Falsificationist Confirmation

Jan Sprenger†

Abstract

Existing accounts of hypothetico-deductive confirmation are able to circumvent the classical objections (e.g. the tacking problems), but the confirmation of conjunctions of hypotheses brings them into trouble. Therefore this paper develops a new, falsificationist account of qualitative confirmation by means of Ken Gemes' theory of content parts. The new approach combines the hypothetico-deductive view with falsificationist and instance confirmation principles. It is considerably simpler than the previous suggestions and gives a better treatment of conjunctive hypotheses while solving the tacking problems equally well.

1 Introduction

In the last two decades, qualitative accounts of confirmation have largely been superseded by probabilistic accounts, in particular Bayesian ones. While probabilities certainly provide a powerful framework for inductive reasoning, this does not imply that qualitative reasoning has become superfluous. In a lot of empirical sciences probabilistic reasoning still plays a minor role, if at all. Qualitative arguments are thus central for the confirmation of scientific hypotheses, and their relativity to a set of primitive predicates of a language does not diminish their significance. We do not aim at a solution of Goodman's new riddle of induction but rather at reconstructing actual cases of and developing normative constraints for theory confirmation. Indeed, the most prominent cases of theory confirmation and replacement are situated in a qualitative framework, e.g. the confirmation of Kepler's laws or Darwin's evolution theory. In order to have a sensible model for such cases, an account of qualitative confirmation is indispensable – introducing subjective probabilities would simply misrepresent the problem. Apart from that, study of qualitative confirmation reveals the role of deductive relations in scientific reasoning and illuminates typical features of confirmation which are helpful for probabilistic accounts, too. All these facts encourage philosophers of science not to give up qualitative confirmation and to keep the Hempel-Glymour tradition in this field alive.

†For further information, contact the author at the Universität Bonn, Institut für Philosophie, Am Hof 1, 53113 Bonn, Germany. Email: jan.sprenger@gmx.net

Most philosophers of science have tried to capture qualitative confirmation in terms of hypothetico-deductive (H-D) confirmation: hypotheses are confirmed by checking their predictions, i.e. by deducing evidential consequences from hypothesis and background knowledge. If these predictions are found to obtain, they will confirm the hypothesis. In fact, scientists often perform several deductive steps to descend from a theoretical hypothesis to an observable consequence. This resemblance to actual practice distinguishes H-D confirmation among qualitative approaches to confirmation. On the other hand, the classical formulation of hypothetico-deductivism

> (H-D): *E H-D confirms H relative to background knowledge K* if and only if (a) $H.K$ is consistent, (b) $H.K$ entails E and (c) K alone does not entail E.

has major drawbacks (cf. Glymour 1980b). First, it is possible to tack *irrelevant conjunctions* to the hypothesis H and to preserve the confirmation relation: If H is confirmed by a piece of evidence E, $H.X$ is confirmed by the same E for an arbitrary X. Second, it is equally possible to tack *irrelevant disjunctions* to the evidence E and to preserve the confirmation relation: If E confirms a hypothesis H, $E \vee E'$ confirms the same H for an arbitrary E'.[1] Both objections exploit the fact that classical H-D confirmation gives no account of *evidential relevance*. This failure of classical H-D confirmation might lead to the conclusion that the entire approach is hopeless and should be replaced by an instance view of confirmation (cf. Glymour 1980a).

Nevertheless, some philosophers have undertaken remarkable efforts to rescue the H-D account of confirmation, most notably Ken Gemes and Gerhard Schurz. They have developed accounts of content part entailment respectively relevant deduction that can be applied to accounts of H-D confirmation (cf. Gemes 1993, 1998 and Schurz 1991). The suugested accounts circumvent the aforementioned tacking paradoxes and a number of other alleged counterexamples to H-D confirmation, thus eliminating a lot of cases of spurious confirmation.[2] Nonetheless, both proposals are not free of problems. In section 2, I introduce and discuss a novel objection to H-D confirmation. Section 3 briefly lays the technical foundations for overcoming the problem while section 4 introduces the new proposal – falsificationist confirmation – and elaborates its merits. The final section 5 summarizes the results.

[1] This inference holds unless K implies $E \vee E'$.

