', E, B\rangle \in Dom_f$:

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

**Observation 1** Let $f$ be an evidence based truth indicator on $\mathcal{L}$. Then we have for all $\langle H, E, B\rangle, \langle \neg H, E, B\rangle, \langle H', E, B\rangle \in Dom_f$:

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

Let $f$ be an evidence neglecting truth indicator on $\mathcal{L}$. Then we have for all $\langle H, E, B\rangle$, $\langle \neg H, E, B\rangle, \langle H', E, B\rangle \in Dom_f$:

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

The range of $f$ is taken to be (a subset of) $\Re$. One could strive for maximal generality by taking the range of $f$ to be any ordered set in which differences can be expressed. 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 $H$. Here the set of possibilities is restricted to those not already ruled out by (the data and) the background information. It follows that logically equivalent theories always have the same plausibility ($f$-value), provided the relevant tuples $\langle H, E, B\rangle$ are in the domain of $f$.

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 assumptions our assessment is based on guarantee (in the sense of logical implication) that the theory to be assessed 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 our theory is false. In the following I will only consider evidence based truth indicators. So whenever I speak of a truth indicator I mean an evidence based truth indicator.

**Definition 3** A possibly partial function $f$: $\mathcal{L} \times \mathcal{L} \times \mathcal{L} \rightarrow \Re$ is an *evidence based strength indicator on* $\mathcal{L}$ iff for all $\langle H, E, B\rangle, \left\langle H', E, B\right\rangle \in Dom_f$:

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

$f$ is an *evidence neglecting strength indicator on* $\mathcal{L}$ iff for all $\langle H, E, B\rangle, \left\langle H', E, B\right\rangle \in Dom_f$:

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

$f$ is a *strength indicator on* $\mathcal{L}$ iff there is an evidence based strength indicator $f_1$, an evidence neglecting strength indicator $f_2$, and a function $g : R_{f_1} \times R_{f_2} \rightarrow \Re$ such that $Dom_f = Dom_{f_1} \cap Dom_{f_2}$, $f(H, E, B) = g(f_1(H, E, B), f_2(H, E, B))$ for all $\langle H, E, B\rangle \in Dom_f$, and $g$ is non-decreasing in both and increasing in at least one of its arguments $f_1$ and $f_2$, where $R_{f_1} \subseteq \Re$ is the range of $f_1$ and $R_{f_2} \subseteq \Re$ is the range of $f_2$.

**Observation 2** Let $f$ be an evidence based strength indicator on $\mathcal{L}$. Then we have for all $\langle H, E, B\rangle$, $\langle \neg H, E, B\rangle$, $\left\langle H', E, B\right\rangle \in Dom_f$:

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

Let $f$ be an evidence neglecting strength indicator on $\mathcal{L}$. Then we have for all $\langle H, E, B\rangle$, $\langle \neg H, E, B\rangle$, $\left\langle H', E, B\right\rangle \in Dom_f$:

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

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

The requirement takes into account that the set of possibilities falsified by a hypothesis $H$ is a subset of the set of possibilities ruled out by any theory logically implying $H$. The set of possibilities is again 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 tuples $\langle H, E, B\rangle$ are in the domain of $f$.

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 (it is inappropriate, so the objection, to ascribe maximal informativeness to any theory logically implying the evidence). Two theories, one might say, both logically implying all of the data can still differ in their informativeness. For instance, consider a complete theory consistent with the data and a theory-like collection of all the data gathered so far.[3] This 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). The discussion of how to measure informativeness will be taken up again when the present paradigm-independent theory is spelt out in terms of subjective probabilities.

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

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

**Definition 4** Let $s$ be a strength indicator, let $t$ be a truth indicator, and let $\beta \in \Re$. A possibly partial function $f\colon \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is *sensitive to informativeness and plausibility in the sense of s and t and with demarcation* $\beta$—or for short: an *s*, *t-function* (*with demarcation* $\beta$)—iff there is a function $g : R_s \times R_t \times X \to \Re$ such that $g$ is a function of, among others, $s$ and $t$, i.e. $f(H,E,B) = g(s(H,E,B), t(H,E,B), x)$ for all $\langle H,E,B\rangle \in Dom_s \cap Dom_t$, such that

1. Continuity
   Any surplus in informativeness succeeds, if the shortfall 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
   $\forall x \in X : \quad g(s_{\max}, t_{\min}, x) = g(s_{\min}, t_{\max}, x) = \beta.$

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. As before, '$R_s$' 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 at other times '$g(s_1,t_1)$', dropping the additional argument place, and at still other times '$f(s_1,t_1)$', treating $f$ as $g(s,t)$.

Continuity implies

3. Weak Continuity

[3] One might want to restrict the term 'theory' to lawlike statements. I do not. Nor do I want to suggest that the collection of all data is lawlike.

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

The difference is that, in its stronger formulation, Continuity requires $\delta$ just to depend on the lower bound $\varepsilon$ of the difference between $s_1$ and $s_2$, and not on the numbers $s_1$ and $s_2$ themselves. Thus, in the case of Weak Continuity, if $s_{1,i} = \frac{1}{i+1} + a$, $a > 0$, and $s_{2,i} = \frac{1}{i+1}$, for $i \in N$, there may be no *common* upper bound $\delta = \delta_{s_{1,i},s_{2,i}}$ by which $t_{1,i}$ must not be smaller than $t_{2,i}$ in order for $g\left(s_{1,i}, t_{1,i}\right) > g\left(s_{2,i}, t_{2,i}\right)$ to hold—the respective upper bounds may be, say, $\delta_i = \frac{1}{n \cdot i}$ for $t_{1,i}$ and $t_{2,i}$. (In case of infinitely many $s_{1,i}$s and $s_{2,i}$s, one cannot always take $\delta = \inf\left\{\delta_{s_{1,i},s_{2,i}} : i \in N\right\}$, because the latter expression may be 0, as is the case in the example.) Continuity demands that $\delta$ depend only on the lower bound $\varepsilon$ by which $s_1$ exceeds $s_2$. Thus, for $s_{1_i}, s_{2,i}$ there must exist a common $\delta$ depending just on the lower bound, say, $\varepsilon = \frac{a}{2}$—there are, of course, uncountably many such $\varepsilon$s for which there exist (not necessarily distinct) $\delta_\varepsilon$s.

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 data one at a time, the plausibility of the $H_i$s in general changes with each new piece of evidence (assuming an evidence based truth indicator). 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 is that the more informative of the two hypotheses, say $H_1$, eventually comes out as the better theory, if $H_1$'s shortfall in plausibility converges to zero (or if $H_1$ becomes more plausible than $H_2$). If the informativeness of the $H_i$ 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* difference between the informativeness values of $H_1$ and $H_2$. One can, however, refer to the minimal difference between the two informativeness values—unless this difference goes itself to 0, in which case $H_1$ should not necessarily come out as the better theory anyway. In case one prefers to work with evidence neglecting strength indicators, one can stick to Weak Continuity.

As just said, the idea behind Continuity is that the more informative of two hypotheses eventually comes out as the better one, if its shortfall in plausibility vanishes. In particular, this should hold if the plausibility of the two hypotheses converges to certainty (more precisely, if their plausibility becomes either arbitrarily close to certainly true or arbitrary close to certainly false).

4. Continuity in Certainty
   Any surplus in informativeness succeeds, if plausibility becomes certainty.
   $\forall \varepsilon > 0 \quad \forall t_i, t_i' \in R_t : t_i, t_i' \to_i \begin{cases} t_{\max} \\ t_{\min} \end{cases} \quad \exists n \forall m \geq n \quad \forall s_m, s_m' \in R_s :$
   $s_m > s_m' + \varepsilon \quad \Rightarrow \quad g(s_m, t_m) > g\left(s_m', t_m'\right).$
5. Weak Continuity in Certainty
   $\forall s_0, s_0' \in R_s : s_0 > s_0' \quad \forall t_i, t_i' \in R_t : t_i, t_i' \to_i \begin{cases} t_{\max} \\ t_{\min} \end{cases} \quad \exists n \forall m \geq n :$
   $g(s_0, t_m) > g\left(s_0', t_m'\right).$

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

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

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

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

Although Continuity implies

7. $s_0 > s_1 \quad \& \quad t_0 \geq t_1 \quad \Rightarrow \quad g(s_0, t_0) > g(s_1, t_1)$,

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

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

($s_0, s_1$ are any values in $R_s$, and $t_0, t_1$ are any values in $R_t$.)

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. By contrast, 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 or subset relation).

Weak Continuity in Certainty implies

8. Maximality: $g(s_0, t_0) = g_{\max} \quad \Rightarrow \quad s_0 = s_{\max}$
9. Minimality: $g(s_0, t_0) = g_{\min} \quad \Rightarrow \quad s_0 = s_{\min}$.

If we additionally assume Plausibility, we get

10. Maximality II: $g(s_0, t_0) = g_{\max} \quad \Rightarrow \quad s_0 = s_{\max} \quad \& \quad t_0 = t_{\max}$
11. Minimality II: $g(s_0, t_0) = g_{\min} \quad \Rightarrow \quad s_0 = s_{\min} \quad \& \quad t_0 = t_{\min}$.

If we finally add that $g$ is a function of $s$ and $t$ only, we get the converse of 10 and of 11.

The conjunction of Continuity and Demarcation does not imply

$$\text{Symmetry: } g(s_1, t_1) = g(t_1, s_1).$$

Assessment functions may consider one aspect, say plausibility, more important than the other. The only thing that is ruled out is to completely 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 $t = 0$, where $R_s = R_t = [0, 1]$. Furthermore, even if Plausibility is assumed and $g$ is a function of $s$ and $t$ only, the conjunction of Continuity and Demarcation does not imply that for a given value $s_0 \in R_s$ there is a value $t_0 \in R_t$ such that $g(s_0, t_0) = \beta$.

