Logique & Analyse **164** (1998), 271–299

QUALITATIVE CONFIRMATION BY THE HD-METHOD

THEO A.F. KUIPERS

Introduction

This paper forms, together with a companion paper, see below, in many respects a systematic exposition of well-known ideas on deductive and non-deductive confirmation. However, the papers present these ideas in a non-standard way and refine and revise several standard solutions of problems associated with these ideas. The present paper deals with qualitative (deductive) confirmation as resulting from applying the HD-method. The companion paper, entitled "Quantitative confirmation, and its qualitative consequences", will appear in one of the next issues of *Logique et Analyse*, and deals with quantitative (deductive and non-deductive) confirmation and its qualitative consequences. The main non-standard aspect is the approach of confirmation from the 'success perspective', according to which confirmation is equated with evidential success, more specifically, with an increase of the plausibility of the evidence on the basis of the hypothesis. Hence, in contrast to standard expositions, confirmation is not equated with an increase of the plausibility of the hypothesis by the evidence. This is merely an additional aspect of confirmation under appropriate conditions.

The present paper starts from the classical idea that testing of an hypothesis by the hypothetico-deductive (HD-)method aims at establishing the truth-value of the hypothesis, and results in confirmation or falsification. A qualitative (classificatory and comparative) theory of deductive confirmation is

This paper and the companion paper "Quantitative confirmation, and its qualitative consequences" (see the *Introduction*) are adapted, semi-final versions of chapters 2 and 3 of the book *From Instrumentalism to Constructive Realism* (Kuipers, 2000). I gratefully acknowledge that, in the final process of writing and rewriting, there were many stimulating and influential exchanges with Roberto Festa. Some discussions with and comments by Kenneth Gemes were also very clarifying. I also like to thank David Atkinson, Anne Boomsma, Job van Eck, Richard Jeffrey, Johannes Heidema, Carl Hempel, Ivo Molenaar, Hans Mooij, Ilkka Niiniluoto, David Papinau, Willem Schaafsma, Abner Shimony, the members of the Groningen Cognitive Structures research group, and the participants of the "Models and Reality" symposium in Brussel (May 1997) for their comments in earlier stages. I thank the NIAS (Wassenaar) for the stimulating environment during the year 1996/97 and Ann Simpson and Jane Colling for correcting my English.

presented, as well as its (in details) non-standard solutions of Hempel's raven paradoxes and Goodman's grue paradox. It is called a *pure* theory of confirmation because it is based on the principle that equally successful hypotheses are equally confirmed by the relevant evidence, that is, independent of the initial plausibility of the hypotheses.

In the companion paper the corresponding quantitative, i.c., probabilistic Bayesian, theory of confirmation is presented, with a decomposition in deductive and non-deductive confirmation. It is pure, and *inclusive*, hence non-standard, in the sense that it leaves room for the confirmation of hypotheses with probability zero. The resulting qualitative theory of (general) confirmation, including the one for deductive confirmation, is also formulated. The resulting quantitative treatment of the raven paradoxes is compared in detail with an analysis in terms of the severity of tests and the standard Bayesian solution as presented by Horwich (Appendix 2). In Appendix 1 it is argued that Popper's ideas about corroboration basically amount to an inclusive and impure variant of the Bayesian approach to confirmation.

According to the leading expositions of the hypothetico-deductive (HD-) method by Hempel (1966), Popper (1934/1959) and De Groot (1961/1969), the aim of the HD-method is to answer the question whether a hypothesis is true or false, that is, it is a method of testing. Note that this formulation of the aim of the HD-method is laden with the epistemological assumption of (theory) realism according to which it generally makes sense to aim at true hypotheses. Though the realist has a clear aim with HD-testing, this does not mean that HD-testing is only useful from that epistemological point of view. Let us briefly review in this respect the main other epistemological positions. Hypotheses may or may not use so-called 'theoretical terms', besides so-called 'observation terms'. What is observational is not taken in some absolute, theory-free sense, but depends greatly on the level of theoretical sophistication. Theoretical terms intended to refer to something in the real world may or may not in fact refer. For the *(constructive) empiricist* the aim of HD-testing is to find out whether the hypothesis is observationally true, i.e., has only true observational consequences, or is observationally or empirically adequate, to use Van Fraassen's favorite expression. For the *instrumentalist* the aim of HD-testing is still more liberal: is the hypothesis observationally true for all intended applications? The *referential realist*, on the other hand, adds to the aim of the empiricist to find out whether the hypothesis is referentially true, i.e., whether its referential claims are correct. In contrast to the *theory realist*, he is not interested in the question whether the theoretical claims, i.e., the claims using theoretical terms, are true as well.

Methodologies are ways of answering epistemological questions. It will turn out that the method of HD-testing, the test methodology, is functional for answering the truth question of all four epistemological positions. For this reason, we will present the test methodology in fairly neutral terms, viz., plausibility, confirmation and falsification.

The expression 'the plausibility of a hypothesis' abbreviates the informal qualification 'the plausibility, in the light of the background beliefs and the evidence, that the hypothesis is true', where 'true' may be specified in one of the four main senses: 1) observationally as far as the intended applications are concerned, 2) observationally, in all possible respects, 3) and, moreover, referentially, 4) and, even, theoretically. Admittedly, despite these possible qualifications, the notion of 'plausibility' remains necessarily vague, but that is what most scientists would be willing to subscribe to. At the end of this paper we will further qualify the exposition for the four epistemological positions when discussing the acceptance of hypotheses. When talking about 'the plausibility of certain evidence', we mean, of course, 'the prior plausibility of the (observational!) hypothesis that the test will result in the reported outcome'. Hence, here 'observationally true' and 'true' coincide by definition of what can be considered as evidential statements.

Regarding the notions of 'confirmation' and 'falsification' the situation is rather asymmetric. 'Falsification' of a hypothesis simply means that the evidence entails that the hypothesis is observationally false, and hence also false in the stronger senses. However, what 'confirmation' of a hypothesis precisely means, is not so clear. The explication of the notion of 'confirmation' of a hypothesis by certain evidence in terms of plausibility is the main target of this paper. It will be approached from the success perspective on confirmation, equating confirmation with an increase of the plausibility of the evidence on the basis of the hypothesis, and implying that the plausibility of the hypothesis is increased by the evidence.

According to the HD-method a hypothesis H is tested by deriving test implications from it, and checking, if possible, whether they are true or false. Each test implication has to be formulated in terms that are considered to be observational terms. A test implication may or may not be of general nature. Usually, a test implication is of a conditional nature, if C then F ($C \rightarrow F$). Here C denotes one or more initial conditions which can be, or have been realized, by nature or artificially, i.e., by experiment. F denotes a potential fact (event or state of affairs) predicted by H and C. If C and F are of individual nature, F is called an individual test implication, and $C \rightarrow F$ a conditional one. When C is artificially realized, it is an experimental test, otherwise it is a natural test.

Neglecting complications that may arise, if the test implication of H is false, H itself must be false, it has been falsified. When the test implication turns out to be true, H has not been (conclusively) verified, but it may still be true. However, we can say more than that. Usually it is said that H has been confirmed. It is important to note that such confirmation by the HD-method is in the strong sense that H has obtained a *success* of (conditional) deductive nature; by entailing the evidence, H makes the evidence as plausible as possible. This will be called *the success perspective* on ((conditional) deductive) *confirmation*.

One of the most influential theories of confirmation is the so-called Bayesian theory (Howson & Urbach 1989; Earman 1992), which is of quantitative nature. Unfortunately, quantitative formal methods necessarily have strong arbitrary elements. However, a quantitative theory may have adequate qualitative features. This paper deals with a qualitative theory of confirmation which is in agreement with the qualitative features of the corresponding version of the Bayesian theory. The non-standard presentation and analysis of the Bayesian theory, and the proof that it has these features, will be postponed to the companion paper. Since the main ideas of Carnap about confirmation (see Kuipers, 1997) concern a special form of the Bayesian theory, there is also qualitative agreement with him. Finally, there is largely agreement with Popper, for it will be argued in Appendix 1 of the companion paper that Popper's qualitative ideas about confirmation (or corroboration, to use his favorite term) are basically in agreement with the Bayesian approach.