[2] For instance, it is *not* the case that the evidence Fa confirms the hypothesis $\forall x\, Fx.\forall x\, Gx$, etc.

2 The new problem of H-D confirmation

Assume the following situation. We would like to test a new antibiotic A which is supposed to kill all bacteria of strain S in the human body within two days. We set up a clinical trial with a group of infected persons who are given the drug. For the sake of simplicity, we do not set up a control group – we might believe the drug to be so effective that it would be irresponsible to give a placebo to the other patients. Then we administer the drug to all patients $a_1, \ldots, a_n$ in the treatment group and wait for the results. Indeed, for n-1 patients everything works fine and all infection markers give negative results after two days. Patient a_n, however, got the antibiotic at the latest time point of all patients, and after being given the drug, she behaves in a way that points to severe mental disorder. Since a causal connection to the taking of A cannot be ruled out, we stop the treatment. In other words, with regard to a_n, we cannot decide whether the drug is indeed as effective as our hypothesis posits. Still, we have to evaluate the experiment. Do the total observations confirm that antibiotic A kills all bacteria of a strain S within two days *and* leads to mental disorder?[3]

Clearly, such a claim would stand on very shaky grounds. First, to confirm the effectiveness of A, we should wait for the results of the last patient. Second and worse, the observations do certainly not confirm that taking A always leads to mental disorder. To confirm that hypothesis properly, we would have to make long-term observations of the other trial persons. In any case, no responsible medical research report would conclude "A clinical trial with n test persons has confirmed that taking antibiotic A kills all bacteria of strain S within two days, but leads to mental disorder, too". It is just outrageous to neglect that both claims are not based on the full treatment group, especially since no single patient has exhibited mental disorder *and* lack of S-bacteria. Therefore we should (and would) not speak of confirmation in this case. However, neither the classical nor the refined versions of H-D confirmation agree. Using the canonical formalization

$$\begin{array}{ll} H_1 = \forall x\,(Ax \rightarrow \neg Sx) & \qquad K = Aa_1.Aa_2 \ldots Aa_n \\ H_2 = \forall x\,(Ax \rightarrow Mx) & \qquad E = \neg Sa_1.\neg Sa_2 \ldots \neg Sa_{n-1}.Ma_n \end{array}$$

the combined hypothesis $H_1.H_2$ is H-D-confirmed by evidence E relative to background knowledge K. This result is highly undesirable and does not depend on whether we choose Gemes', Schurz' or the classical formulation of H-D confirmation. The observations $\neg Sa_1, \neg Sa_2, \ldots, \neg Sa_{n-1}$ on the one hand and Ma_n on the other hand are completely unrelated so that they should not jointly confirm a composite hypothesis which the single parts clearly fail to confirm. We strongly feel that the evidence should contain

[3] Ken Gemes brought this problem (in a slightly different setting) to my attention. Interestingly, we have opposing views on how to resolve it.

at least one *instance* of $H_1.H_2$, i.e. a patient who, after being given the drug, exhibits both absence of S-bacteria and mental disorder. Although E is deductively entailed by $H_1.H_2$, such a kind of "confirming evidence" is far too easy to obtain, thus opening the door to deliberate manipulation of scientific experiments. More generally, we could examine a group of objects, finding property P_1 in some of them and property P_2 in others. According to all existing variations of H-D confirmation, these observations would confirm that all objects share property P_1 *and* property P_2. This is a fallacy similar to the tacking paradox since the examined objects do not count as evidence for the *conjunctive hypothesis*. Whatever the classical and refined accounts of H-D confirmation capture, it differs from standard confirmation in science. Thus, the proposed accounts of H-D confirmation are in severe trouble, the more so as for a deductive explication of scientific confirmation, being too permissive is much worse than being too restrictive.

It might be objected that the failure of H-D confirmation should not disturb us too much – there is no completely perfect account of confirmation, and we should simply adjust our intuitions and learn to live with the counterexamples. Such a reply would be fair if we could not set up a better account of qualitative confirmation. But I believe that such an account is available, and I would like to sketch it in the subsequent sections.