The functions $r$ and $l$ 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 evidence based 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,

$$d = t + s + c, \quad R_t = R_s = \left[0, 1\right],$$

is sensitive to informativeness and plausibility with demarcation $c + 1$, and thus does not exhibit the discontinuity of $r$ and $l$. If $c = -1$, then

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

with $f$ a positive function not depending on $H$, 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.

## 4 Assessing theories, Bayes style

### 4.1 The Bayesian plausibility-informativeness theory

What has been seen so far is the general plausibility-informativeness theory of theory assessment. In a nutshell, its message is (1) that there are two values a theory should exhibit: truth and informativeness—measured respectively by a truth indicator $t$ and a strength indicator $s$; (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 assessed; and (3) that in assessing a given theory one should weigh between these two conflicting aspects in such a way that any surplus in informativeness succeeds, if the shortfall in plausibility is small enough. Particular accounts arise by inserting particular strength indicators and truth indicators.

The theory can be spelt out in terms of Spohn's (1980, 1990) ranking theory (Huber, 2007a), 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 degrees of belief; and (iii) they can be measured by her betting behavior.

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 $\Pr\left(H \mid E \wedge B\right), \ldots$, and the whole problem reduces to the definition of a suitable function of Pr.[4]

In this paradigm it is natural to take

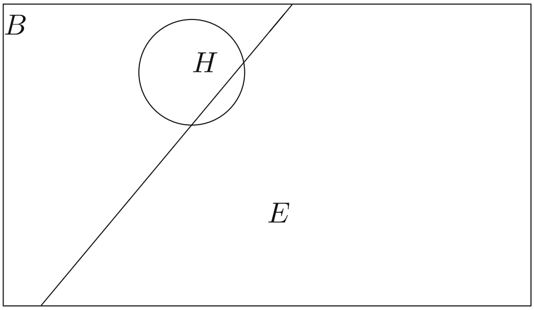
$$t_{\Pr}\left(H, E, B\right) = \Pr\left(H \mid E \wedge B\right) = p$$

as truth indicator, and

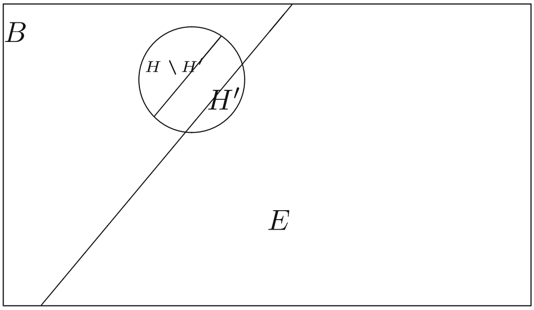
[4] This is not the case in the Hempel paradigm. There the numbers have to be squeezed out of the logical structure of $H$, $E$, and $B$ and nothing else. As a consequence, these values are uniquely determined by $H$, $E$, and $B$ and the logical consequence relation. In particular, they are independent of the underlying language (Huber, 2004).

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

as evidence based and evidence neglecting strength indicators, respectively. Here Pr is a *regular* probability measure on the underlying language or field of propositions.[5] The choice of $p$ hardly needs any discussion. For the choice of $i$ consider the following figure with hypothesis $H$, evidence $E$, and background information $B$ (conceived of as propositions).




Suppose we want to strengthen $H$ by deleting possibilities verifying it, that is, by shrinking the area representing $H$. In this case $i$ recommends to delete possibilities outside $E$. The reason is that, given $E$, those are exactly the possibilities known not to be the actual one, whereas the possibilities inside $E$ are still alive options. Thus, when $H$ shrinks to $H'$ as depicted below, the probabilistic evidence based strength indicator $i$ increases.




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 $i^*_\alpha$ with parameter $\alpha \in [0,1]$ is given by

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

[5] Regularity is often paraphrased as open-mindedness (Earman, 1992), because it demands that no consistent statement be 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 expressible purely in terms of the logical consequence relation are not swept under the carpet.

For $i^*_\alpha$, it depends on $\alpha$ 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[6] are

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

These measures, all of which assign minimal informativeness to any theory entailed by the data and the background assumptions, do pretty bad on this count. They require the deletion of possibilities inside $E$. They measure how much the information in $H$ goes beyond the information provided by $E$. This is not the appropriate notion of informativeness for present purposes, though (see section 4.3 for more on this).

The background information $B$ plays a role different from that of the data $E$ for $i^*_\alpha$, but not for $i''$, cont, or inf. If there is a difference between data on the one hand and background assumptions on the other, then this difference should show up somewhere. According to one view (Hendricks, 2006), background assumptions determine the set of possibilities. Seen this way they are nothing but restrictions on the set of possible worlds over which inquiry has to succeed. Evidence based strength indicators reflect this difference. They measure how much a theory informs about the data, but not how much a theory informs 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.

## 4.2 Incremental confirmation

Let us see how this approach compares to Bayesian confirmation theory. The following notion is central in this literature (Fitelson, 2001).

**Definition 5** A possibly partial function $f = f_{\Pr} : \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is a *$\beta$-relevance measure based on* Pr just in case it holds for all $H, E, B \in \mathcal{L}$ with $\Pr(E \wedge B) > 0$:

$$f(H, E, B) \overset{>}{\underset{<}{=}} \beta \quad \Leftrightarrow \quad \Pr(H \mid E \wedge B) \overset{>}{\underset{<}{=}} \Pr(H \mid B).$$

As

$$\Pr(H \mid E \wedge B) > \Pr(H \mid B) \Leftrightarrow \Pr(\neg H \mid \neg E \wedge B) > \Pr(\neg H \mid B) \tag{1}$$

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

$$s^*_c = p + i^* + c, \quad c \in \Re,$$

[6] Cf. Carnap and Bar-Hillel (1952), Bar-Hillel and Carnap (1953), and Hintikka and Pietarinen (1966). Cf. also Bar-Hilllel (1952, 1955). 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 $E \wedge B$.

is a $c+1$-relevance measure, where $i^*$ is a strength indicator based on $i$ and $i'$. For $c=-1$ and $i^*=i'$, one gets the distance measure $d$,

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

(Earman, 1992). For $c=-1$ and $i^*=i$, 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 3, for positive $f$ not depending on $H$, the functions

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

are $i,p$-functions with demarcation 0. 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[7]:

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

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

So $d$ and $s$ are exactly alike in the way they combine or weigh between informativeness and plausibility. They both 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.

[7] Cf. Hintikka and Pietarinen (1966), Levi (1961, 1963), but also Hempel (1960).

## 4.3 Expected informativeness

What results do we get from the decision-theoretic way of setting confirmation equal to the expected informativeness for the measures $i''$, cont, and inf mentioned in section 4.1? Let '$i''(H)$' be short for '$i''(H, E, B)$', and similarly for '$\text{cont}(H)$' and '$\inf(H)$'.

$$\begin{aligned} E\left(i''(H)\right) &= i''(H) \cdot \Pr(H \mid E \wedge B) - i''(\neg H) \cdot \Pr(\neg H \mid E \wedge B) \\ &= \Pr(\neg H \mid E \wedge B) \cdot \Pr(H \mid E \wedge B) - \\ &\quad - \Pr(H \mid E \wedge B) \cdot \Pr(\neg H \mid E \wedge B) \\ &= 0 \\ E(\text{cont}(H)) &= \text{cont}(H) \cdot \Pr(H \mid E \wedge B) - \text{cont}(\neg H) \cdot \Pr(\neg H \mid E \wedge B) \\ &= \Pr(E) \cdot \Pr(\neg H \mid E \wedge B) \cdot \Pr(H \mid E \wedge B) - \\ &\quad - \Pr(E) \cdot \Pr(H \mid E \wedge B) \cdot \Pr(\neg H \mid E \wedge B) \\ &= 0 \\ E(\inf(H)) &= \inf(H) \cdot \Pr(H \mid E \wedge B) - \inf(\neg H) \cdot \Pr(\neg H \mid E \wedge B) \\ &= -\log_2 \Pr(\neg H \mid E \wedge B) \cdot \Pr(H \mid E \wedge B) + \\ &\quad + \log_2 \Pr(H \mid E \wedge B) \cdot \Pr(\neg H \mid E \wedge B) \\ &\begin{matrix} > \\ = \\ < \end{matrix} 0 \\ &\Leftrightarrow \Pr(H \mid E \wedge B) \begin{matrix} > \\ = \\ < \end{matrix} \Pr(\neg H \mid E \wedge B) \end{aligned}$$

Hence only inf gives a non-trivial answer, viz. to maximize probability. Maximizing probability is also what the "Acceptance rule based on relative-content measure of utility" from Hempel (1960) requires (I have dropped the body of background information $B$, because Hempel does not have it, and I took his content measure $m(\cdot)$ to be $1 - \Pr(\cdot)$, which is in accordance with his remarks on p. 76 of Hempel (1965) and with Hempel (1962) and Hempel and Oppenheim (1948)). Hempel's "Relative-content measure of purely scientific utility" is this:

$$\begin{aligned} \text{rc}(H, E) &= i_H(H, E) \cdot \Pr(H \mid E) - i_H(H, E) \cdot \Pr(\neg H \mid E) \\ &= \frac{\Pr(\neg H \wedge E)}{\Pr(\neg E)} \cdot \Pr(H \mid E) - \frac{\Pr(\neg H \wedge E)}{\Pr(\neg E)} \cdot \Pr(\neg H \mid E) \\ &= \frac{\Pr(\neg H \wedge E)}{\Pr(\neg E)} (2 \cdot \Pr(H \mid E) - 1). \end{aligned}$$

However, as noted by Hintikka and Pietarinen (1966, fn. 12), it seems more adequate to consider

$$\begin{aligned} E(i_H(H, E)) &= i_H(H, E) \cdot \Pr(H \mid E \wedge B) - i_H(\neg H, E) \cdot \Pr(\neg H \mid E \wedge B) \\ &= \frac{\Pr(\neg H \wedge E)}{\Pr(\neg E)} \cdot \Pr(H \mid E) - \frac{\Pr(H \wedge E)}{\Pr(\neg E)} \cdot \Pr(\neg H \mid E) \\ &= 0. \end{aligned}$$

Given this result, it is clear why Hintikka and Pietarinen (1966) choose $i' = \Pr(\neg H)$ as measure of information, and thus arrive at the distance measure $d$ as shown above.