However, there is at least one influential author about confirmation, viz., Glymour, whose ideas do not seem to be compatible with the qualitative theory to be presented. In a note to Section 2.3 we will briefly discuss Glymour's project, and question it as a project of explication of the concept of confirmation, rather than of theoretical measurement, which it surely is.

Guided by the success perspective on confirmation, Section 1 gives an encompassing qualitative theory of deductive confirmation by adding a comparative supplement to the classificatory basis of deductive confirmation. The comparative supplement consists of two principles. In Section 2 several classical problems of confirmation are dealt with. First, it is shown that the theory has a plausible solution of Hempel's raven paradoxes, roughly, but not in detail, in agreement with an influential Bayesian account. Second, it is shown that the theory has a plausible, and instructive, solution of Goodman's problem with 'grue' emeralds. Third, it is argued that further arguments against 'deductive confirmation' do not apply when conditional deductive confirmation is also taken into account. Section 3 concludes with some remarks about the problem of acceptance of well-confirmed hypotheses.

1. *A qualitative theory of deductive confirmation*

Traditionally, a qualitative theory of deductive confirmation is conceived as merely a classificatory theory. However, we conceive it from the outset as a combination of classificatory and comparative principles. We start with the classificatory principles.

1.1. *Classificatory basis*

The classificatory notions of evaluation of a hypothesis H by evidence E, are generated by considering the four possible, mutually exclusive (except for some extreme cases), deductive relations ($\models$: logical entailment) in a systematic order, assuming that H *is consistent and* E is true. They are depicted in the following *(Deductive) Confirmation Matrix* (Figure 1). The names will first be elucidated. To speak of falsification of H by E in the case H entails $\neg E$, $H \models \neg E$, which is equivalent to $E \models \neg H$, and of (conclusive) verification of H by E in the case $\neg H$ entails $\neg E$, $\neg H \models \neg E$, which is equivalent to $E \models H$, will be plausible enough. The other names are suggested by the success perspective on confirmation, that is, confirmation amounts to a success of a hypothesis. The first case, $H \models E$, is a paradigm case in which scientists speak of confirmation, viz., when E is a (hypothetico-)deductive success of H, for that reason called (the principle of) deductive (d-)confirmation. Note that only in this case E is a test implication of H (which came true). Just as verification (of H) amounts to falsification of $\neg H$, the remaining case, $\neg H \models E$, amounts to a deductive success of $\neg H$, hence to deductive confirmation of $\neg H$. However, speaking, by way of shorthand, precisely in this case of deductive *dis*confirmation seems also very plausible. Hence, whereas philosophers and scientists sometimes use the term disconfirmation as an euphemism for falsification, we use the formal opportunity to qualitatively distinguish sharply between (deductive) disconfirmation and falsification.

	E(true)	$\neg E$(false)
H	$H \models E(\neg E \models \neg H)$ Deductive Confirmation DC(E, H)	$H \models \neg E(E \models \neg H)$ Falsification F(E, H)
$\neg H$	$\neg H \models E(\neg E \models H)$ Deductive Disconfirmation DD(E, H)	$\neg H \models \neg E(E \models H)$ Verification V(E, H)

Figure 1: The (Deductive) Confirmation Matrix

The Confirmation Matrix specifies the basic cognitive structure governing hypothesis testing. However, although it gives the core of the classificatory part of the qualitative success theory of confirmation, it does not yet reflect standard scientific practice in detail. Recall that there occur as premises in hypothetico-deductive (HD-)prediction and deductive-nomological (DN-)explanation of individual events, besides the hypothesis, the so-called 'initial conditions'. The above definition of 'deductive confirmation', as confirmation by a deductive success, directly suggests the notion of 'conditional deductive (cd-)confirmation', i.e., confirmation of H by E, assuming a certain condition C, defined as follows:

$H\&C \models E$ conditional Deductive Confirmation: DC($E, H; C$)

Here it is assumed that H and C are logically independent (LI(H, C)), for reasons becoming clear in Section 2.3, and that C does not entail E, whereas C may or may not be entailed by E. Hence, in view of the fact that hypotheses and initial conditions usually are logically independent, successful HD-prediction and DN-explanation of individual events form the paradigm cases of cd-confirmation, for they report 'conditional deductive successes'.

When dealing with specific examples, several expressions related to cd-confirmation will be used with the following interpretation:

> "*E C-confirms H*": E cd-confirms H on the condition (indicated by) C,
> "*E cd-confirms H*": there is a (non-tautological) condition C, such that E entails C and E C-confirms H,
> "E is *deductively (d-)neutral (evidence)* for H": H and E are logically independent (i.e., none of the four deductive relations holds),
> "E is *conditional deductively (cd-)neutral (evidence)* for H": E is d-neutral for H and *in addition* E does not cd-confirm H nor $\neg H$.

Note that 'cd-neutrality' trivially implies 'd-neutrality', but not vice versa.

Of course, 'conditionalization' of the other three deductive relations, by just adding C as a premise, is also possible and more realistic, leading to the corresponding *conditional* (Deductive) Confirmation Matrix.

1.2. *Comparative supplement*

As announced, we explicate 'confirmation by the HD-method', in this paper, as (conditional) deductive confirmation including some crucial comparative principles. However, in order to do so it is useful to introduce first some main principles of general confirmation, that is, confirmation of deductive and non-deductive nature. The success perspective on confirmation in general may be explicated by the following definition of (general) confirmation:

SDC: Success definition of confirmation
E confirms H iff (E is a success of H in the sense that) H makes E more plausible

This definition is satisfied by d-confirmation in the extreme sense that E is made true, hence maximally plausible, by (the truth of) H. The second principle deals with the updating of the plausibility of hypotheses:

UPP: Updating principle of plausibility
E makes H more plausible[1] iff E confirms H

For d-confirmation the updating principle implies an increase of the plausibility of the hypothesis.

Note that the two general principles imply:

PS: Principle of symmetry
H makes E more plausible iff E makes H more plausible

Note, moreover, that the combination of the first two principles makes comparative expressions of the form "$E*$ confirms $H*$ more than E confirms H" essentially ambiguous. It can express that the plausibility increase of $E*$ by $H*$ is higher than that of E by H, or that the plausibility increase of $H*$ by $E*$ is higher than that of H by E. However, we exclude this ambiguity by adopting as the third and last general principle, now of a comparative nature:

PCS: Principle of comparative symmetry
$E*$ confirms $H*$ (much) more than (as much as) E confirms H
iff $H*$ increases the plausibility of $E*$ (much) more than (as much as) H increases the plausibility of E
iff $E*$ increases the plausibility of $H*$ (much) more than (as much as) E increases the plausibility of H

For our purposes, these four principles of general confirmation are sufficient. From now on in this paper, comparative confirmation claims will always pertain to (conditional) *deductive* success and (conditional) *deductive* confirmation.

[1] In the companion paper we will see that the 'if-direction' fails in the quantitative theory of confirmation for hypotheses with probability zero. Similarly, the 'only if-direction' in PS below fails for such hypotheses. Hence, one might refine UPP and PS by imposing the condition that H has some initial plausibility.

Now we are able to propose two comparative principles concerning d-confirmation, one (P.1) for comparing two different pieces of evidence with respect to the same hypothesis, and one (P.2) for comparing two different hypotheses in the light of the same piece of evidence.

P.1 if E and E^* d-confirm H then E d-confirms H more than E^* iff E is less plausible than E^* in the light of the background beliefs

P.2 if E d-confirms H and H^* then E d-confirms H^* as much as H

To be sure, P.1 and P.2 are rather vague, but we will see that they have some plausible applications, called the special principles. Hence, if one finds the general principles too vague, it is suggested that one primarily judges the special principles.