3 Content parts

This section introduces Ken Gemes' account of content parts which is required for the definition of a new account of qualitative confirmation.
Well-formed forms sometimes have irrelevant consequences, e.g. the conclusion in $Fa \models (Fa \vee Ga)$ contains the irrelevant element Ga. For this purpose, I would like to sketch Gemes' account of *content parts*. It is based on the idea that the consequens is a part of the *content* of the antecedens only if every *relevant model* of the consequens can be extended to a relevant model of the antecedens. Only such content-preserving entailments count as relevant. For the sake of simplicity, I presuppose a first order predicate language L without identity.[4] The following definition captures our intuitive view of relevance relations between wffs:

Definition 3.1 *An atomic well-formed form (wff) β is* relevant to a wff α *if and only if there is some model M of α such that: if M' differs from M only in the value β is assigned, M' is not a model of α.*

[4]The account given in this paper uses the abridged definition of content parts in Gemes 2006. See Gemes 1997, 451-60, for a thorough model-theoretic introduction of content parts that also covers languages with identity.

So intuitively, β is relevant for α if at least in one model of α the truth value of β cannot be changed without making α false. Now we can define the notion of a relevant model:

Definition 3.2 *A* relevant model *of a wff α is a model of α that assigns values to all and only those atomic wffs that are relevant to α.*

The idea is that the underlying structure assigns meaning only to those non-logical constants that occur in the relevant atomic wffs of α. For example, every relevant model of $Fa.Ga$ is a relevant model of $Fa \to Ga$, but not all relevant models of $Fa \vee Ga$ are relevant models of $Fa.Ga$.[5] This allows us to define the notion of a *content part*:

Definition 3.3 *For two wffs α and β, β is a* content part *of α ($\alpha \models_{cp} \beta$) if and only if (1) α and β are contingent, (2) α logically entails β and (3) every relevant model of β has an extension which is a relevant model of α.*

The content part relation is a means of detecting *irrelevant conclusions*. For instance, $Fa \vee Ga$ is no content part of Fa because the model that assigns 'false' to Fa and 'true' to Ga is a relevant model of $Fa \vee Ga$ but no model of Fa. The content part relation marks such deductions as irrelevant.[6] We will take advantage of this property of content parts in the remainder. Finally, we define the notion of the domain of a wff:

Definition 3.4 *The* domain of a wff *α, denoted by $dom(\alpha)$, is the set of singular terms which occur in the atomic (!) well-formed formulas (wffs) of L that are relevant for α.*

Informally, $dom(\alpha)$ specifies the scope of a wff α. For instance, the domain of $Fa.Fb$ is $\{a, b\}$ whereas the domain of $Fa.Ga$ is $\{a\}$. Quantifiers are treated substitutionally, e.g. the domain of $\forall x\, Fx$ is $\{a, b, c, \ldots\}$.

4 Falsificationist confirmation

The new account of qualitative confirmation stands on three pillars: First, a hypothesis is confirmed by verifying its predictions. This is also the basic tenet of hypothetico-deductivism. Second, the evidence has to put the hypothesis to a serious test, in other words, had a result different from the

[5] Assign 'true' to Fa and 'false' to Ga.

[6] Similarly, Fa is a content part of $\forall x\, Fx$, but $Fa \vee Fb$ is no content part of $\forall x\, Fx$ since the model that assigns 'true' to Fa and 'false' to Fb cannot be extended to a model of $\forall x\, Fx$.

actual one obtained, the hypothesis would have been locally falisified. This condition incorporates the falsificationist principle of conjecture and refutation. Third and last, instances of a hypothesis have a distinguished position in the confirmation of hypotheses.

Now we can proceed to the formalization of those principles. First, evidential predictions are deductively obtained by conjoining hypothesis H with background assumptions K. Hence, $H.K \models E$ is a necessary condition for the new account. As we would like to circumvent the problem of tacking by disjunction, we restrict ourselves to relevant entailments and proceed to the stronger condition $H.K \models_{cp} E$. Second, falsificationists deduce bold and risky predictions from theory and background knowledge – predictions that would fall back on the hypothesis under test if they failed to obtain. In other words, if E confirms H, $\neg E$ will falsify H. Therefore we demand that

$$\neg E.K \models_{cp} \neg H_{|dom(E)}.K \tag{1}$$

where $H_{|dom(E)}$ denotes the restriction of H to the domain of E.[7] Note in particular that deductively gained instances confirm a hypothesis. And vice versa: if we do not get a full instance of H, we normally fail to confirm H.[8] So we automatically cover the third principle which demands us to give special weight to instance confirmation. Finally, we write down the definition of falsificationist or F-confirmation:

Definition 4.1 ***(Falsificationist Confirmation (FC)):*** E F-confirms H relative to K *if and only if*

- *E is a content part of $H.K$ ($H.K \models_{cp} E$) and*
- *$\neg H_{|dom(E)}.K$ is a content part of $\neg E.K$ ($\neg E.K \models_{cp} \neg H_{|dom(E)}.K$)*

We will see that this new account is able to deal both with the classical and the novel objections to H-D confirmation.[9] A main challenge for deductive theories of confirmation consists in the tacking paradoxes that clash with our view that confirmation must not be arbitrarily transmitted (cf. section 1).

The second condition of (FC) ensures that, in the case of irrelevant conjunctions, there are relevant models of $\neg H_{|dom(E)}.K$ that cannot be extended to

[7] The idea to restrict a hypothesis to the domain of the evidence was introduced by Carl G. Hempel (1965). Omitting the restriction would not work, because, for a sufficiently general H, (1) would never be satisfied.

[8] This is especially pronounced when the evidence is a truth-functional compound of atomic wffs.

[9] Furthermore, the definition makes clear that (FC) satisfies the Equivalence Condition which is a basic requirement for all formal accounts of confirmation: if H is logically equivalent to H', then E confirms H relative to K if and only if E confirms H'.

relevant models of $\neg E.K$. For instance, if $H = \forall x\, Fx$, $X = \forall x\, Gx$, $K = \emptyset$ and $E = Fa$, E should not confirm $H.X$ because it is irrelevant to X. Indeed, $\neg(H.X)_{|\{a\}}.K = \neg Fa \vee \neg Ga$ is no content part of $\neg E.K = \neg Fa$, thus avoiding the undesirable result. Partial confirmation, though, is still possible: If we add the background knowledge $K = Ga$ (instead of tautologous K), $E = Fa$ F-confirms $H.X$ relative to K. This corresponds to our intuitions that partial confirmation is sound as long as the background knowledge provides the missing piece of evidence that we need for a full instance of the hypothesis. Put another way, partial confirmation counts as F-confirmation whenever the background knowledge covers that part of the predictions of the hypothesis which the evidence does not confirm itself.

A corresponding problem for H-D confirmation arises if the evidence is logically weakened, i.e. if irrelevant disjunctions are tacked to the evidence. Assume that a hypothesis H is confirmed by a certain piece of evidence E. If we tack an arbitrary disjunct E' to the evidence, classical H-D confirmation of H remains intact because $\models$ is a transitive relation and $H \models E \models (E \vee E')$. But we do not think that such an $E \vee E'$ is still a relevant prediction of H because E' could be anything. Indeed, the first condition of (FC) requires confirming evidence to be a content part of $H.K$. If an irrelevant disjunction is tacked to the evidence, there will be relevant models of the compound evidence which cannot be extended to relevant models of $H.K$. For instance, if $H = \forall x\, Fx$, $K = \emptyset$, $E = Fa$ and $E' = Gb$, $E \vee E' = Fa \vee Gb$ is no content part of H.

Finally, the behavior of falsificationist confirmation with regard to conjunctive confirmation avoids the problems of conjunctive confirmation. Let $K = \emptyset$. Assume that E_1 confirms H_1 and E_2 confirms H_2 due to deductive entailment, in the very spirit of H-D confirmation. Then it is not necessarily the case that $E_1.E_2$ F-confirms $H_1.H_2$, too. The existing H-D accounts of confirmation instantiate that scheme, but we have seen the problems in section 2. In this respect, falsificationist confirmation differs from all previous accounts. Nonetheless this does *not* rule out the confirmation of conjunctions of independent hypotheses. Consider the following case:

$$\begin{aligned} H_1 &= \forall x\,(Rx \rightarrow Bx) & \qquad E_1 &= Ba \\ H_2 &= \forall x\,(Dx \rightarrow Wx) & \qquad E_2 &= Wb \\ K &= Ra.Db.(\forall x\,\neg(Rx.Dx)) \end{aligned}$$