Forming assessment values by taking the expected informativeness is thus allowed, but not required by the Bayesian version of the plausibility-informativeness theory. Here is the expected informativeness for the measures $i^*_\alpha, \alpha \in [0,1]$:

$$E\left(i^*_\alpha(H,E,B)\right) = \alpha \cdot s_{\Pr} + (1-\alpha) \cdot d_{\Pr}.$$

## 4.4 Is it likely to be lovely?

Lipton (2004) suggests the view that a theory which is lovely in his sense (which provides a lot of good explanations) is also likely to be true. Loveliness, as understood here, is an indicator of the informativenss of a theory, and thus need not have anything to do with explanation. Still, one might ask whether "it is likely to be lovely".

The first way to make this question more precise is to ask whether, given no data at all, a lovely theory is also a likely one. This is, of course, not the case, as is clear from the fact that loveliness and likeliness 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 assessed. However, the equivalence in (1) gives rise to another way of putting this question. Given that a piece of evidence $E$ raises the loveliness of $H$ relative to $B$, does that piece of evidence also raise the likeliness of $H$ relative to $B$?[8]

Let $E_0, \ldots, E_{n-1}, E_n$ be the evidence seen up to stage $n+1$ of the inquiry. Then the answer is affirmative if, at stage $n+1$, one considers the *total available evidence* $E = E_0 \wedge \cdots \wedge E_{n-1} \wedge E_n$ and asks whether the likeliness of $H$ given $E$ and background information $B$ is greater than the likeliness of $H$ at stage 0 before the first datum came in, i.e. whether

$$\Pr(H \mid E \wedge B) > \Pr(H \mid B).$$

As we have seen, this holds just in case the loveliness of $H$ relative to $E$ and $B$, $\Pr(\neg H \mid \neg E \wedge B)$, is greater than $H$'s loveliness at stage 0, when it may be set equal to $\Pr(\neg H \mid B)$.[9] So on the global scale, lovely theories are likely to be true. However, the answer is negative on the local scale where one considers just the single datum $E_n$. At stage $n$, the loveliness and the likeliness of $H$ relative to $B$ and the data seen so far are given by

$$s_n = \Pr\left(\neg H \mid \neg\left(E_0 \wedge \cdots \wedge E_{n-1}\right) \wedge B\right), t_n = \Pr\left(H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge B\right).$$

Now suppose the next datum $E_n$ at stage $n+1$ raises the loveliness of $H$ relative to $B$ and the data seen so far, $s_{n+1} > s_n$, i.e.

$$\Pr\left(\neg H \mid \neg\left(E_0 \wedge \cdots \wedge E_{n-1} \wedge E_n\right) \wedge B\right) > \Pr\left(\neg H \mid \neg\left(E_0 \wedge \cdots \wedge E_{n-1}\right) \wedge B\right).$$

Does it follow that $t_{n+1} > t_n$, i.e.

$$\Pr\left(H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge E_n \wedge B\right) > \Pr\left(H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge B\right)?$$

[8] According to $i'$, the informativeness of a theory is independent of the data, and so it does not make sense to ask whether a piece of evidence $E$ raises the loveliness—in the sense of $i'$—of some hypothesis $H$ relative to a body of background information $B$. Therefore only $i$ is considered in the following.

[9] It may justifiedly be argued that the loveliness of $H$ at stage 0 before the first datum came in is *not* $\Pr(\neg H \mid B)$, but rather is not defined. This follows if the "empty datum", i.e. the one before the first datum came in, is represented by $\top$. Stipulating that $s_0$ is defined and equal to $\Pr(\neg H \mid B)$ should only enable me to make sense of the question whether it is likely to be lovely.

It does not. What holds true is that

$$t_{n+1} > t_n$$
$$\Leftrightarrow$$
$$\Pr\left(\neg H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge \neg E_n \wedge B\right) > \Pr\left(\neg H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge B\right),$$

given that the relevant probabilities are non-negative. But $t_{n+1}$ may be smaller than $t_n$, even if $s_{n+1} > s_n$.[10] Thus, although on the global scale a lovely theory is also a likely one, this does not hold true on the local scale, where single pieces of evidence are considered.

## 5 The logic of theory assessment

In Huber (2007b, sct. 6) I briefly indicate how the plausibility-informativeness theory sheds new light on some problems in the philosophy of science. Here I will restrict myself to a discussion of Hempel's conditions of adequacy and the question of a logic of confirmation or theory assessment. This topic is treated in more detail in Huber (2007a) and Huber (submitted).

### 5.1 Hempel's conditions of adequacy

In his "Studies in the Logic of Confirmation" (1945) Carl G. Hempel presents the following conditions of adequacy for any relation of confirmation $|\!\sim\ \subseteq \mathcal{L} \times \mathcal{L}$ on some language $\mathcal{L}$ (the names of 3.1 and 3.2 are not used by Hempel):

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

[10] The same holds true on both the local *and the global* scale, if one takes the measure $i'' = \Pr(\neg H \mid E \wedge B)$ instead of $\Pr(\neg H \mid \neg E \wedge B)$. The reason is that

$$\Pr\left(\neg H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge E_n \wedge B\right) < \Pr\left(\neg H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge B\right) \quad \text{and}$$
$$\Pr(\neg H \mid E \wedge B) < \Pr(\neg H \mid B),$$

if

$$\Pr\left(H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge E_n \wedge B\right) > \Pr\left(\neg H \mid E_0 \wedge \cdots \wedge E_{n-1} \wedge B\right) \quad \text{and}$$
$$\Pr(H \mid E \wedge B) > \Pr(H \mid B),$$

respectively. Though $i''$ is a decreasing function of the logical strength of $H$, it is not an evidence based strength indicator in the sense defined, because $\Pr(\neg H \mid E \wedge B)$ need not equal 1 if $H, B \vdash E$. Moreover, according to the $i'', p$-function $s_c'' = i'' + p + c$, every theory $H$ has the same value $c + 1$ independently of the given evidence $E$ and background information $B$.

As I learned in September 2003, Levi (personal correspondence) now favors $i' = \Pr(\neg H \mid B)$ as a measure of the informativeness of $H$ given $B$. According to this measure, informativeness is a virtue of a theory $H$ relative to background information $B$ which is independent of the data $E$. This is not true for $i = \Pr(\neg H \mid \neg E \wedge B)$. Interestingly $i'$ violates a condition of adequacy Levi himself holds (Levi 1986): any two theories which are logically equivalent given evidence $E$ and background knowledge $B$ should be assigned the same value. This condition does not hold of $i, p$-functions and has the consequence that any two refuted theories are assigned the same value. Given the history of science, this is inappropriate for a theory of theory assessment.

3.2 Consistent Selectivity: $E \nvdash \bot, \quad E \mid\!\sim H \quad \Rightarrow \quad E \nvdash \neg H$
4. Converse Consequence Condition: $E \mid\!\sim H, \quad H' \vdash H \quad \Rightarrow \quad E \mid\!\sim H'$

Hempel then shows that 1, 2, and 4 entail that every sentence (observation report) $E$ confirms every sentence (hypothesis or theory) $H$, 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.[11] One reason seems to be that up to now the predominant view on Hempel's conditions is the analysis Carnap gives in §87 of his *Logical Foundations of Probability* (1962).

## 5.2 Carnap's analysis 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, p. 475; notation adapted)

In discussing the Consistency Condition, Carnap mentions that

> 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, pp. 476–477; emphasis in the original)

which is the concept of "initially confirming evidence", as Carnap calls it in §86 of his (1962), that he explicates by positive probabilistic relevance.

> But 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, pp. 477–478.)

Carnap's analysis can be summarized as follows. In presenting his first three conditions of adequacy Hempel was mixing up two distinct concepts of confirmation, two distinct explicanda in Carnap's terminology. The first concept is explicated by incremental confirmation (positive probabilistic relevance) according to which $E$ incrementally confirms $H$ iff $\Pr(H \mid E) > \Pr(H)$. The second concept is explicated by absolute confirmation according to which $E$ absolutely confirms $H$ iff $\Pr(H \mid E) > r$, for some $r \in [.5, 1)$. The special versions of Hempel's second and third conditions hold true for the second explicatum, 2.1 and 3.1, respectively, but they do not hold true for the first explicatum. On the other hand, Hempel's first condition 1 holds true for the first

[11] The exceptions I know of are Flach (2000), Milne (2000), and Zwirn and Zwirn (1996). Roughly, Zwirn and 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 them, viz. confirmatory and explicatory induction (corresponding to Hempel's conditions 1–3 and 4, respectively). Finally, 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.

explicatum, 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 one explicandum (explicated by incremental confirmation) for the Entailment Condition. Then he had in mind another explicandum (explicated by absolute confirmation) for the Special Consequence and the Special Consistency Conditions. And when Hempel finally 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 this third explicandum might have been.

### 5.3 Hempel vindicated

As to Hempel's Entailment Condition, note that it is satisfied by 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 (this proviso can be dropped by using Popper measures instead of classical probabilities). So the following more charitable reading of Hempel seems plausible. When presenting his first three conditions, Hempel had in mind Carnap's second explicandum that Carnap explicates by 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 3). Consequently he abandoned the informativeness concept of confirmation in favor of the plausibility concept aiming at true theories.

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 that $\Pr(H' \mid E) < r$ whenever $\Pr(H \mid E) > r$ and $H \vdash \neg H'$ 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 (cf. pp. 103-107, esp. pp. 104–105 of his *Studies*), 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.

### 5.4 The logic of theory assessment

As we have seen, HD says that a good theory is informative, whereas IL says good theories are probable or true. According to the above analysis, the driving force behind Hempel's conditions is the idea that *a good theory is both true and informative*. Hempel can thus be seen as the champion of the plausibility-informativeness theory.

As I will show now we can have his cake and eat it too. There is a logic that takes into account both of these two conflicting concepts.