The motivation of the plausibility principles is twofold. First, our claim is that P.1 and P.2 are roughly in agreement with scientific common sense concerning confirmation by successes obtained from HD-tests. For P.1 this is obvious: less expected evidence has more 'confirmation value' than more expected evidence. P.2 amounts to the claim that hypotheses should be equally praised for the same success, which is pretty much in agreement with (at least one version of) scientific common sense.

The second motivation of the principles P.1 and P.2 will be postponed to the companion paper, where we will show that they result, with some qualifications, from certain quantitative confirmation considerations of Bayesian nature applied to HD-tests, Bayesian considerations for short, when 'more plausible' is equated with 'higher probability'.

P.2 amounts to the claim that the 'amount of confirmation', more specifically, the increase of plausibility by evidence E is independent of differences in initial plausibility between H and H^* in the light of the background beliefs. In the companion paper we will also deal with some quantitative theories of confirmation for which holds that a more probable hypothesis profits more than a less probable one. This may be seen as a methodological version of the so-called Matthew-effect, according to which the rich profit more than the poor. It is important to note, however, that P.2 does not deny that the resulting posterior plausibility of a more plausible hypothesis is higher than that of a less plausible one, but the increase does not depend on the initial plausibility. For this reason, P.2 and the resulting theory are called neutral with respect to equally successful hypotheses or, simply, *pure*[2] ,

[2] In view of Sober (1995, p. 193), P.2 might be called a Humean principle. However, it certainly does not cover all of Humean skepticism regarding confirmation. As we will see in the companion paper, P.2 leaves perfectly room for (systems of) *inductive* confirmation.

whereas theories having the Matthew-effect, or the reverse effect, are called *impure*. The first type of impure theories are said to favor plausible hypotheses, whereas theories of the second type favor implausible hypotheses. In Section 2.1 of the companion paper, we will give an urn-model, hence objective probabilistic, illustration and defence of P.2.

The condition 'in the light of the background beliefs' in P.1 is in line with the modern view (e.g., Sober 1988, p. 60) that confirmation is a three place relation between evidence, hypothesis and background beliefs, since without the latter it is frequently impossible to make differences between the strength of confirmation claims. As a matter of fact, it would have been better to include the background beliefs explicitly in all formal representations, but we have refrained from doing so to make reading the formulas more easy.

This does not mean that background beliefs always play a role. For instance, they do not play a role in the following two applications of the principles, that is, the first two special principles (for non-equivalent E and E^*):

S.1 if $H \models E \models E^*$ then E d-confirms H more than E^*

S.2 if $H^* \models H \models E$ then E d-confirms H^* as much as H

These special principles obviously are applications of the corresponding general principles since (logically) weaker evidence and weaker hypotheses may be assumed to be more plausible than stronger versions. They can be motivated in the same two ways as the general principles themselves: indirectly, by showing (in the companion paper) that they result from certain Bayesian considerations as soon as 'logically weaker' guarantees a higher prior probability, and directly, by referring to scientific common sense around HD-tests. S.1 states, in line with the success perspective, that a (logically) stronger deductive success of a hypothesis confirms that hypothesis more than a weaker success. S.2 states that a logically stronger hypothesis is as much confirmed by a deductive success as a weaker one which shares that success.

Since S.1 and S.2 directly pertain to the rebuttal of two standard objections in the literature against deductive confirmation, we will deal already now with these objections, but postpone the treatment of some other ones to Section 2.3. It is easily checked that deductive confirmation has the so-called 'converse consequence property' with respect to the hypothesis (CC-H) and the 'consequence property' with respect to the evidence (C-E). The first property amounts to:

CC-H: converse consequence property with respect to H
if E d-confirms H then it also d-confirms any stronger H^*

This property generates the 'irrelevant conjunction objection', i.e., CC-H has the *prima facie* absurd consequence that if $H \models E$ then E d-confirms $H\&H'$, for any H', which is compatible with H, but not entailed by H (Hempel 1945/1965); Glymour 1980a/b). From the classificatory-cum-comparative point of view, this consequence is not at all absurd as soon as we are aware of all relevant (qualitative) confirmation aspects, which we like to call the *proper (conjunction) connotation:*

if E d-confirms H then
- E d-confirms $H\&H'$, for any H' compatible with H
- even as much as H (due to S.2)
- but E does not at all necessarily d-confirm H'
- hence, the d-confirmation remains perfectly localizable

If one finds it strange that E confirms $H\&H'$ as much as H, it is important to realize that the prior plausibility of $H\&H'$ will be less than that of H, and hence, because of PCS and S.2, this will hold for the corresponding posterior plausibilities. Hence, if H' is very irrelevant and implausible, e.g. the moon is made of cheese, $H\&H'$ will be implausible as well, a priori as well as a posteriori.

The situation is to some extent similar for the standard objection against the second property, C-E, the consequence property with respect to the evidence. This property amounts to:

C-E: consequence property with respect to the evidence
if E d-confirms H, then also any weaker E*

This property generates what might be called the "irrelevant disjunction objection", i.e., C-E (see above) has the *prima facie* absurd consequence that if $H \models E$ then $E \vee E'$ d-confirms H, for any E' (Grimes 1990). Again, from our point of view this consequence is not at all absurd as soon as we are aware of all relevant (qualitative) confirmation aspects, i.e., the *proper (disjunction) connotation:*

if E d-confirms H then
- $E \vee E'$ d-confirms H, for any E' compatible with E
- though (much) less (due to S.1)
- but E' does not at all necessarily d-confirm H
- hence, the d-confirmation remains perfectly localizable

It will be useful to conditionalize the general and special principles:

P.1c if E C-confirms H and $E*$ $C*$-confirms H then E C-confirms H more than $E*$ $C*$-confirms H iff $E*$ is, given $C*$, more plausible than E, given C, in the light of the background beliefs

S.1c if $H \models C \rightarrow E \models C* \rightarrow E*$ then E C-confirms H more than $E*$ $C*$-confirms H

P.2c if E C-confirms H and $H*$ then E C-confirms $H*$ as much as H

S.2c if $H \models C \rightarrow E$, $H* \models C \rightarrow E$ and $H* \models H$ then E C-confirms $H*$ as much as H

Later we will introduce two other applications of the principles, more specifically of P.1c and P.2c, or, if you want, new special principles. They will provide the coping-stones for the solution of the raven paradoxes and the grue problem. In contrast to S.1 and S.2, they presuppose background beliefs and concern *conditional* (deductive) confirmation. Moreover, they deal with special types of hypotheses, pieces of evidence and conditions. They will be indicated by $S^{\#}$.1c(-ravens) and S^{Q}.2c(-emeralds), respectively.

2. *Ravens, grue emeralds, and other problems and solutions*

In this section it will be shown how the qualitative theory of deductive confirmation resolves the famous paradoxes of confirmation presented by Hempel and Goodman. Moreover, we will deal with the main types of criticism in the literature against the purely classificatory theory of deductive confirmation, that is, when a comparative supplement is absent.

2.1. *The raven paradoxes*

Hempel (1945/1965) discovered two paradoxes about confirmation of the hypothesis "all ravens are black" (RH) on the basis of two *prima facie* very plausible conditions. Assuming that a black raven confirms RH (so-called Nicod's criterion (NC)) and that the logical formulation of RH may not matter (equivalence condition (EC)), Hempel derived not only that a non-black non-raven (the first paradox), but, even more counter-intuitive, also a black non-raven (the second paradox) confirms RH in the same sense as a black raven. First, according to NC, a non-black non-raven confirms "all non-blacks objects are non-ravens" and hence, according to EC, RH itself. Second, again according to NC, a black non-raven confirms "all objects are non-ravens or black" and hence, according to EC, RH itself.