For instance, H_1 could mean that all ravens are black and H_2 could mean that all doves are white. Then the background knowledge K asserts that a is a raven, b is a dove and nothing is both a raven and a dove. Now, the observation $E_1.E_2$ that a is black and b is white F-confirms the composite hypothesis $H_1.H_2$ that all ravens are black and all doves are white. So F-confirmation satisfactorily captures conjunctive confirmation. But if we had omitted the background knowledge that the sets of ravens and doves are disjoint, we would not have got confirmation (a could have also been a non-white dove). Falsificationist confirmation is thus more fine-grained and sensitive to the pecularities of a specific case[10] than its predecessors which unanimously affirm confirmation in the above case as well as in the example of section 2. By combining deductivist and instantial views of confirmation, (FC) avoids a lot of contentious properties of a purely H-D approach and complies with our intuitions about evidential relevance.

Apart from that, (FC) is considerably simpler than other deductive accounts of confirmation. Although the definition of (FC) has some parallels to Gemes' account of H-D confirmation (see, for instance, Gemes 1998), it is clearly more parsimonious: Gemes suggests a criterion which involves the *natural axiomatization* of a hypothesis. But it is open to serious discussion how to fix the notion of a natural axiomatization, so much the more as Schurz (2005) has pointed out that Gemes' natural axiomatizations are not very fine-grained and in some cases far from being the "natural" representations of a hypothesis. On the other hand, Schurz' own suggestion – hypotheses must be represented as conjunctions of their *relevant consequence elements* – has similar drawbacks. Moreover, Schurz does not assign any role to the background knowledge. Thus, (FC) has a simpler and less contentious definition than the rival proposals: Neither natural axiomatizations nor relevant consequence elements are introduced. This contributes to the overall attractivity of (FC).

5 Summary and conclusions

This paper has criticized preceding accounts of H-D confirmation because they yield spurious confirmation when conjunctive compounds of hypotheses are tested. To overcome this problem, I have introduced a new account of qualitative confirmation which combines Gemes' theory of content parts with falsificationist principles. The falsificationst account of confirmation (FC) is able to rebut the novel as well as the classical objections to H-D confirmation, in particular the tacking paradoxes. Besides, it gives a new and convincing reading of the confirmation of conjunctive hypotheses and considers the value of confirmation by instances. Nonetheless (FC) is very simple and does not require a natural decomposition of a hypothesis, unlike the preceding accounts. Therefore (FC) improves in several respects upon the accounts of H-D confirmation discussed in the literature.

[10] These pecularities are usually encoded in the background knowledge.

References

[1] KEN GEMES (1993): “Hypothetico-Deductivism, Content and the Natural Axiomatisation of Theories”, *Philosophy of Science* **60**, 477-87.

[2] KEN GEMES (1997): “A New Theory of Content II: Model Theory and Some Alternatives”, *Journal of Philosophical Logic* **26**, 449-76.

[3] KEN GEMES (1998): “Hypothetico-Deductivism: The Current State of Play”, *Erkenntnis* **49**, 1-20.

[4] KEN GEMES (2006): “Content and Watkins’ Account of Natural Axiomatizations”, *dialectica* **60**, 85-92.

[5] CLARK GLYMOUR (1980a): *Theory and Evidence*. Princeton University Press, Princeton.

[6] CLARK GLYMOUR (1980b): “Discussion: Hypothetico-Deductivism is Hopeless”, *Philosophy of Science* **47**, 322-25.

[7] CARL G. HEMPEL (1965): “Studies in the Logic of Confirmation”, in: *Aspects of Scientific Explanation*, 3-46. The Free Press, New York. (Reprint from *Mind* **54**, 1945.)

[8] GERHARD SCHURZ (1991): “Relevant Deduction”, *Erkenntnis* **35**, 391-437.

[9] GERHARD SCHURZ (2005): “Bayesian H-D Confirmation and Structuralistic Truthlikeness: Discussion and Comparison with the Relevant-Element and the Content-Part Approach”, in: R. Festa (ed.), *Logics of Scientific Discovery. Essays in Debate with Theo Kuipers*, Rodopi, Amsterdam 2005, 141-59.