According to the logic of theory assessment (Huber, 2007a), $H$ 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[12] this means

$$E \mid\!\sim_{i^*,\mathrm{Pr}} H \Leftrightarrow \begin{array}{l} \Pr(H \mid E) \geq \Pr(\neg H \mid E) \quad \& \quad i^*(H,E) > i^*(\neg H,E), \\ \text{or} \\ \Pr(H \mid E) > \Pr(\neg H \mid E) \quad \& \quad i^*(H,E) \geq i^*(\neg H,E), \end{array}$$

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. $\mid\!\sim_{i^*,\mathrm{Pr}}$ is the ($i^*$-) assessment relation induced by Pr on $\mathcal{L}$.

The term 'accept' is used as a qualitative counterpart to the quantitative assessment value, and not in the sense of 'believe' or 'hold to be true'. Loosely speaking, the logic of theory assessment has it that the attitude towards hypotheses is like the attitude towards bottles of wine. One would like to buy a good bottle of wine for a small price. On the one hand, one wants to spend as little money as possible (one's theory should be as plausible or riskless as possible). On the other hand, one wants to drink reasonably good wine (one's theory should be sufficiently informative). Sometimes one need not care much about the quality of the wine (say, when one is mixing it with juice anyway), and the main focus is on the price — 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. Usually, though, quality does matter. Likewise, in normal situations the most plausible theories just won't do, because they are too uninformative to answer our questions.

The trade-off between price and quality characterizes a pool of candidate bottles of wine from which to choose. Call them *favorable deals*. For instance, a good bottle of wine for free is a favorable deal. And if a bottle of wine is a favorable deal, then so is any equally good or better bottle for the same price or less. The logic of theory assessment similarly characterizes the pool of *acceptable* hypotheses. For instance, a sufficiently informative theory that is certainly true is acceptable. And if a theory is acceptable, then so is any equally or more informative theory that is equally or more plausible.

Another, perhaps more natural way of defining a qualitative counterpart to the quantitative assessment value is to say that $H$ is acceptable relative to $E$ iff the overall assessment value of $H$ relative to $E$ is greater than that of its negation. The reason why I prefer the stronger notion of acceptability is that the weaker notion is heavily dependent on the way one weighs between informativeness and plausibility. Note, though, that there may be hypotheses $H_1$, $H_2$, data $E$, and assessment functions $a$ such that $H_1$ is an acceptable theory for $E$, but $H_2$ is not, even though, relative to $E$, $a$ assigns a greater assessment value to $H_2$ than to $H_1$.[13]

12 The logic of theory assessment in Huber (2007a) is spelt out in terms of ranking functions. While there are many formal parallels between ranking functions and probability measures, there are also important conceptual differences. One of them is that, conceptually, the rank-theoretic notion of acceptability is weaker than its probabilistic counterpart. In the probabilistic case an acceptable $H$ cannot have a probability of less than .5, which is a requirement that is hardly ever satisfied in examples from science. In the rank-theoretic case an acceptable $H$ merely cannot be disbelieved.

13 I am grateful to Alexander Moffett for pointing this out to me at *FEW 2004* in Berkeley, CA.

Let us see how acceptability relates to Carnap's concept of qualitative confirmation. Positive probabilistic relevance between $E$ and $H$ is necessary in order for $H$ to be an acceptable theory for $E$. Here is why. First,

$$i^*(H,E) \mathrel{\substack{>\\ \geq}} i^*(\neg H,E) \Rightarrow \begin{array}{c} \Pr(\neg H \mid \neg E) \mathrel{\substack{>\\ \geq}} \Pr(H \mid \neg E) \\ \text{or} \\ \Pr(\neg H) \mathrel{\substack{>\\ \geq}} \Pr(H)\,. \end{array}$$

Second,

$$\begin{array}{c} \left[\begin{array}{lcll} \Pr(H \mid E) \geq \Pr(\neg H \mid E) & \& & \Pr(\neg H \mid \neg E) > \Pr(H \mid \neg E)\,, & \text{or} \\ \Pr(H \mid E) > \Pr(\neg H \mid E) & \& & \Pr(\neg H \mid \neg E) \geq \Pr(H \mid \neg E) & \end{array}\right] \\ \text{or} \\ \left[\begin{array}{lcll} \Pr(H \mid E) \geq \Pr(\neg H \mid E) & \& & \Pr(\neg H) > \Pr(H)\,, & \text{or} \\ \Pr(H \mid E) > \Pr(\neg H \mid E) & \& & \Pr(\neg H) \geq \Pr(H) & \end{array}\right] \end{array}$$

entails

$$\Pr(H \mid E) > \Pr(H)\,.$$

However, the converse is not true, because positive probabilistic relevance is symmetric, whereas acceptability is not—which, as noted by Christensen(1999, 437f), is as it should be.

## 6 What is the point?

### 6.1 Revealing the true assessment structure

An important question a theory of theory assessment faces is this. What is the point of having theories that are given high assessment values? That is, why are theories with high assessment values better than other theories? In terms of confirmation the question is: what is the point of having well confirmed theories? That is, why should we stick to well confirmed theories rather to any other theories?[14]

The traditional answer to this question is that science aims at truth, and that one should stick to well confirmed theories because, in the long run, confirmation takes one to the truth. Yet, as we have seen, truth is only one side of the coin. Therefore, a different answer is called for. It will be that, as epistemic agents, we (all of us, not only scientists) aim at informative truth, and that we should stick to theories with high assessment values because, in the medium run, theory assessment takes us to the most informative among all true theories.

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

- $H_1$ is true in $\omega$, but $H_2$ is false in $\omega$; or
- $H_1$ and $H_2$ have the same truth value in $\omega$, but $H_1$ logically implies $H_2$, whereas $H_2$ does not logically imply $H_1$.

In case $H$ is logically false, it is worse in $\omega$ than every contingent theory that is true in $\omega$ (because they are all true in $\omega$, whereas $H$ is false in $\omega$). However, $H$ is better than

[14] I discuss this question for absolute and incremental confirmation in Huber (2005).

every contingent theory that is false in $\omega$ (because $H$ is more informative than any one of them). Similarly, if $H$ is logically true, it is worse in $\omega$ than every contingent theory that is true in $\omega$ (because they all are more informative than $H$), but better than every contingent theory that is false in $\omega$ (because they all are false in $\omega$, whereas $H$ is true in $\omega$). Let us define accordingly.

**Definition 6** A possibly partial function $f\colon \mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ *reveals the true assessment structure in world $\omega$ with demarcation $\beta$* iff for any hypotheses $H, H_1, H_2 \in \mathcal{L}$, every body of background information $B \in \mathcal{L}$ which is true in $\omega$, and any data stream $e_0, \ldots, e_n, \ldots, e_i \in \mathcal{L}$, from $\omega$ (i.e. a sequence of sentences all of which are true in $\omega$):

1. If $H_1$ is contingently true in $\omega$ and $H_2$ is contingently false in $\omega$, then there is $n$ such that for all $m \geq n$: $f(H_1, E_m, B) > \beta > f(H_2, E_m, B)$.
2. If $H_1$ and $H_2$ are contingently true in $\omega$, but $H_1$ is logically stronger than $H_2$, then there is $n$ such that for all $m \geq n$: $f(H_1, E_m, B) > f(H_2, E_m, B) > \beta$.
3. If $H_1$ and $H_2$ are contingently false in $\omega$, but $H_1$ is logically stronger than $H_2$, then there is $n$ such that for all $m \geq n$: $\beta > f(H_1, E_m, B) > f(H_2, E_m, B)$.
4. If $H$ is logically determined, then it holds for all $m$: $f(H, E_m, B) = \beta$.

Here $E_m = e_0 \wedge \cdots \wedge e_{m-1}$.

So $f$ must *stabilize* to the correct answer. That is, $f$ must 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).[15] The smallest $n$ for which the above holds is called *the point of stabilization*.

The central question is whether assessment functions do in fact reveal the true assessment structure and thus lead to informative true theories. As shown in more detail below, the answer is affirmative: 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.[16]

## 6.2 Making the point more precise

This section develops the claim of the last section. The framework adopted here is that of Gaifman & Snir (1982). $\mathcal{L}_0$ is a first order language for arithmetic. It contains all numerals '1', '2', ... as individual constants, and countably many individual variables '$x_1$', ... taking values in the set of natural numbers $N$. Furthermore, $\mathcal{L}_0$ contains the common symbols '+', '·', and '=' for addition, multiplication, and identity, respectively. In addition, there may be finitely many predicates and function symbols

[15] Stabilization to the correct answer is a stronger requirement than convergence to the correct answer (see Kelly, 1996). The latter is a bit odd to formulate for revealing the true assessment structure. In general it says that for any $\varepsilon > 0$ (as small as you like) there exists a point $n$ (depending on $\varepsilon$) such that for all later points $m > n$, $f$'s conjecture differs form "the truth" only by an amount smaller than $\varepsilon$. The difference between stabilization and convergence was the reason for appealing to the medium run (stabilization) as compared to the long run (convergence). Note, however, that the Gaifman and Snir convergence theorem can be used to obtain an almost-sure stabilization result by assigning 1 to $H$, if the probability of $H$ is above .5 (or any other positive threshold that is smaller than 1), and 0 otherwise (cf. section 7).

[16] The result stated below holds only for almost every world and is restricted to data sequences that separate $Mod_{\mathcal{L}}$. This flaw is serious (Kelly, 1996, ch. 13), but not inevitable. 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. It is fair to say that measure one results are not problematic in this case.

denoting certain fixed relations over $N$. Finally, $\mathcal{L}_0$ contains the quantifiers '$\forall$', '$\exists$', the unary sentential connective '$\neg$', and the binary sentential connectives '$\wedge$', '$\vee$', '$\rightarrow$', and '$\leftrightarrow$'. The language $\mathcal{L}$ is obtained from $\mathcal{L}_0$ by adding finitely many predicates and function symbols.