It is easy to check that the previous section leads to the following classificatory results:

(1) a black raven, a non-black non-raven and a black non-raven are all deductively (d-)neutral evidence for RH

(2) a black raven and a non-black non-raven both cd-confirm RH, more specifically, a black raven on the condition of being a raven and a non-black non-raven on the condition of being non-black

(3) a black non-raven is even cd-neutral evidence for RH, i.e., not only just deductively, but even conditional deductively, for a black non-raven does not cd-confirm RH on the condition of being a non-raven, nor on the condition of being black.

All three results are in agreement with scientific common sense, provided the following comparative claim can also be justified:

(4) a black raven cd-confirms RH (much) more than a non-black non-raven

However, assuming that the background beliefs include or imply the assumption, where #R indicates the number of R's, etc.:

A-ravens: the number of ravens is much smaller than the number of non-black objects ($\#R \ll \#\overline{B}$)

the desired result, i.e., (4), immediately follows from the third special principle, viz., the following general application of P.1c. Though the symbols are suggested by the raven example (e.g., an RB may represent a black raven, i.e., a raven which is black), they can get any other interpretation.

$S^{\#}$.1c(-ravens): an RB R-confirms "all R are B" more than an $\overline{RB}$ $\overline{B}$-confirms it iff the background beliefs imply that $\#R < \#\overline{B}$

$S^{\#}$.1c realizes in a precise sense the intuition that a black raven confirms "all ravens are black" (much) more than a non-black non-raven. That $S^{\#}$.1c is an application of P.1c can be shown as follows. If $\#R < \#\overline{B}$ and "all R are B" is false, then the percentage of RB's among the R's is lower than the percentage of $\overline{RB}$'s among the $\overline{B}$'s, hence hitting among the R's at an R which is B is less plausible (hence more surprising) than hitting among the $\overline{B}$'s at a $\overline{B}$ which is $\overline{R}$. For example, even in the extreme case of just one non-black raven, the first percentage is #R/(#R+1), which is indeed smaller than the second percentage, $\#\overline{B}/(\#\overline{B}+1)$, if and only if #R is smaller than $\#\overline{B}$. In other words, the higher the percentage of individuals with a certain trait in a population, the more it is (to be) expected that it applies to an arbitrary member, where only the comparison of percentages is relevant, and not the trait nor the (size of the) population.[3] Hence, $S^{\#}$.1c can be motivated by referring directly to scientific common sense, but also indirectly by showing, in the companion paper, that it results from Bayesian considerations when the evidence is assumed to be obtained by random sampling in the relevant universe. Of course, we speak of *much* more confirmation in $S^{\#}$.1c when the background beliefs imply that #R is *much* smaller than $\#\overline{B}$ ($\#R \ll \#\overline{B}$), as in the case of A-ravens, and hence, if RH is false, the percentage of RB's among the R's is (relatively speaking) *much* lower than the percentage of $\overline{RB}$'s among the $\overline{B}$'s. Precisely because typical applications of $S^{\#}$.1c concern such cases, it is defensible to call it a qualitative application and principle, despite its explicit reference to numbers of individuals and the reference to percentages in the motivation.[4]

In sum, cd-confirmation solves the raven paradox concerning black non-ravens by (3) and the one concerning non-black non-ravens by (4), which is guaranteed by applying P.1c, or its application $S^{\#}$.1c-ravens, to the background assumption (A-ravens) that the number of ravens is much smaller than the number of non-black objects.

There remains the question of what to think of Hempel's principles used to derive the paradoxes of confirmation. It is clear that the equivalence condition was not the problem, but Nicod's criterion that a black raven confirms RH *un*conditionally. Whereas Nicod's criterion is usually renounced unconditionally, we may conclude that it is (only) right in a sophisticated sense: a black raven is a case of cd-confirmation, viz., on the condition of being a raven.

[3] It is also possible to give a 'syntactic' motivation of S.1c. Note first that "all R are B" has two conjunctive versions corresponding to the two types of conditional confirmation, viz. the finite conjunction ranging over all R, telling for each that it is a B, hence with #R conjuncts, or as the finite conjunction ranging over all $\overline{B}$, telling for each that it is a $\overline{R}$, with $\#\overline{B}$ conjuncts. If $\#R < \#\overline{B}$ then each of the first kind of conjuncts has, relatively speaking, a greater share in the first finite conjunction than each of the second kind of conjunctions has in the second finite conjunction. Hence, the first ones contribute more to the complete verification of "all R are B" than the second ones.

[4] The presented solution of the raven paradox is the qualitative version of an improved version of Horwich's Bayesian solution, to be presented in the companion paper.

2.2. *Grue emeralds*

The qualitative theory of deductive confirmation generates an instructive analysis of the other famous riddle of confirmation, i.e., the problem with 'grue' emeralds, discovered by Goodman (1955). This problem is also called the grue 'paradox', for the same reason as one speaks about the raven paradoxes. They both concern counter-intuitive consequences of certain principles of confirmation.

The problem with grue emeralds is that a green emerald found before the year 3000 seems to confirm not only the hypothesis that all emeralds are green but also that all emeralds are 'grue', where grue is defined as the following queer predicate: green if examined before 3000, and blue if not examined before 3000. Goodman's generally accepted account roughly is as follows: the predicate 'grue' is not well-entrenched in predictively successful scientific generalizations, hence, as it stands, it is below the mark of scientific respectability to be used in generalizations that can be confirmed, i.e., to use Goodman's other favorite expression, the 'grue-hypothesis' is not (yet) projectable. We will give a related, but more detailed diagnosis of the problematic aspect. It may well be conceived as a formal explication and justification of Goodman's informal account. It can best be presented by using from time to time a formally similar, but less queer, definition of 'grue', which is only applicable to living beings, say, eagles: 'being male and green, or being female and blue'. This will be called the 'gender' reading, as opposed to the former 'temporal' reading.

Recall that we use the abbreviation: "...C-confirms ..." as a shorthand for "... cd-confirms ... on the condition (indicated by) C". We add the abbreviations: E: emerald/eagle (in this subsection not to be confused with 'evidence'), M: (known to be) examined before 3000/male, $\overline{\text{M}}$: not (known to be) examined before 3000/female, G: green, B: blue, and Q: grue (queer), i.e., MG or $\overline{\text{M}}$B. G and B are supposed to be mutually exclusive, but they are not supposed to be exhaustive.

We will first specify in detail to what extent 'green' and 'grue' are similar, and show that additional assumptions are needed to create the intuitively desirable asymmetry. More specifically, it will be shown that not only a strong, but also a weak irrelevance assumption is suitable for this purpose. Both are in line with the entrenchment analysis of Goodman.

The basic intuition

The basic intuition of Goodman is, of course, that, though a green emerald investigated before 3000 confirms 'the green hypothesis' ("all E are G"), it does not confirm 'the grue hypothesis' ("all E are Q"). It postulates an asymmetry in confirmation behavior between 'green' and 'grue'. However, from the unconditional version of Nicod's criterion, 'Nicod-confirmation', we not only get

(1) an EMG Nicod-confirms "all E are G"

but also, as is easy to check,

(2) an EMG Nicod-confirms "all E are Q"

Hence Nicod's criterion excludes an account of the asymmetric intuition, which shows an additional problematic feature of that criterion.

As may be expected, from the straightforward, that is, unconditional deductive point of view, both cases are invalid, for in both cases the evidence is deductively neutral for the hypothesis. Hence, unconditional deductive confirmation also fails to account for the basic intuition, but it does not exclude a 'conditional deductive' account. So, what about conditional deductive confirmation?

