The Logic of Confirmation

Franz Huber
Centre for Junior Research Fellows
University of Konstanz

Penultimate version: please cite the paper in Philosophische Perspektiven

Abstract

The paper presents a new analysis of Hempel’s conditions of adequacy (Hempel 1945), differing from the one in Carnap (1962). Hempel, so it is argued, felt the need for two concepts of confirmation: one aiming at true theories, and another aiming at informative theories. However, so the analysis continues, he also realized that these two concepts were conflicting, and so he gave up the concept of confirmation aiming at informative theories. It is then shown that one can have the cake and eat it: There is a logic of confirmation that accounts for both of these two conflicting aspects.

1. Introduction

In his “Studies in the Logic of Confirmation” (1945) Carl G. Hempel presented the following conditions of adequacy for any relation of confirmation $\models \subseteq L \times L$ on a language (set of sentences closed under negation and conjunction) $L$: For all $E, H$ in $L$,

(1) Entailment: $E \models \neg H \Rightarrow E \models H$

(2) Consequence: $\{H : E \models \neg H\} \models H' \Rightarrow E \models H'$

(2.1) Special Consequence: $E \models \neg H, H \models H' \Rightarrow E \models H'$

(2.2) Equivalence: $E \models \neg H, H \models H' \Rightarrow E \models H'$

(3) Consistency: $\{H : E \models \neg H\} \models \neg \neg E$

(3.1) $E \models \bot, E \models \neg H \Rightarrow E \models \neg \neg H$

(3.2) $E \models \bot, E \models H, H \models \neg H' \Rightarrow E \models \neg H'$

(4) Converse Consequence: $E \models \neg H, H' \models \neg H \Rightarrow E \models \neg H'$

Hempel then showed that (1-4) entail that every sentence (observation report) $E$ confirms every sentence (theory) $H$, i.e. for all $E, H$ in $L$: $E \models H$. 


Since Hempel’s negative result, there has hardly been any progress in constructing a logic of confirmation. The only two articles I know of are Zwirn/Zwirn (1996) and Milne (2000). Roughly, Zwirn/Zwirn (1996) argue that there is no logic of confirmation taking into account all of the partly conflicting aspects of confirmation, whereas Milne (2000) argues that there is a logic of confirmation, but that it does not deserve to be called a logic. One reason for this seems to be that up to now the predominant view on Hempel’s conditions is the analysis Carnap gave in his *Logical Foundations of Probability* (1962), § 87.

Carnap’s analysis can be summarized as follows: In presenting his first three conditions Hempel was mixing up two distinct concepts of confirmation, two distinct explicanda in Carnap’s terminology, viz.

(i) the concept of incremental confirmation (positive statistical relevance, initially confirming evidence in Carnap’s terminology) according to which $E$ confirms $H$ iff $E$ increases the probability of $H$, $\Pr(H \mid E) > \Pr(H)$, and

(ii) the concept of absolute confirmation according to which $E$ confirms $H$ iff the probability of $H$ given $E$ is greater than some value $r$, $\Pr(H \mid E) > r$.

Hempel’s second and third conditions hold true for the second explicandum, but they do not hold true for the first explicandum. On the other hand, Hempel’s first condition holds true for the first explicandum, but it does so only in a qualified form (Carnap 1962, 473) – namely only if $H$ is not already assigned probability 1.

This, however, means that Hempel first had in mind the explicandum of incremental confirmation for the entailment condition; then he had in mind the explicandum of absolute confirmation for the consequence and the consistency conditions; and then, when Hempel presented the converse consequence condition, he got completely confused, so to speak, and had in mind still another explicandum or concept of confirmation. Apart from not being very charitable, Carnap’s reading of Hempel also leaves open the question what the third explicandum might have been.

2. Conflicting Concepts of Confirmation
We present another analysis based on the following two notions. A relation $\models \subseteq L \times L$ is a *likeliness* relation on the language $L$ iff

$$E \models H, H \models H' \Rightarrow E \models H';$$

and $\models$ is a *loveliness* relation on $L$ iff

$$E \models H, H \models H' \Rightarrow E \models H'.$$

These two concepts underlie the two main approaches to confirmation that have been put forth in the last century: qualitative *Hypothetico-Deductivism* HD, and quantitative probabilistic *Inductive Logic* IL.

According to HD, $E$ HD-confirms $H$ if(f) $H$ logically implies $E$. Hence, if $E$ HD-confirms $H$ and $H'$ logically implies $H$, then $E$ HD-confirms $H'$. So HD-confirmation is a loveliness relation.

According to IL, the degree of *absolute* confirmation of $H$ by $E$ equals the probability of $H$ given $E$, $\Pr(H \mid E)$. The natural *qualitative* counterpart of this quantitative notion is that $E$ “absolutely confirms” $H$ iff $\Pr(H \mid E) > r$, for some value $r$ in $[.5, 1]$ (Carnap’s second *explicandum*). If $E$ absolutely confirms $H$ and $H$ logically implies $H'$, then $E$ absolutely confirms $H'$. So absolute confirmation is a likeliness relation.