A model $\omega$ for $\mathcal{L}$ consists of an interpretation $\varphi$ of the empirical symbols which assigns every $k$-ary predicate '$P$' a subset $\varphi(\text{'}P\text{'}) \subseteq N^k$, and every $k$-ary function symbol '$f$' a function $\varphi(\text{'}f\text{'})$ from $N^k$ to $N$. The interpretation of the symbols in $\mathcal{L}_0$ is the standard one and is kept the same in all models. $Mod_{\mathcal{L}}$ is the set of all models for $\mathcal{L}$. '$\omega \models A$' says that formula $A$ is true in model $\omega \in Mod_{\mathcal{L}}$. $A[x_1,\ldots,x_k]$ is valid, $\models A[x_1,\ldots,x_k]$, iff $\omega \models A[n_1/x_1,\ldots,n_k/x_k]$ for all $\omega \in Mod_{\mathcal{L}}$ and all numerals $n_1,\ldots,n_k \in L_0$. Here, '$A[n_1/x_1,\ldots,n_k/x_k]$' results from '$A[x_1,\ldots,x_k]$' by uniformously substituting '$n_i$' for '$x_i$' in '$A$', $1 \leq i \leq k$. '$A[x_1,\ldots,x_k]$' indicates that '$x_1$', …, '$x_k$' are the only individual variables occurring free in '$A$'.

**Definition 7** A function $\Pr\colon \mathcal{L} \rightarrow \Re_{\geq 0}$ is a *probability on* $\mathcal{L}$ iff for all $A, B \in \mathcal{L}$:

1. $\models A \leftrightarrow B \quad \Rightarrow \quad \Pr(A) = \Pr(B)$
2. $\models A \quad \Rightarrow \quad \Pr(A) = 1$
3. $\models \neg(A \wedge B) \quad \Rightarrow \quad \Pr(A \vee B) = \Pr(A) + \Pr(B)$
4. $\Pr(\exists x A[x]) = \sup\left\{\Pr(A[n_1/x] \vee \ldots \vee A[n_k/x]) : n_1,\ldots,n_k \in N, k \in N_{\geq 1}\right\}$

Iff $\Pr(B) > 0$, the conditional probability $\Pr(\cdot \mid A) : \mathcal{L} \rightarrow \Re_{\geq 0}$ based on the probability $\Pr(\cdot) : \mathcal{L} \rightarrow \Re_{\geq 0}$ is defined as

$$\Pr(A \mid B) = \frac{\Pr(A \wedge B)}{\Pr(B)}.$$

A set of sentences $S$ *separates* a set of models $X \subseteq Mod_{\mathcal{L}}$ just in case for any two distinct $\omega_1, \omega_2 \in X$ there is a sentence $A \in S$ such that $\omega_1 \models A$ and $\omega_2 \not\models A$. The set of all atomic empirical sentences separates $Mod_{\mathcal{L}}$. Gaifman & Snir (1982, p. 507) prove the following theorem.

**Theorem 1** (Gaifman and Snir Convergence Theorem) *Let the set of sentences $S = \{A_i : i = 0, 1, \ldots\}$ separate $Mod_{\mathcal{L}}$, and let $[B](\omega)$ be 1 if $\omega \models B$ and 0 otherwise. Then for every $B \in \mathcal{L}$:*

$$\Pr\left(B \mid \bigwedge_{0 \leq i < n} A_i^{\omega}\right) \rightarrow [B](\omega) \textit{ almost everywhere as } n \rightarrow \infty.$$

Based on the Gaifman and Snir convergence theorem we can now prove

**Theorem 2** *Let $e_0,\ldots,e_n,\ldots$ be a sequence of sentences of $\mathcal{L}$ which separates $Mod_{\mathcal{L}}$, and let $e_i^{\omega}$ be $e_i$, if $\omega \models e_i$, and $\neg e_i$ otherwise, where $\omega \in Mod_{\mathcal{L}}$. Let* Pr *be a regular probability on $\mathcal{L}$, and let a be a function of, among others, i and p which satisfies Continuity in Certainty and Demarcation for i and p. Finally, let* $\Pr^*$ *be the unique probability measure on the smallest $\sigma$-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 $Mod(A) = \{\omega \in Mod_{\mathcal{L}} : \omega \models A\}$. Then there exists $X \in \mathcal{A}$ with $\Pr^*(X) = 1$ such that the following holds for every $\omega \in X$, any two contingent $H_1, H_2 \in \mathcal{L}$, and every $H \in \mathcal{L}$:*

1. $\omega \models H_1, \omega \not\models H_2 \quad \Rightarrow \quad \exists n \forall m \geq n : a\left(H_1, E_m^{\omega}\right) > \beta > a\left(H_2, E_m^{\omega}\right)$
2. $\omega \models H_1, H_1 \vdash H_2 \not\vdash H_1 \Rightarrow \exists n \forall m \geq n : a\left(H_1, E_m^{\omega}\right) > a\left(H_2, E_m^{\omega}\right) > \beta$

3. $\omega \not\models H_2, H_1 \vdash H_2 \nvdash H_1 \Rightarrow \exists n \forall m \geq n : \beta > a\left(H_1, E_m^\omega\right) > a\left(H_2, E_m^\omega\right)$
4. $\models H \quad or \quad \models \neg H \quad \Rightarrow \quad \forall m : a\left(H, E_m^\omega\right) = \beta.$

*Proof* 1. Assume the conditions stated in theorem 2, and suppose $\omega \models H_1$ and $\omega \not\models H_2$, where $\omega \in X'$ for some $X' \in \mathcal{A}$ with $\Pr^*\left(X'\right) = 1$ such that for all $B \in \mathcal{L}$ and all $\omega' \in X'$:

$$\Pr\left(B \mid E_n^{\omega'}\right) \to [B]\left(\omega'\right) \quad \text{as} \quad n \to \infty$$

(such $X'$ exists by the Gaifman and Snir convergence theorem). So

$$\Pr\left(H_1 \mid E_n^\omega\right) \to 1 \quad \text{as} \quad n \to \infty, \quad \text{and} \quad \Pr\left(H_2 \mid E_n^\omega\right) \to 0 \quad \text{as} \quad n \to \infty.$$

First, observe that there exists $n_1$ such that for all $m \geq n_1$:

$$\Pr\left(\neg H_1 \mid \neg E_m^\omega\right) > \Pr\left(\neg H_1\right) > 0 \quad \text{and} \quad \Pr\left(\neg H_2 \mid \neg E_m^\omega\right) < \Pr\left(\neg H_2\right) < 1.$$

The reason is that Pr is regular, the $H_i$ are contingent ($i = 1, 2$), and, provided $0 < \Pr\left(E_m^\omega\right) < 1$,

$$\Pr\left(\neg H_i \mid \neg E_m^\omega\right) \gtrless \Pr\left(\neg H_i\right) \Leftrightarrow \Pr\left(H_i \mid E_m^\omega\right) \gtrless \Pr\left(H_i\right).$$

If $\Pr\left(E_m^\omega\right) = 0$, then $\Pr^*\left(Mod\left(E_m^\omega\right)\right) = 0$ (Gaifman & Snir, 1982, p. 504, Basic Fact 1.3). The union of all such sets $Mod\left(E_m^\omega\right)$ of probability 0 is also of probability 0 (there are just countably many such sets), i.e.

$$\Pr{}^*(A) = 0, \quad A := \bigcup \left\{Mod\left(E_m^\omega\right) \in \mathcal{A} : \Pr{}^*\left(Mod\left(E_m^\omega\right)\right) = 0\right\} \in \mathcal{A}.$$

Similarly, if $\Pr\left(E_m^\omega\right) = 1$, then $\Pr^*\left(Mod\left(\neg E_m^\omega\right)\right) = 0$. The union of all such sets $Mod\left(\neg E_m^\omega\right)$ of probability 0 is also of probability 0, i.e.

$$\Pr{}^*(B) = 0, \quad B := \bigcup \left\{Mod\left(\neg E_m^\omega\right) \in \mathcal{A} : \Pr{}^*\left(Mod\left(E_m^\omega\right)\right) = 1\right\} \in \mathcal{A}.$$

As a consequence, $X := X' \setminus (A \cup B) \in \mathcal{A}$ and $\Pr^*(X) = 1$. Assume therefore that $\omega \in X$. As $\Pr\left(H_1 \mid E_n^\omega\right) \to_n 1$ and $\Pr\left(H_2 \mid E_n^\omega\right) \to_n 0$, there is $n_1$ such that for all $m \geq n_1$:

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

and thus

$$\Pr\left(\neg H_1 \mid \neg E_m^\omega\right) > \Pr\left(\neg H_1\right) > 0 \quad \text{and} \quad \Pr\left(\neg H_2 \mid \neg E_m^\omega\right) < \Pr\left(\neg H_2\right) < 1.$$

Hence

$$\Pr\left(\neg H_1\right) \leq \inf_{m \geq n_1} \left\{\Pr\left(\neg H_1 \mid \neg E_m^\omega\right)\right\}, \quad \Pr\left(\neg H_2\right) \geq \sup_{m \geq n_1} \left\{\Pr\left(\neg H_2 \mid \neg E_m^\omega\right)\right\}.$$

By Continuity in Certainty, for $\varepsilon = \frac{\Pr(\neg H_1)}{2}$ and the sequences $t_i = \Pr\left(H_1 \mid E_i^\omega\right)$ and $t_i' = 1$ with $t_i, t_i' \to_i t_{\max} = 1$ there exists $n_2$ such that for all $m \geq n_2$ and all $s_m, s_m' \in R_s = [0, 1]$:

$$s_m > s_m' + \varepsilon \quad \Rightarrow \quad a\left(s_m, t_m\right) > a\left(s_m', t_m'\right).$$