Cd-confirmation, however, is also in conflict with the intuition, for it is easy to check that the following classifications obtain:

(3) an EMG EM-confirms "all EM are G"
(4) an EMG EM-confirms "all E are G"
(5) an EMG EM-confirms "all E are Q"

Note first that (3) is an unproblematic formal analogue of the equally valid claim that an EG E-confirms "all E are G", where E is replaced by EM. More importantly, whereas (4) fits the basic intuition, (5) is in conflict with it. However, before rejecting cd-confirmation because of (5), it is important to study the validity of (5) in detail. To begin with, it is important to note that the following claim is invalid:

(6*) an EMG EM-confirms "all $E\overline{M}$ are B"

The invalidity of (6*) amounts to the claim that a green object does not confirm the hypothesis "all $E\overline{M}$ are B" on the condition that it is an emerald investigated before 3000. In this light, we suggest that the problem with (5) derives from the wrong impression that it implies (6*) and the intuition that it would indeed be absurd if (6*) were to obtain. In other words, we take the invalidity of (6*) as the proper interpretation of the confirmation-rejecting side of the basic intuition, instead of the originally, but wrongly, suggested invalidity of (5).

It is interesting to see in more detail how (5), (6*), and (3) are related. Note that the following equivalence holds:

(7) “all E are Q” $\Leftrightarrow$ “all EM are G” & “all E$\overline{\text{M}}$ are B”

which makes clear, by the way, that “all E are Q” may be less strange than the temporal reading suggests, for in the gender reading it is the conjunction of two descent hypotheses, viz., “all male eagles are green” and “all female eagles are blue”. In view of (7), (5) is equivalent to:

(5) an EMG EM-confirms “all EM are G” & “all E$\overline{\text{M}}$ are B”

Now it is easy to see that (3), (5) and (6*) form an illustration of the conditional version of the ‘proper conjunction connotation’ of Section 1. That is, whereas cd-confirmation according to (3) transmits to a conjunction with some other hypothesis according to (the equivalent version of) (5), and the latter confirmation is, according to S.2c, even as much as the former, no confirmation transmits to the added conjunctive hypothesis, for (6*) is invalid.

The question why (6*) would be absurd if valid, however, remains interesting. Is it, *prima facie* in line with Goodman’s entrenchment considerations, because it amounts to a surprising prediction across some clearly defined border (a year, gender), breaking the continuity of nature? In this case, it would be plausible to expect that the additional hypothesis suggested by continuity considerations, viz., “all E$\overline{\text{M}}$ are G”, is EM-confirmed by an EMG, since G is well-entrenched, and the trouble with (5) would merely be caused by the queer, non-entrenched character of Q. Or is it because the ‘grue-induced’ additional hypothesis “all E$\overline{\text{M}}$ are B” reaches incautiously over a border that might be a relevant distinction? In this case also the ‘green-induced’ additional hypothesis should not be EM-confirmed by an EMG, and the usual, but wrong, assumption that this is implied by (4), is brought to light by the queer predicate.

It is easy to check that the second option is the proper answer from the perspective of cd-confirmation. That is,

(8*) an EMG EM-confirms “all E$\overline{\text{M}}$ are G”

is invalid, for the same reason as (6*): the antecedence E$\overline{\text{M}}$ of the hypothesis ‘cannot be put to work’ by the condition EM to derive G and B, respectively. This suggests that (8*) can also provide an illustration of the proper conjunction connotation. Note, for this purpose, that the following equivalence obtains:

(9) “all E are G” $\Leftrightarrow$ “all EM are G” & “all E$\overline{\text{M}}$ are G”

and hence that (4) is equivalent to

(4) an EMG EM-confirms "all EM are G" & "all $E\overline{M}$ are G"

Accordingly, in spite of the validity of (3) and (4), (8*) is invalid, precisely for the same reason that (6*) is invalid as opposed to (3) and (5), viz., being another triple of instances of the proper conjunction connotation. Conditional deductive confirmation (3) transmits to a conjunction with some other hypothesis (4) and this confirmation is as much as that of (3), according to S.2, but the confirmation does not transmit to the added conjunctive hypothesis (8*).

In sum, the *prima facie* absurdity of (5) has a hidden analogue in (4), which is also due to improper connotations. The invalidity of (8*) is a formal blockade against confirmation claims that cross a border that may be relevant:

> *cd-confirmation blockade:* For all E, M and G, although an EMG EM-confirms "all EM are G"(3) and even "all E are G"(4), it does not EM-confirm "all $E\overline{M}$ are G" (8*).

Whereas the blockade may seem superfluous in the temporal reading, it is clear that, for instance, in the gender reading it is highly plausible and desirable. A green male eagle does not (cd-)confirm the hypothesis that all female eagles are green. Awareness of the blockade is, for instance, expressed in the feminist criticism of male oriented drug research. Results of testing drugs on male subjects have frequently been extrapolated to women in an irresponsible way (see e.g. Cotton, 1990; Ray et al., 1993).

So far, however, the results of the conditional deductive perspective are symmetric with respect to green and grue: (4) and (5) on the one hand, and (6*) and (8*) on the other. Moreover, (3) is a common feature of both. Hence, the question remains to account for the asymmetric basic intuition. The foregoing analysis shows that an additional assumption is needed to create an asymmetric situation.

There are at least two possible ways. In the first way, a hypothesis is added which removes the blockade, at the expense of the grue hypothesis. In the second way, the blockade is not removed, but the grue hypothesis is downgraded, without excluding it.

Asymmetry by an extra assumption

According to the first way, we explicitly assume as an extra hypothesis, that the border is irrelevant for the kind of properties at stake, that is, that they can be extrapolated across that border. In the temporal reading of M this amounts to adding a strong irrelevance assumption of time for the color of emeralds, formally:

SIA(-emeralds): for all colors C, "all EM are C" implies "all E are C", and hence "all E$\overline{\text{M}}$ are C", and vice versa.

This assumption may well be a background belief. The 'implication' in SIA is stronger than just a material implication and weaker than a purely logical implication. In particular, it may be an analytical (or semantical) implication. However, it may also be a 'physical implication', in the sense of a physical necessity. In the latter case, it may be based on the underlying assumption that emeralds constitute a natural kind with respect to color. Note first that SIA, in view of (7), excludes the grue hypothesis "all E are Q". Moreover, an EMG now not only EM-confirms "all EM are G" (3), and hence also "all E are G" (4), but even "all E$\overline{\text{M}}$ are G", i.e, the adapted version of (8*) becomes valid:

(8-SIA) an EMG (SIA & EM)-confirms "all E$\overline{\text{M}}$ are G"

This is in agreement with the intuition behind the grue problem that the artificial time barrier will not change the color. However, according to the cd-analysis, this is only guaranteed when we take this formally into account by the auxiliary assumption SIA. In conjunction with SIA it is even guaranteed that an EMG not only falsifies "all EM are B", and hence "all E are B", but also "all E$\overline{\text{M}}$ are B", for the latter and SIA now imply "all E are B", and hence, in view of (7), an EMG falsifies "all E are Q". Hence, instead of the invalid confirmation claim (6*) that an EMG EM-confirms "all E$\overline{\text{M}}$ are B", we may now even conclude that an EMG 'SIA-falsifies' that hypothesis in the plausible sense that (EMG & SIA) is incompatible with the hypothesis:

(6-SIA) an EMG SIA-falsifies "all E$\overline{\text{M}}$ are B"

Accordingly, one way to achieve an asymmetry between green and blue in the line of the cd-perspective is by assuming SIA. This is very much in the spirit of Goodman's entrenchment analysis in terms of so-called projectible predicates.[5] However, whereas Goodman's notion remains rather vague, the required irrelevance assumption to remove the surprising blockade highlighted by the invalid (8*), and giving rise to the valid confirmation claim (8-SIA), is crystal clear. At the same time, it allows the replacement of the appealing blockade reported by the invalid (6*) even by the valid falsification claim (6-SIA). Hence, we have obtained an asymmetric (cd-)explication of Goodman's basic intuition and a dichotomous reading of his entrenchment analysis. However, this is not satisfactory in all cases. In the gender reading,

[5] It may even be argued that SIA explicates Goodman's projectibility condition: a necessary condition for being a color (with respect to emeralds) is satisfying SIA.

we do not want to exclude relevance of sex for color, although we may have good reasons to find irrelevance more likely than relevance. For this reason, it is plausible to look for a possible refinement of the basic intuition, which also leads to asymmetry, without totally excluding the grue hypothesis from the confirmation game. For, as is clear from the gender reading of the grue hypothesis, there may well be formally similar cases where the exclusion is highly debatable.