This is not the way Carnap defined qualitative IL-confirmation in his (1962), ch. VII. There he required that $E$ raises the probability of $H$ in order for $E$ to qualitatively IL-confirm $H$, i.e. $\Pr(H \mid E) > \Pr(H)$. Nevertheless, the above seems to be the natural qualitative counterpart of the degree of absolute confirmation. The reason is that later on, the difference between $\Pr(H \mid E)$ and $\Pr(H)$ was taken as degree of *incremental confirmation*, and Carnap’s proposal is the natural qualitative counterpart of this notion of incremental confirmation. Let us say that $E$ *incrementally confirms* $H$ iff $\Pr(H \mid E) > \Pr(H)$.

The loveliness concept underlying HD aims at *informative* theories, whereas the likeliness concept underlying IL aims at *true* (probable) theories. These two concepts are conflicting in the sense that the first increases, whereas the second decreases with the logical strength of the theory to be assessed.

3. Hempel Vindicated
Turning back to Hempel’s conditions, note first that Carnap’s second *explicandum* satisfies the entailment condition *without* qualification: If $E$ logically implies $H$, then $\Pr(H \mid E) = 1 \geq r$, for any value $r$ in $[.5, 1]$.

So the following more charitable reading of Hempel seems plausible: When presenting his first three conditions, Hempel had in mind Carnap’s second *explicandum*, the concept of absolute confirmation. But then, when discussing the converse consequence condition, Hempel also felt the need for a second concept of confirmation: the loveliness 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 the second concept of confirmation, Hempel also realized that these two concepts were *conflicting*, and so he abandoned the loveliness concept in favour of the likeliness concept.

4. The Logic of Theory Assessment

However, in a sense one can have the cake and eat it: There is a logic of confirmation that takes into account both of these two conflicting concepts.

Roughly speaking, HD says that a good theory is informative, whereas IL says that a good theory is true (probable). The driving force behind Hempel’s conditions seems to be the insight that *a good theory is both true and informative*. Hence, in assessing a given theory by the available data, one should account for these two conflicting aspects. This is done in the following.

Let $\langle W, A, \kappa \rangle$ be a ranking space, where $W$ is a non-empty set of possibilities, $A$ is a field over $W$, i.e. a set of subsets of $W$ containing the empty set $\emptyset$ and closed under complementation and finite intersections, and $\kappa: W \rightarrow N \cup \{\infty\}$ is a *ranking function* (Spohn 1988, 1990), i.e. a function from $W$ into the set of extended natural numbers $N \cup \{\infty\}$ such that at least one possibility $w$ in $W$ is assigned rank 0. $\kappa$ is extended to a function on $A$ by setting

$\kappa(\emptyset) = \infty$ and defining, for each non-empty $A$ in $A$,

$\kappa(A) = \min \{\kappa(w): w \in A\}$.

The conditional rank of $B$ given $A$, $\kappa(B \mid A)$, is defined as
\[ \kappa(B \mid A) = \kappa(A \cap B) - \kappa(A) \text{ if } \kappa(A) < \infty, \]

\[ = 0 \quad \text{ if } \kappa(A) = \infty. \]

A ranking function represents an ordering of disbelief. For \( A, B \) in \( A \), \( \kappa(B^C \mid A) \) measures how likely \( B \) is given \( A \), whereas \( \kappa(B \mid A^C) \) measures how much \( B \) informs about \( A \), where \( A^C \) is the set-theoretical complement of \( A \) with respect to \( W \).

A ranking space \(<W, A, \kappa>\) is an assessment model for the language \( L \) iff \( W \) is the set \( \text{Mod}_L \) of all models for \( L, \text{Mod}(\alpha) \in A \) for each \( \alpha \) in \( L \), and \( \kappa(\text{Mod}(\alpha)) < \infty \) for each consistent \( \alpha \) in \( L \).

The consequence relation \( \models_{\kappa} \) on the language \( L \) induced by an assessment model \(<\text{Mod}_L, A, \kappa>\) is defined as follows: For all \( H, E \) in \( L \),

\[ E \models_{\kappa} H \iff \kappa(\text{Mod}(H) \mid \text{Mod}(E)) \leq \kappa(\text{Mod}(\neg H) \mid \text{Mod}(E)) \& \kappa(\text{Mod}(\neg H) \mid \text{Mod}(\neg E)) \leq \kappa(\text{Mod}(H) \mid \text{Mod}(\neg E)) \]

where at least one of these inequalities is strict. In words: \( H \) is an acceptable theory given \( E \) (according to \( \kappa \)) iff \( H \) is at least as likely as and more informative than \( \neg H \) given \( E \), or \( H \) is more likely than and at least as informative as \( \neg H \) given \( E \).