For $s_i = \Pr\left(\neg H_1 \mid \neg E_i^\omega\right)$ and $s_i' = 0$ we thus get for every $m \geq \max\left\{n_1, n_2\right\}$: $a\left(s_m, t_m\right) > a(0, 1) = \beta$. Similarly, for $\varepsilon = \frac{1 - \Pr(\neg H_2)}{2}$ and the sequences $t_i = 0$ and $t_i' = \Pr\left(H_2 \mid E_i^\omega\right)$

with $t_i, t'_i \to_i t_{\min} = 0$ there exists $n_3$ such that for all $m \geq n_3$ and all $s_m, s'_m \in R_s = [0,1]$:

$$s_m > s'_m + \varepsilon \quad \Rightarrow \quad a\left(s_m, t_m\right) > a\left(s'_m, t'_m\right).$$

For $s_i = 1$ and $s'_i = \Pr\left(\neg H_2 \mid \neg E^\omega_i\right)$ we thus get for every $m \geq \max\{n_1, n_3\}$: $\beta = a(1,0) > a\left(s_m, t_m\right)$. Hence for every $m \geq \max\{n_1, n_2, n_3\}$:

$$a\left(\Pr\left(\neg H_1 \mid \neg E^\omega_m\right), \Pr\left(H_1 \mid E^\omega_m\right)\right) > \beta > a\left(\Pr\left(\neg H_2 \mid \neg E^\omega_m\right), \Pr\left(H_2 \mid E^\omega_m\right)\right).$$

2. Suppose now that $\omega \models H_1$, $\omega \models H_2$, and $H_1 \vdash H_2 \nvdash H_1$, where $\omega \in X$ for some $X \in \mathcal{A}$ as before. So

$$\Pr\left(H_1 \mid E^\omega_n\right) \to 1 \quad \text{as} \quad n \to \infty, \quad \text{and} \quad \Pr\left(H_2 \mid E^\omega_n\right) \to 1 \quad \text{as} \quad n \to \infty,$$

and we can safely assume that $0 < \Pr\left(E^\omega_m\right) < 1$ for all $m$. As before, there exists $n_1$ such that for all $m \geq n_1$: $\Pr\left(\neg H_2 \mid \neg E^\omega_m\right) > \Pr\left(\neg H_2\right) > 0$. Observe that

$$\begin{aligned}
\Pr\left(\neg H_1 \mid \neg E^\omega_m\right) - \Pr\left(\neg H_2 \mid \neg E^\omega_m\right) &= \frac{1 - \Pr\left(H_1\right) - \Pr\left(E^\omega_m\right) + \Pr\left(H_1 \wedge E^\omega_m\right)}{\Pr\left(\neg E^\omega_m\right)} - \\
&\quad - \frac{1 - \Pr\left(H_2\right) - \Pr\left(E^\omega_m\right) + \Pr\left(H_2 \wedge E^\omega_m\right)}{\Pr\left(\neg E^\omega_m\right)} \\
&= \frac{\Pr\left(H_2\right) - \Pr\left(H_1\right)}{\Pr\left(\neg E^\omega_m\right)} - \\
&\quad - \frac{\left[\Pr\left(H_2 \mid E^\omega_m\right) - \Pr\left(H_1 \mid E^\omega_m\right)\right] \cdot \Pr\left(E^\omega_m\right)}{\Pr\left(\neg E^\omega_m\right)} \\
&> \frac{\Pr\left(H_2\right) - \Pr\left(H_1\right)}{\Pr\left(\neg E^\omega_m\right)} - \\
&\quad - \frac{\Pr\left(H_2 \mid E^\omega_m\right) - \Pr\left(H_1 \mid E^\omega_m\right)}{\Pr\left(\neg E^\omega_m\right)}.
\end{aligned}$$

By the above, for $\varepsilon = \frac{\Pr(H_2) - \Pr(H_1)}{2} > 0$ there exists $n_\varepsilon$ such that for all $m \geq n_\varepsilon$: $\Pr\left(H_2 \mid E^\omega_m\right) - \Pr\left(H_1 \mid E^\omega_m\right) < \varepsilon$. Consequently it holds for all $m \geq n_\varepsilon$:

$$\Pr\left(\neg H_1 \mid \neg E^\omega_m\right) - \Pr\left(\neg H_2 \mid \neg E^\omega_m\right) > \frac{2\varepsilon - \varepsilon}{\Pr\left(\neg E^\omega_m\right)} > \varepsilon,$$

i.e. $\Pr\left(\neg H_1 \mid \neg E^\omega_m\right) > \Pr\left(\neg H_2 \mid \neg E^\omega_m\right) + \varepsilon$.

By Continuity in Certainty[17], for $\varepsilon > 0$ and the sequences $t_i = \Pr\left(H_1 \mid E_i^\omega\right)$ and $t_i' = \Pr\left(H_2 \mid E_i^\omega\right)$ with $t_i, t_i' \to_i t_{\max} = 1$ there is $n_2$ such that for all $m \geq n_2$ and all $s_m, s_m' \in R_s = [0,1]$:

$$s_m > s_m' + \varepsilon \quad \Rightarrow \quad a\left(s_m, t_m\right) > a\left(s_m', t_m'\right).$$

For $s_i = \Pr\left(\neg H_1 \mid \neg E_i^\omega\right)$ and $s_i' = \Pr\left(\neg H_2 \mid \neg E_i^\omega\right)$ we thus get for all $m \geq \max\left\{n_1, n_2, n_\varepsilon\right\}$: $a\left(s_m, t_m\right) > a\left(s_m', t_m'\right)$.

It follows from 1 that there is $n_3$ such that for all $m \geq n_3$:

$$a\left(\Pr\left(\neg H_2 \mid \neg E_m^\omega\right), \Pr\left(H_2 \mid E_m^\omega\right)\right) > \beta.$$

Hence for all $m \geq \max\left\{n_1, n_2, n_3, n_\varepsilon\right\}$:

$$a\left(\Pr\left(\neg H_1 \mid \neg E_m^\omega\right), \Pr\left(H_1 \mid E_m^\omega\right)\right) > a\left(\Pr\left(\neg H_2 \mid \neg E_m^\omega\right), \Pr\left(H_2 \mid E_m^\omega\right)\right) > \beta.$$

3. Similarly.
4. This follows from Demarcation $\beta$. □

**Corollary 1** *The same holds true if* $i = \Pr\left(\neg H \mid \neg E\right)$ *is replaced by* $i' = \Pr\left(\neg H\right)$*, even if Continuity in Certainty is weakened to Weak Continuity in Certainty.*

**Corollary 2** *The same holds true if* $i$ *is replaced by any function of* $i$ *and* $i'$ *that is non-decreasing in both arguments, and increasing in at least one.*

The relativization to the body of background information $B$ has been dropped. 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 $\omega \in Mod\left(B\right) \cap X$.

Continuity in its general form is not needed for these theorems to hold. In fact, even Continuity in Certainty is not necessary. The necessary and sufficient condition for revealing the true assessment structure in almost every world when presented separating data is this ($\beta$ is assumed to be 0).

**Definition 8** A possibly partial function $f$: $\mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ is a *Gaifman–Snir* assessment function iff for every Gaifman-Snir language $\mathcal{L}$, every probability Pr on $\mathcal{L}$, and every $\left\{e_i : i \in N\right\} \subseteq \mathcal{L}$ separating $Mod_{\mathcal{L}}$ there is $X \in \mathcal{A}$ with $\Pr^*\left(X\right) = 1$ such that for all $\omega \in X$ and all $m \in N$:

$$\text{I.} \quad \begin{array}{l} H_1 \vdash H_2 \nvdash H_1 \\ \Pr\left(H_1 \mid E_m^\omega\right) \to_m \begin{cases} 1 \\ 0 \end{cases} \end{array} \quad \Rightarrow \quad \exists n \forall m \geq n : f\left(H_1, E_m^\omega\right) > f\left(H_2, E_m^\omega\right).$$

$$\text{II.} \ \vdash H_1, H_2 \vdash \bot, \Pr\left(E_m^\omega\right) > 0 \quad \Rightarrow \quad f\left(H_1, E_m^\omega\right) = f\left(H_2, E_m^\omega\right) = 0.$$

**Definition 9** Let $\mathcal{L}$ be a Gaifman–Snir language, let Pr be a probability on $\mathcal{L}$, and let $\left\{e_i : i \in N\right\} \subseteq \mathcal{L}$ separate $Mod_{\mathcal{L}}$. A possibly partial function $f$: $\mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ *reveals*

[17] It is here where the assumption enters that the $\delta$ in Continuity and the $n$ in Continuity in Certainty depend only on $\varepsilon$, which is a lower bound of the difference between $s_m = \Pr\left(\neg H_1 \mid \neg E_m^\omega\right)$ and $s_m' = \Pr\left(\neg H_2 \mid \neg E_m^\omega\right)$. Otherwise, i.e. when $\delta$ or $n$ depend on $s_m$ and $s_m'$, it is possible that $n_{s_m, s_m'} = m + 1$. In this case there is no $n$ such that for all $m \geq n$:

$$a\left(\Pr\left(\neg H_1 \mid \neg E_m^\omega\right), \Pr\left(H_1 \mid E_m^\omega\right)\right) > a\left(\Pr\left(\neg H_2 \mid \neg E_m^\omega\right), \Pr\left(H_2 \mid E_m^\omega\right)\right).$$

In case of $s_m = \Pr\left(\neg H_1\right)$ and $s_m' = \Pr\left(\neg H_2\right)$ it suffices to assume Weak Continuity in Certainty, because the informativeness values $s_m$ and $s_m'$ do not change with $m$.