Asymmetry by refinement of the basic intuition
The following refinement of Goodman's basic intuition is (the qualitative analogue of a quantitative refinement) inspired by Sober (1994). Reconsider first:

(4) an EMG EM-confirms "all E are G"
(5) an EMG EM-confirms "all E are Q"

which are equivalent to

(4) an EMG EM-confirms "all EM are G" & "all E$\overline{\text{M}}$ are G"
(5) an EMG EM-confirms "all EM are G" & "all E$\overline{\text{M}}$ are B"

The *prima facie* version of the basic intuition stated that, whereas (4) is correct, (5) is problematic. Above we have seen that (5) is not problematic, but also that (6*) would be problematic, if valid, which it is not. This creates the room for the following refinement of the basic intuition: a green emerald investigated before 3000 may confirm "all E are Q" as well as "all E are G", even as much as, but the resulting plausibility of the former is (remains) much lower than that of the latter, i.e., in the relevant conditional sense:

(4&5) although an EMG EM-confirms "all E are G" as much as "all E are Q", the resulting plausibility of the former is much less than that of the latter

In order to justify this refined intuition we introduce the fourth special principle, a general application of P.2c, with symbols suggesting the example, but of course intended for general use (as in the case of S^Q.1c):

S^Q.2c(-emeralds): an EMG EM-confirms "all E are Q" as much as "all E are G"

It is clear that S^Q.2c is a straightforward application of P.2c, using the fact that an EMG EM-confirms both hypotheses. Moreover, in the emerald version, it is safe to assume as background belief the following weak irrelevance assumption:

WIA(-emeralds): for all colors C and C′, C≠C′, "all E are C" is (much) more plausible than the conjunction "all EM are C" & "all E$\overline{\text{M}}$ are C′" (which is equivalent to "all E are Q" when C=G and C′=B)

In view of the fact that (4) and (5) hold, we may apply the general principle of comparative symmetry (PCS) and P.2c, or its application S^Q.2c, to our background belief WIA, which directly leads to the asymmetric cd-explication of the refined intuition, that is, (4&5).

Note that SIA implies WIA as soon as we assume that SIA amounts to the implication that grue-like hypotheses lack any plausibility, whereas green-like hypotheses have at least some plausibility. Hence, in the light of SIA, (4&5) provides an asymmetry additional to that between (6-SIA) and (8-SIA).

In the gender reading, however, only WIA may have some plausibility, but not SIA. That is, it may well be that we would like to subscribe to the background belief that the green hypothesis is (much) more plausible than the grue hypothesis, without excluding the latter. The reason would be, of course, that a systematic color difference between the sexes of a species regularly occurs, though supposedly not as frequently as sex irrelevance for color. Even in the temporal reading WIA defensible, and SIA not. It is surely the case that, as far as we know, there are no types of stones that have changed color at a certain moment in history. However, this does not exclude the possibility that this might happen at a certain time for a certain type of stone, by some cosmic event. To be sure, given what we know, any hypothesis which presupposes the color change is much less plausible than any hypothesis which does not.

It is important to note that we cannot simply take as an assumption that "all E are G" is more plausible than "all E are Q", that is, without reference to background beliefs, but only with the motivation that the former generalization expresses more uniformity or continuity of nature than the latter. That is, it may seem that adding "all E$\overline{\text{M}}$ are G" to the common generalization of "all E are G" and "all E are Q", viz., "all EM are G", giving rise to "all E are G", is more in line with that common generalization than adding "all E$\overline{\text{M}}$ are B", giving rise to "all E are Q". If one thinks this way one does so because one assumes that E$\overline{\text{M}}$G is more similar to EMG than E$\overline{\text{M}}$B. However, this 'uniformity argument' hinges upon the particular E/M/G-language. As is well-known from the discussion of the grue problem (and of the problem of language dependence of definitions of verisimilitude, see Zwart (1995)) such arguments are language dependent. More specifically, in the E/M/G-language the suggested uniformity argument would imply that "all E are G" is, for example, more plausible than “for all E: M iff G”. However, in the E/M/X-language, with X $=_{\mathrm{def}}$ M iff G, the suggested uniformity argument would imply the opposite plausibility claim, viz., that “all E are X” is more plausible than “for all E: M iff X”. Hence, reference to background beliefs is unavoidable.

In sum, the basic intuition can be justified in terms of cd-confirmation in two ways. In the first way, a strong assumption is added which excludes the grue hypothesis. In the second way, the basic intuition is refined by downgrading the grue hypothesis, without excluding it. Both are very much in the spirit of Goodman’s entrenchment approach. Although Goodman’s specific example of grue primarily suggests the first, dichotomous way, since green is, and grue is not, well-entrenched (in the temporal reading). However, his general exposition in terms of more or less entrenched predicates primarily suggests the second, gradual way. As has been noted, the first way is a kind of extreme version of the second.[6] Hence, the above analysis is highly congenial to Goodman’s informal account in terms of entrenchment. However, the cd-analysis localizes in formal detail the symmetric point of departure for two asymmetric explications, a stronger and a weaker one.

2.3. *Objections to (conditional) deductive confirmation*

In the literature several objections have been expressed to the very idea of (conditional) deductive confirmation. Hence, our specific account of (un-)conditional deductive confirmation is also subject to them. The reader is invited to himself evaluate our rebuttals. We begin with unconditional deductive confirmation. Two of the three standard objections against this idea have already been dealt with in Section 1.2, viz., this type of confirmation is transmitted to a stronger hypothesis (CC-H), and to weaker evidence (C-E), leading to the two ‘proper connotations’ which also take comparative claims into account.

An important remaining objection against *unconditional* deductive confirmation is that d-confirmation is not transmitted to a weaker hypothesis (lacks the consequence property with respect to the hypothesis). Hempel (1945/1965) and others point at the supposed intuition among scientists that confirmation transmits to consequences of the hypothesis. From the success perspective on deductive confirmation, i.e., deductive confirmation as deductive success, this objection also disappears, for a consequence of H

[6] In the extreme quantitative case, we will get back the dichotomous way by assigning the grue hypothesis the prior probability 0 such that the posterior probability remains 0 (see the companion paper).

need not yield the success, it even need not be co-responsible for the success. Hence, the straightforward consequence property, viz. transmission of d-confirmation to a weakening of the hypothesis, is invalid as well as implausible.

One might argue that a liberation of the property holds: d-confirmation of H by E implies confirmation, not necessarily deductive, of any consequence of H by E. However, if deductive confirmation were to have the liberated consequence property, any success, i.e., any established true consequence of a hypothesis would confirm any other consequence of it, even if the two consequences use two non-overlapping (sub-)languages, and would hence not be recognized as confirmationally related at all without the hypothesis. More generally, together with the converse consequence property with respect to the hypothesis (CC-H) of d-confirmation, the liberated property would have the absurd consequence that any two hypotheses with no overlapping vocabulary would always be confirmationally related in the sense that any d-confirming evidence of the one would be, via the conjunction of the two hypotheses, confirming evidence of the other. In short, most scientists are well aware that a success usually is a joint venture of most, if not all, of the components of the hypothesis. Hence, if some scientists and philosophers nevertheless have also the intuition that (deductive) confirmation transmits somehow to all consequences of the hypothesis, the foregoing analysis, dealing with deductive successes as the paradigm cases of confirmation, provides an invitation to reconsider that connotation of (deductive) confirmation seriously.[7]

As to the predictable objections against *conditional* deductive confirmation, we start with a purely technical objection (Gemes 1990, and formally related, Glymour 1980b). It is easy to check that the definition of DC(E, H; C), i.e., $H\&C \models E$, *prima facie* implies that, for any E and H, E cd-confirms H on a condition that is entailed by E, hence true, viz., the condition $H \rightarrow E$, i.e., DC($E, H; H \rightarrow E$), for trivially $H\&(H \rightarrow E) \models E$. Hence, *prima facie*, any E cd-confirms any H. However, we have added the requirement LI(H, C) to the definition, i.e., the condition that C does not logically depend on H. Since not only E but also $\neg H$ (trivially) entails the relevant condition $H \rightarrow E(\neg H \vee E)$, the purported trivialization of cd-confirmation has thus been excluded. Note that the same component of LI(H, C), that is, $\neg H$ may not entail C, is formally required to prevent

[7] The tension between the two potential principles, C-E and CC-H, is for Flach (1995) the reason to distinguish and elaborate, in his terms, two kinds of induction, confirmatory induction, obeying C-E and not CC-H, and explanatory induction, obeying CC-H but not C-E. Since the latter concept is in our view the closest to the scientific common sense concept of confirmation, we restrict attention to that type of explication.