On the other (the syntactical) hand, a relation \( \models \subseteq L \times L \) is an assessment relation on \( L \) iff \( \models \) satisfies the following principles, for all \( E, H \) in \( L \):

(A1) \( E \models E \) \quad Reflexivity*

(A2) \( E \models H, E \not\models E', H \not\models H' \Rightarrow E' \models H' \) \quad L-Equivalence*

(A3) \( E \models H \Rightarrow E \models H \) \quad Selectivity*

(A4) \( E \models H \Rightarrow E \models E \land H \) \quad Weak Composition*

(A5) \( E \models H \Rightarrow \neg E \models \neg H \) \quad Loveliness and Likeliness

(A6) \( \models E \lor H \Rightarrow E \lor H \models E \) or \( E \lor H \models H \) \quad Either-Or

(A7) \( E \lor H \models E, \models E \lor H \Rightarrow E \lor \neg E \models \neg E \) \quad Negation 1

(A8) \( E \lor F \models E, F \lor H \models F, \models E \lor H \Rightarrow E \lor H \models E \) \quad quasi Nr 21

(A9) \( E \lor F \models E, F \lor H \models F, \models E \lor H \Rightarrow E \lor H \models \neg E \) \quad suppl. Nr 21

(A10) \( E \lor E_{i+1} \models E_{i+1}, \models E \lor E_j \Rightarrow \exists n \forall m \geq n: E_m \lor E_{m+1} \models E_m \) \quad Minimum

(A11) \( E \models E \land H, E \models E \land H \Rightarrow E \models \neg H \)

(A12) \( E \models E \land \neg H, E \models E \land H, \models E \models \neg E \Rightarrow E \models \neg H \)

(A13) \( E \lor \neg E \models E, E \models H \Rightarrow E \lor \neg E \models H \)

(A14) \( E \land \neg E \models E, E \lor H \models E \Rightarrow E \land \neg E \models H \)
The * starred principles are among the core principles in Zwirn/Zwirn (1996). quasi Nr 21 without the restriction \( \vdash E \lor H \) is the derived rule (21) of the system \( P \) in Kraus/Lehmann/Magidor (1990).

Theorem: The consequence relation \( \models_\kappa \) induced by an assessment model \( <Mod_L, A, \kappa> \) for \( L \) is an assessment relation on \( L \). For each assessment relation \( \models \) on \( L \) there is an assessment model \( <Mod_L, A, \kappa> \) for \( L \) such that \( \models = \models_\kappa \).

The following principles are admissible:

(A15) \( E \models F \Rightarrow H, E \models F \Rightarrow E \models H \)  
MPC

(A16) \( E \models F, E \models H \Rightarrow E \models F \lor H \) or \( E \models F \lor H \)  
quasi Composition

(A17) \( E \models E \land \neg E \Rightarrow E \models E \land \neg E \)  
Consistency*

(A18) \( E \models E \lor \neg E \Rightarrow E \models E \lor \neg E \)  
Informativeness

(A19) \( E \land F \models H, E \land \neg F \models H \Rightarrow E \models H \)  
Proof by Cases

The following principles are not admissible:

(N1) \( E \models H \Rightarrow E \models H \)  
Entailment (Supraclassicality)*

(N2) \( H \models E \Rightarrow E \models H \)  
Conversion

(N3) \( F \models H, E \models F \Rightarrow E \models H \)  
Left Monotonicity

(N4) \( E \models F, E \models \neg H \Rightarrow E \models H \)  
Strong Selectivity

(N5) \( E \land F \models H, E \models F \Rightarrow E \models H \)  
Cut

(N6) \( E \models H, E \models \neg F \Rightarrow E \land F \models H \)  
Cautious Monotonicity

In comparing the present approach with standard nonmonotonic logic in the KLM-tradition (Kraus/Lehmann/Magidor 1990), we note two points:

First, the present system is genuinely nonmonotonic in the sense that not only Left, but also Right Monotonicity is not admissible:

(N7) \( E \models F, F \models H \Rightarrow E \models H \)  
Right Monotonicity (Right Weakening)

Not only arbitrary strengthening of the premises, but also arbitrary weakening of the conclusion is not allowed. The reason is this: By arbitrary weakening of the conclusion
information is lost – and the less informative conclusion need not be worth taking the risk being led to a false conclusion.

Second, the present approach can explain why everyday reasoning is satisfied with a standard that is weaker than truth-preservation in all possible worlds (e.g. truth-preservation in all normal worlds): We are willing to take the risk of being led to a false conclusion, because we want to arrive at informative conclusions.

Finally, one might wonder how the present logic of theory assessment compares to Carnap’s *dictum* that qualitative confirmation is positive statistical relevance. A first answer is given by

**Observation:** For every regular probability $Pr$ on a language $L$,

$$\neg Pr = \perp^+ Pr \cup \{<E, H>: E \models H \models E \lor \neg E \} \cup \{<E, H>: E \models \neg H \models E \land \neg E \}$$

is an assessment relation on $L$, where $\perp^+ Pr$ is the relation of positive statistical relevance in the sense of $Pr$.

However, theory assessment is not the same as positive statistical relevance, for Symmetry is not admissible:

$$(N8) \ E \models H \Rightarrow H \models \neg E \quad \text{Symmetry}$$

**Acknowledgments**

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 to Luc Bovens.

**References**