*the true assessment structure of* $\Pr^*$*-almost every world* $\omega \in Mod_{\mathcal{L}}$ *when presented separating* $\{e_i\}$ iff there is $X \in \mathcal{A}$ with $\Pr^*(X) = 1$ such that for all $\omega \in X$, all contingent $H_1, H_2 \in \mathcal{L}$, and all $H \in \mathcal{L}$:

1. $\omega \models H_1, \omega \not\models H_2 \quad \Rightarrow \quad \exists n \forall m \geq n : f\left(H_1, E_m^{\omega}\right) > 0 > f\left(H_2, E_m^{\omega}\right).$
2. $\omega \models H_1, H_1 \vdash H_2 \nvdash H_1 \Rightarrow \exists n \forall m \geq n : f\left(H_1, E_m^{\omega}\right) > f\left(H_2, E_m^{\omega}\right) > 0.$
3. $\omega \not\models H_2, H_1 \vdash H_2 \nvdash H_1 \Rightarrow \exists n \forall m \geq n : 0 > f\left(H_1, E_m^{\omega}\right) > f\left(H_2, E_m^{\omega}\right).$
4. $H \vdash \bot \quad \text{or} \quad \vdash H \quad \Rightarrow \quad \forall m : f\left(H, E_m^{\omega}\right) = 0.$

*f reveals the true assessment structure in almost every world when presented separating data* iff for any language $\mathcal{L}$, any probability on $\mathcal{L}$, and any $\{e_i : i \in N\} \subseteq \mathcal{L}$ separating $Mod_{\mathcal{L}}$: $f$ reveals the true assessment structure in $\Pr^*$-almost every world $\omega \in Mod_{\mathcal{L}}$ when presented separating $\{e_i\}$.

**Theorem 3** *A possibly partial function* $f$: $\mathcal{L} \times \mathcal{L} \times \mathcal{L} \to \Re$ *reveals the true assessment structure in almost every world when presented separating data iff f is a Gaifman–Snir assessment function.*

*Proof* Suppose $f$ is a Gaifman–Snir assessment function. Let $\mathcal{L}$ be a language and Pr a probability on $\mathcal{L}$. Suppose $\{e_i : i \in N\} \subseteq \mathcal{L}$ separates $Mod_{\mathcal{L}}$. We show that $f$ reveals the true assessment structure of $\Pr^*$-almost every world $\omega \in Mod_{\mathcal{L}}$ when presented separating $\{e_i\}$. By the Gaifman and Snir convergence theorem, there is $X' \in \mathcal{A}$ with $\Pr^*\left(X'\right) = 1$ such that for all $\omega \in X'$ and all $H \in \mathcal{L}$: $\Pr\left(H \mid E_m^{\omega}\right) \to_m [H](\omega)$. By assumption, there is $X'' \in \mathcal{A}$ such that for all $\omega \in X''$ and all $m \in N$:

$$\text{I.} \quad \begin{array}{l} H_1 \vdash H_2 \nvdash H_1 \\ \Pr\left(H_1 \mid E_m^{\omega}\right) \to_m \begin{cases} 1 \\ 0 \end{cases} \end{array} \quad \Rightarrow \quad \exists n \forall m \geq n : f\left(H_1, E_m^{\omega}\right) > f\left(H_2, E_m^{\omega}\right).$$

$$\text{II.} \quad \vdash H_1, H_2 \vdash \bot, \Pr\left(E_m^{\omega}\right) > 0 \quad \Rightarrow \quad f\left(H_1, E_m^{\omega}\right) = f\left(H_2, E_m^{\omega}\right) = 0.$$

Hence, $X' \cap X''$ is an element of $\mathcal{A}$ with $\Pr^*\left(X' \cap X''\right) = 1$ and such that I and II are satisfied for all $\omega \in X' \cap X''$ and all $m \in N$. Furthermore, $A := \left\{\omega \in X' : \exists m : \Pr\left(E_m^{\omega}\right) = 0\right\}$ is of $\Pr^*$-measure 0, i.e. there is $B \in \mathcal{A}$ with $A \subseteq B$ and $\Pr^*(B) = 0$. Hence $X := \left(X' \cap X''\right) \setminus B$ is an element of $\mathcal{A}$ with $\Pr^*(X) = 1$ such that I and II are satisfied for all $\omega \in X$ and all $m \in N$.

So suppose $\omega \models H_1$, for $\omega \in X$ and contingent $H_1 \in \mathcal{L}$. Then there is $n$ such that for all $m$: $f\left(H_1, E_m^{\omega}\right) > f\left(\top, E_m^{\omega}\right) = 0$. Furthermore, if $\omega \not\models H_2$, for the same $\omega \in X$ and some contingent $H_2 \in \mathcal{L}$, then there is $n$ such that for all $m$: $f\left(H_2, E_m^{\omega}\right) < f\left(\bot, E_m^{\omega}\right) = 0$. If $\omega \models H_1$, for some $\omega \in X$, and $H_1 \vdash H_2 \nvdash H_1$, for contingent $H_1, H_2 \in \mathcal{L}$, then $\Pr\left(H_1 \mid E_m^{\omega}\right) \to_m 1$, and hence $\Pr\left(H_2 \mid E_m^{\omega}\right) \to_m 1$. So there is $n$ such that for all $m \geq n$: $f\left(H_1, E_m^{\omega}\right) > f\left(H_2, E_m^{\omega}\right) > f\left(\top, E_m^{\omega}\right) = 0$. Similarly, if $\omega \not\models H_2$, for some $\omega \in X$, and $H_1 \vdash H_2 \nvdash H_1$, for contingent $H_1, H_2 \in \mathcal{L}$, then $\Pr\left(H_2 \mid E_m^{\omega}\right) \to_m 0$, and hence $\Pr\left(H_1 \mid E_m^{\omega}\right) \to_m 0$. So there is $n$ such that for all $m \geq n$: $0 = f\left(\bot, E_m^{\omega}\right) > f\left(H_1, E_m^{\omega}\right) > f\left(H_2, E_m^{\omega}\right)$. Finally, for all $\omega \in X$: $f\left(H, E_m^{\omega}\right) = 0$ for any logically determined $H$ and all $m$.

Conversely, suppose $f$ reveals the true assessment structure in almost every world when presented separating data. We show that $f$ is a Gaifman–Snir assessment function. Suppose not. Then there exist $\mathcal{L}$, Pr on $\mathcal{L}$, and $\{e_i : i \in N\} \subseteq \mathcal{L}$ separating $Mod_{\mathcal{L}}$ such that for all $X \in \mathcal{A}$ with $\Pr^*(X) = 1$ there is $\omega \in X$ such that:

i. There are $H_1, H_2 \in \mathcal{L}$ with $H_1 \vdash H_2 \nvdash H_1$ and $\Pr\left(H_i \mid E_m^{\omega}\right) \to_m \begin{cases} 1 \\ 0 \end{cases}$ such that for all $n$ there is $m \geq n$: $f\left(H_1, E_m^{\omega}\right) \leq f\left(H_2, E_m^{\omega}\right)$; or

ii. there are logically determined $H \in \mathcal{L}$ and $m$ such that $\Pr\left(E_m^\omega\right) > 0$ and $f\left(H, E_m^\omega\right) \neq 0$.

By the Gaifman and Snir convergence theorem, there is $X \in \mathcal{A}$ with $\Pr^*(X) = 1$ such that for all $\omega \in X$ and all $H \in \mathcal{L}$: $\Pr\left(H \mid E_m^\omega\right) \rightarrow_m [H](\omega)$. For any such $X$ there is $\omega \in X$ such that i or ii hold.

**Case 1** If $\Pr\left(H_i \mid E_m^\omega\right) \rightarrow_m 1$, then $H_1$ and $H_2$ are true in $\omega$. Hence, $H_1$ is contingent and $\Pr\left(E_m^\omega\right) > 0$ for all $m \in N$. If $H_2$ is contingent, then 2 fails; if $H_2$ is logically determined, then 4 fails for $H_2$ or 1 fails for $H_1$. If $\Pr\left(H_i \mid E_m^\omega\right) \rightarrow_m 0$, then $H_1$ and $H_2$ are false in $\omega$. Hence, $H_2$ is contingent and $\Pr\left(E_m^\omega\right) > 0$ for all $m \in N$. If $H_1$ is contingent, then 3 fails; if $H_1$ is logically determined, then 4 fails for $H_1$ or 1 fails for $H_2$.

**Case 2** Obviously 4 fails. □

One reason why I nevertheless stick to the more general Continuity conditions is that it depends on the underlying convergence theorem which conditions are necessary and sufficient for revealing the true assessment structure in so and so many worlds when presented such and such data. More importantly, the idea behind the use of these limit considerations is that they provide a theoretical justification for obeying the proposed normative conditions in the here and now. When assessing several alternative theories we cannot wait until we have arrived at the point of stabilization. We need to make our evaluations when the plausibility and informativeness values are somewhere in between their maximal and minimal values. Continuity tells us what to do in such a situation; Continuity in Certainty does not.

However, I also need to justify this answer. And I do so by appealing to the fact that when we satisfy Continuity in the special case when the plausibility values converge to certainty, we reveal the true assessment structure in almost every world when presented separating data. Of course, as long as the relevant probabilities are non-extreme, this is compatible with any funny behavior in the short run. One response to this objection is to look at the necessary and sufficient conditions for revealing the true assessment structure (in almost every world when presented separating data) *as soon as possible* (Kelly, 1996). Then we vindicate the normative conditions of the plausibility-informativeness theory relative to the goal of eventually arriving at the most informative true theory as soon as possible. Another response is to say that the very fact that we do not know when the point of stabilization occurs is reason enough to always be prepared for it to take place. While I think that only the first answer is conclusive, I cannot offer a proof to the effect that Continuity and Demarcation are necessary and sufficient for eventually arriving at the most informative true theory as soon as possible.

Finally I should mention that the present approach is also viable if truth and informativeness are not the only epistemic values. Whatever these values besides truth are, and however they are measured; if there is a function $u$ such that $u(H, E, B)$ measures the overall value *without* truth of $H$ in view of $E$ and $B$; and if for any two theories $H_1$ and $H_2$, any separating data sequence $e_0, \ldots, e_n, \ldots$ from any world $\omega$, and any body of background information $B$ true in $\omega$ there is a point $j$ such that for all later points $k > j$: $u(H_1, E_k, B) > u(H_2, E_k, B) + \varepsilon$, for some $\varepsilon > 0$; then the following holds for every $f$ satisfying the two conditions corresponding to Continuity in Certainty and Demarcation for $u$ (instead of $s$) and $t$. There is a point $m$ such that for all later points

$n > m$: $f(H_1, E_n, B) > f(H_2, E_n, B)$, where both of these values are greater than $\beta$, if both $H_1$ and $H_2$ are contingently true in $\omega$, both of these values are smaller than $\beta$ if both $H_1$ and $H_2$ are contingently false in $\omega$, and $\beta$ lies between these two values if $H_1$ is contingently true, but $H_2$ is contingently false in $\omega$.