HD-prediction and DN-explanation of individual events from similar trivialization. Of course, the other alternative is just to consider these related trivial cases as illustrations of the trivialization phenomenon that frequently occurs under logically extreme conditions. However, the other three components of LI(H, C) serve already to exclude more transparent improper cases, viz., incompatible hypothesis and condition (H may not entail $\neg C$), or one of them being redundant (H may not entail C and C may not entail H), so why not just include the fourth component?

Let us now turn to a *prima facie* more fundamental objection. First of all, it is time to mention that the notion of cd-confirmation was already essentially considered, and rejected, by Hempel (1945/1965) as a general criterion of confirmation, under the heading 'prediction criterion'. Later, it was reconsidered by Horwich (1983), and rejected as too narrow, which it of course is when taken as *pars pro toto*, i.e., without taking unconditional confirmation, also treated in this paper, and non-deductive confirmation, treated in the companion paper, into account. Hempel argues that in interesting cases (he elaborates a case of plane-polarized light) the plausible condition is already a universally quantified statement, of which the acceptance presupposes, what he calls, a quasi-induction. Assuming moreover that the idea of quasi-induction presupposes the idea of confirmation, he concludes that the predictive approach to confirmation becomes circular. From our analysis it is clear, however, that not only unconditional deductive confirmation but also conditional deductive confirmation can, without problems, be defined in general, and can be straightforwardly applied to non-universal conditions. Hence, it may be true that the acceptance of a universal condition requires criteria for the 'inductive jump' and it is even plausible that these criteria will be phrased in terms of straightforward cases of (conditional) deductive confirmation. However, it does not follow from this that the analysis of (conditional) deductive confirmation is circular.[8]

To conclude, we have argued that confirmation by the HD-method can be explicated as (conditional) deductive confirmation, provided some comparative principles are added to this classificatory point of departure. The resulting theory not only leads to qualitative solutions of the raven paradoxes

[8] Hempel's argument against cd-confirmation resembles a general, but invalid, circularity objection to confirmation. In the next section we will argue that confirmation increases plausibility and that high plausibility leads to acceptance and acceptance to inclusion in the background beliefs. Although background beliefs are frequently presupposed in confirmation claims, this does not at all imply that confirmation of a particular hypothesis presupposes that that hypothesis itself already belongs to the background beliefs. On the contrary, if E only d-confirms H assuming background beliefs B, it means by definition that we have to presuppose B *in addition to* H in order to derive E. As far as B would presuppose, hence entail, H, it would be redundant as an extra.

and the grue problem, but can be defended against the objections normally raised against the purely classificatory approach of qualitative confirmation as deductive confirmation.

As already suggested, many philosophers who are reserved about the possibility of an acceptable theory of deductive confirmation subscribe to (a version of) the quantitative Bayesian theory of confirmation. As will be shown in the companion paper, that theory perfectly leaves room for HD-testing of hypotheses and a plausible specification of it implies the pure theory of confirmation presented in this paper.[9]

3. *Acceptance of hypotheses*

Finally, we will briefly deal with the problem of the acceptance of hypotheses. Explicating the idea of deductive confirmation of a hypothesis is one thing, explicating the idea of being sufficiently confirmed to be accepted is another. For the latter issue it is important to recognize that a hypothesis may be highly confirmed, without having become very plausible, since it may have been very *im*plausible at the start. For acceptance we need something like 'being sufficiently confirmed to have become sufficiently plausible to be accepted'. If it was already plausible at the start, the acquired confirmation may have been not very important for this purpose. Hence, crucial is 'being sufficiently plausible for being accepted'. Acceptance criteria may depend on the nature of the hypothesis: is it of an individual or a general nature? does it contain theoretical terms? etc. Moreover, they will depend

[9] In contrast to the above mentioned authors, Glymour (1980a) has, in response to the supposed objections to deductive confirmation, developed the so-called bootstrap method of hypothesis testing, which deviates in several respects from the HD-method, and he has based a new explication of confirmation on that method. The bootstrap method is primarily intended to solve the problem of the underdetermination of theory by data by an attempt to localize the support data provide for the separate hypotheses constituting a complex theory. As a consequence, the relativized clause "evidence E (bootstrap) confirms hypothesis H with respect to theory T" becomes the crucial statement to explicate. In view of the criticism of that explication by Christensen (1983), Glymour (1983) had to change his original proposal of (1980a). These revisions were criticized and revised by Zytkow (1986), leading to another revision by Earman and Glymour (1988). However, according to Christensen (1990) all these revisions are still defective, even so much that he concludes that the bootstrap method is irrelevant for the discrimination of relevant and irrelevant confirmations. Hence, in the light of all this the perspectives for bootstrap confirmation are at least uncertain and confused. The bootstrap method certainly has relevance for measuring theoretical terms and hence may indeed reduce the problem of the underdetermination of theory by data. However, as has been pointed out by Forge (1984), using the structuralist approach to measuring theoretical terms, the perspectives for localization of support by the bootstrapping method are nevertheless problematic.

on one's epistemological position, for there are, of course, various types of being accepted, roughly corresponding to the epistemological positions. To begin with the latter, for the realist acceptance of a hypothesis amounts to accepting the hypothesis as literally true, including its observational, referential and theoretical consequences. The referentialist will drop the theoretical consequences, and the empiricist, in addition, the referential consequences. Finally, the instrumentalist will drop the observational consequences concerning non-intended applications. In all cases, acceptance of a hypothesis means that it is to be added to the body of background beliefs, of which the general status, of course, also depends on the relevant epistemological position.

Within each of the above mentioned types of accepting a hypothesis as true we could also distinguish between at least four 'kinds of truth': true simpliciter, approximately true, near to the truth, and nearer to the truth than another hypothesis. The above, epistemologically induced, qualifications were primarily intended for the first kind of truth. The fourth kind is explicated elsewhere (Kuipers, 2000) for the various epistemological types in great detail. Although we use there informally expressions referring to the second and the third kind of truth, such as 'approximately true' and 'near to the truth' themselves, no precise explications of them are given. It is clear that 'approximately true' would need some conventional threshold for deviations from being true. Similarly, 'near to the truth' would need some threshold for being sufficiently near to the truth.

Returning to 'true simpliciter', and assuming that the hypothesis is of a general nature, with general test implications in different directions, and using theoretical terms, the three positions beyond the instrumentalist one (which we will further neglect as far as acceptance is concerned) have to make *inductive jumps* of various kinds, that is, from a threshold plausibility in one of the distinguished epistemological senses to the corresponding type of truth. The empiricist has to make *elementary or first order* inductive jumps, i.e., inductive generalizations (in observation terms), and *second order* inductive jumps, generalizations of inductive generalizations. (This distinction is not a very sharp one). The referentialist has to add *referential* inductive jumps, and the realist, in addition, *theoretical* ones. Hence, even if we neglect further refinements of referential and theoretical inductive jumps, there are at least four different questions of explication and justification regarding 'true simpliciter'.

To be sure, normally speaking, an inductive jump reflects a deductive fallacy. The recent AI-literature has made it clear that deductive fallacies may be very useful default rules of reasoning. The reader is referred to (Tan, 1992) and (Marek and Truszczynski, 1993). Here, however, we prefer to indicate another direction. Unfortunately, there only seem to be debatable conditions for elementary inductive jumps, and similarly for the other three types. Be this as it may, it turns out to be fruitful to consider the non-elementary jumps in the comparative perspective of empirical progress and truth approximation. In (Kuipers, 2000) it is argued that the conditions for accepting comparative success and truth approximation claims are essentially of the same non-elementary nature. Hence, the assessment of empirical progress and truth approximation claims is fundamentally of the same nature as the assessment of certain ‘truth simpliciter claims’.

Concluding remarks

As already indicated, in (Kuipers, 2000) several matters concerning hypothesis testing are dealt with. It becomes clear that the role of falsification and confirmation has to be relativized in several respects. To begin with, as is well-known, *prima facie* falsification may be disputed in several ways, e.g. by questioning the description of the counter-example or the truth of the auxiliary hypotheses needed to derive the relevant test implication. More fundamentally, it turns out that ‘being false’, and ‘being true’ for that matter, is from the point of view of empirical progress and truth approximation rather irrelevant, hence falsification will have to play a more modest role than frequently is assumed. We also show that the realist may even claim against the empiricist that one theory may be closer to the truth in the encompassing theoretical sense than another, even though the first has some counter-examples which are no counter-examples to the second.

Similarly, the role of confirmation can be relativized along the same lines and roughly at the same places. Since ‘confirmation’ has the connotation of not yet being falsified, that is, the hypothesis may still be true, and since it turns out to make perfectly sense to continue the ‘HD-evaluation’ of a theory, even though it has been falsified, the confirmation of a theory is not so important, but the more general notion of obtaining (general) successes is very important. *Prima facie* successes may be disputed in similar ways as *prima facie* falsification. Moreover, the obtainment of a success plays a modest role similar to that of a counter-example. However, now the realist cannot claim that a theory can be closer to the theoretical truth than another despite the fact that the other has one or more extra successes.

Although the role of confirmation and falsification can be strongly relativized, this does not mean that there is no need of a qualitative theory of deductive confirmation and falsification, as developed in this paper. On the contrary, the notion of confirmation and falsification remain of crucial importance for testing at least three types of hypotheses: 1) general test implications and similar general observational hypotheses, 2) comparative success hypotheses, and 3) truth approximation hypotheses. From the last point it may, however, not be concluded that there is a strong relation between confirmation and truth approximation. On the contrary, as mentioned before, it turns out that there is no direct link between ‘being true or false’ and ‘truth approximation’. This does not exclude that there is some sophisticated link between confirmation and truth approximation, but this is not explored by us.

Given that a qualitative theory of deductive confirmation remains highly relevant, it is also useful to have a satisfactory quantitative corresponding to it. In the companion paper such a quantitative, i.c., probabilistic Bayesian, theory of confirmation will be presented, with a decomposition in deductive and non-deductive confirmation. It is pure, and inclusive in the sense that it leaves room for the confirmation of hypotheses with probability zero. The resulting qualitative theory of (general) confirmation, including the one for deductive confirmation, will also be formulated.

University of Groningen
E-mail: T.A.F.Kuipers@Philos.rug.nl

REFERENCES

Christensen, D., (1983), “Glymour on evidential relevance”, *Philosophy of Science*, 50, 471–481.

Christensen, D., (1990), “The irrelevance of bootstrapping”, *Philosophy of Science*, 57, 644–662.

Cotton, P., (1990), “Is there still too much extrapolation from data on middle-aged white men?”, *JAMA (Journal of the American Medical Association)*, 263.8, 1049–1050.

Earman, J., (1992), *Bayes or bust. A critical examination of Bayesian confirmation theory*, MIT-press, Cambridge Ma.

Earman, J. & Glymour, C., (1988), “What revisions does bootstrap testing need?”, *Philosophy of Science*, 55, 260–264.

Flach, P., (1995), *Conjectures. An inquiry concerning the logic of induction*, ITK-series Tilburg University, Tilburg.

Forge, J., (1984), “Theoretical functions, theory and evidence”, *Philosophy of Science*, 51, 443–463.

Gemes, K., (1990), “Horwich and Hempel on hypothetico-deductivism”, *Philosophy of Science*, 56, 609–702.

Glymour, C., (1980a), *Theory and evidence*, Princeton University Press, Princeton.

Glymour, C., (1980b), “Hypothetico-deductivism is hopeless”, *Philosophy of Science*, 47, 322–325.

Glymour, C., (1983), "Revisions of bootstrap testing", *Philosophy of Science*, 50, 626–629.

Goodman, N., (1955), *Fact, fiction, and forecast*, Harvard UP, Cambridge Ma.

Grimes, T.R., (1990), "Truth, content, and the hypothetico-deductive method", *Philosophy of Science*, 57, 514–522.

Groot, A.D. de, (1961/1969), *Methodologie*, Mouton, Den Haag, 1961; translated as *Methodology*, Mouton, New York, 1969.

Hempel, C.G., (1945/1965), "Studies in the logic of confirmation", *Mind*, 54, 1945, 1–26/97–121. Reprinted in C.G. Hempel *Aspects of scientific explanation*, The Free Press, New York, 1965, 3–47.

Hempel, C.G., (1966), *Philosophy of natural science*, Prentice-Hall, Englewood Cliffs.

Horwich, P., (1983), "Explanations of irrelevance", in J. Earman (ed.), *Testing scientific theories*, University of Minnesota Press, Minneapolis, 55–65.

Howson, C. & Urbach, P., (1989), *Scientific reasoning: the Bayesian approach*, Open Court, La Salle.

Kuipers, T., (1997), "The Carnap-Hintikka programme in inductive logic", in: Matti Sintonen (ed.), *Knowledge and inquiry: Essays on Jaakko Hintikka's epistemology and philosophy of science*, Poznan Studies, Volume 51, Rodopi, Amsterdam, pp. 87–99.

Kuipers, T., (companion paper), "Quantitative confirmation, and its qualitative consequences", to appear in *Logique et Analyse*.

Kuipers, T., (2000), *From Instrumentalism to Constructive Realism*, Synthese Library, Vol. 287, Kluwer AP, Dordrecht.

Marek, W. & Truszczynski, M., (1993), *Nonmonotonic logic: context-dependent reasoning*, Springer, Berlin.

Popper, K.R., (1934/1959), *Logik der Forschung*, Vienna, 1934, translated as *The logic of scientific discovery*, Hutchinson, London, 1959.

Ray, W., Griffin, M. & Avorn, J., (1993), "Evaluating drugs after their approval for clinical use", *The New England Journal of Medicine*, 329.27, 2029–2032.

Sober, E., (1988), *Reconstructing the past: parsimony, evolution, and inference*, MIT-press, Cambridge Ma.

Sober, E., (1994), "No model, no inference: a Bayesian primer on the grue problem", in: Stalker, D. (ed.), *Grue! The new riddle of induction*, Open Court, La Salle, 225–240.

Sober, E., (1995), *Core questions in philosophy*, Prentice Hall, Englewood Cliffs.

Tan, Y.-H., (1992), *Non-monotonic reasoning: logical architecture and philosophical applications*, Free University, Amsterdam.

Zwart, S.D., (1995), "A hidden variable in the discussion about 'language dependency' of truthlikeness", in: *Cognitive patterns in science and common sense*, T. Kuipers and A.R. Mackor, (eds.), *Poznan Studies Vol. 45*, Amsterdam, Rodopi, 249–272.

Zytkow, J.M., (1986), "What revisions does bootstrapping need?", *Philosophy of Science*, 53, 101–109.