## 7 Relevance measures and their exclusive focus on truth

As shown in the preceding section, all one needs to do to reveal the true assessment structure in almost every world when presented separating data is to stick to a 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 $\beta$-relevance measure?

The connection to the $i,p$-function $s_c = i + p + c$ for $c = -1$, and the function $d_f$ for $f = \Pr(\neg E \mid B)$ respectively $f = \Pr(\neg E \mid B) \cdot \Pr(B) \cdot \Pr(E \wedge B)$ has already been pointed out. So for any strict probability Pr, $s_{\Pr}$ and $c_{\Pr}$ and $d_{\Pr}$ reveal the true assessment structure in almost every world when presented separating data. However, there are many other relevance measures. Do they all further the goal of eventually arriving at the most informative true theory?

If $H_1$ is contingently true in $\omega$, and $H_2$ is contingently false in $\omega$, then, after finitely many steps, $H_1$ has to get a greater value in $\omega$ than the demarcation parameter $\beta$ which in turn has to be greater than the value of $H_2$ in $\omega$. Any $\beta$-relevance measure $r$ reveals this part of almost any $\omega$'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 $m \geq n$:

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

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

$$r\left(H_1, E_m^{\omega}\right) > \beta > r\left(H_2, E_m^{\omega}\right).$$

Moreover, the value (in $\omega$) of any logically determined hypothesis is always equal to $\beta$.

So far, so good. But the definition of a $\beta$-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 $\omega$. This exclusive focus on truth—in contrast to the weighing between the conflicting goals of informativeness and truth of an $s,t$-function—is what prevents relevance measures from revealing the true assessment structure in general. As we have seen, $\beta$-relevance measures sometimes do weigh between $i^*$ and $p$. Yet, $\beta$-relevance measures are not required to weigh between informativeness and truth. In concluding, let us briefly look at the most popular relevance measures all of which are 0-relevance measures. It is assumed throughout that Pr is strict.

As already mentioned, the Joyce–Christensen measure $s$, the distance measure $d$, and the Carnap measure $c$ get it right in all four cases (in case of Carnap's $c$, note that the union of all sets $Mod\left(\pm E_n^{\omega}\right)$ with $\Pr\left(\pm E_n^{\omega}\right) = 0$ has probability 0 in the sense of $\Pr^*$, whence $f = \Pr\left(\neg E_n^{\omega} \mid B\right) \cdot \Pr(B) \cdot \Pr\left(E_n^{\omega} \wedge B\right)$ is 0 only for a set of measure 0).

The log-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 $\omega$, and $H_1 \vdash H_2 \nvdash H_1$. In this case

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

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

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

However, $r$ does not get it right when both $H_1$ and $H_2$ are contingently false in $\omega$, and such that $H_1 \vdash H_2 \nvdash H_1$. In this case,

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

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

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

So even if both $\Pr\left(H_1 \mid E_m^{\omega}\right)$ and $\Pr\left(H_2 \mid E_m^{\omega}\right)$ converge to 0, the logically weaker $H_2$ may always have a greater $r$-value than $H_1$, as is the case when $\Pr\left(H_1 \mid E_m^{\omega}\right) = 1/2^m$ and $\Pr\left(H_2 \mid E_m^{\omega}\right) = 1/m$. The failure of $r$ is even clearer when both $H_1$ and $H_2$ are eventually falsified. In this case the only thing that matters is the minimal plausibility value, and they both get the 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 $l$,

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

(Fitelson, 1999, 2001). When $H_1$ and $H_2$ are contingently true or contingently false in $\omega$ and such that $H_1 \vdash H_2 \nvdash H_1$, it need not be the case that there is $n$ such that for all $m \geq n$:

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

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

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

So even if both $\Pr\left(H_1 \mid E_m^{\omega}\right)$ and $\Pr\left(H_2 \mid E_m^{\omega}\right)$ converge to 1 or to 0, the logically weaker $H_2$ may always have a greater $l$-value than the logically stronger $H_1$. For instance, this is the case when $\Pr\left(H_1 \mid E_m^{\omega}\right) = 1 - 1/m$ and $\Pr\left(H_2 \mid E_m^{\omega}\right) = 1 - 1/2^m$, or when $\Pr\left(H_1 \mid E_m^{\omega}\right) = 1/2^m$ and $\Pr\left(H_2 \mid E_m^{\omega}\right) = 1/m$. The failure of $l$ is even clearer when both $H_1$ and $H_2$ are eventually verified or falsified. In this 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, $l$ gets it right, if it is stipulated 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 assessment function (or measure of confirmation) furthers the goal of eventually arriving at informative true theories. Due to their focus on truth, relevance measures—like $s, t$-functions—separate true from false theories. However, 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 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, while I was a member of the *Philosophy, Probability, and Modeling* research group at the Center for Junior Research Fellows at the University of Konstanz.

## References

Bar-Hillel, Y. (1952). Semantic information and its measures. In *Transactions of the tenth conference on cybernetics* (pp. 33–48). New York: Josiah Macy, Jr. Foundation. (Reprinted in Bar-Hillel (1964), 298-310.)

Bar-Hillel, Y. (1955). An examination of information theory. *Philosophy of Science, 22*, 86–105. (Reprinted in Bar-Hillel (1964), 275–297.)

Bar-Hillel, Y. (1964). *Language and information. Selected essays on their theory and application*. Reading, MA: Addison-Wesley.

Bar-Hillel, Y., & Carnap, R. (1953). Semantic information. *British Journal for the Philosophy of Science, 4*, 147–157.

Carnap, R. (1952). *The continuum of inductive methods*. Chicago: University of Chicago Press.

Carnap, R. (1962). *Logical foundations of probability* (2nd ed). Chicago: University of Chicago Press.

Carnap, R., & Bar-Hillel, Y. (1952). *An outline of a theory of semantic information*. Technical Report No. 247 of the Research Laboratory of Electronics, MIT. (Reprinted in Bar-Hillel (1964), 221–274.)

Christensen, D. (1999). Measuring confirmation. *Journal of Philosophy, 96*, 437–461.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science, 66*, S362–S378.

Fitelson, B. (2001). *Studies in Bayesian confirmation theory*. Madison, WI: University of Wisconsin-Madison.

Flach, P. A. (2000). Logical characterisations of inductive learning. In D. M. Gabbay & R. Kruse (Eds.), *Abductive reasoning and learning* (pp. 155–196). Dordrecht: Kluwer Academic Publishers.

Gaifman, H., & Snir, M. (1982). Probabilities over rich languages, testing, and randomness. *Journal of Symbolic Logic, 47*, 495–548.

Hempel, C. G. (1943). A purely syntactical definition of confirmation. *Journal of Symbolic Logic, 8*, 122–143.

Hempel, C. G. (1945). Studies in the logic of confirmation. *Mind, 54*, 1–26, 97–121. (Reprinted in Hempel (1965), 3–51.)

Hempel, C. G. (1960). Inductive inconsistencies. *Synthese, 12*, 439–469. (Reprinted in Hempel (1965), 53–79.)

Hempel, C. G. (1962). Deductive-nomological vs. statistical explanation. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (vol. 3., pp. 98–169). Minneapolis: University of Minnesota Press.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: The Free Press.
Hempel, C. G., & Oppenheim, P. (1945). A definition of "degree of confirmation". *Philosophy of Science, 12*, 98–115.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 135–175. (Reprinted in Hempel (1965), 245–295.)
Hendricks, V. F. (2006). *Mainstream and formal epistemology*. Cambridge: Cambridge University Press.
Hintikka, J., & Pietarinen, J. (1966), Semantic information and inductive logic. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic*. Amsterdam: North-Holland.
Huber, F. (2004). *Assessing theories. The problem of a quantitative theory of confirmation*. PhD Dissertation. Erfurt: University of Erfurt.
Huber, F. (2005). What is the point of confirmation? *Philosophy of Science 72*, 1146–1159.
Huber, F. (2007a). The logic of theory assessment. *Journal of Philosophical Logic*.
Huber, F. (2007b). The plausibility-informativeness theory. In V. F. Hendricks & D. Pritchard (Eds.), *New waves in epistemology*. Aldershot: Ashgate.
Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press.
Kelly, K. T. (1996). *The logic of reliable inquiry*. Oxford: Oxford University Press.
Kelly, K. T. (1999). Iterated belief revision, reliability, and inductive amnesia. *Erkenntnis, 50*, 11–58.
Levi, I. (1961). Decision theory and confirmation. *Journal of Philosophy, 58*, 614–625.
Levi, I. (1963). Corroboration and rules of acceptance. *British Journal for the Philosophy of Science, 13*, 307–313.
Levi, I. (1967). *Gambling with truth. An essay on induction and the aims of science*. London: Routledge.
Levi, I. (1986). Probabilistic pettifoggery. *Erkenntnis, 25*, 133–140.
Lipton, P. (2004). *Inference to the best explanation* (2nd ed). London: Routledge.
Milne, P. (2000). Is there a logic of confirmation transfer? *Erkenntnis, 53*, 309–335.
Spohn, W. (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* (pp. 105–134). Dordrecht: Kluwer.
Spohn, W. (1990). A general non-probabilistic theory of inductive reasoning. In R. D. Shachter et al. (Eds.), *Uncertainty in artificial intelligence 4* (pp. 149–158). Amsterdam: North-Holland.
van Fraassen, B. C. (1983). Theory comparison and relevant Evidence. In J. Earman (Ed.), *Testing scientific theories* (pp. 27–42). Minneapolis: University of Minnesota Press.
Zwirn, D., & Zwirn, H. P. (1996). Metaconfirmation. *Theory and Decision, 41*, 195–228.