

Theo A. F. Kuipers

## The Hypothetico-Probabilistic (HP)-Method as a Concretization of the HD-Method

**Abstract.** As Ilkka Niiniluoto happens to have suggested already in 1973, analogous to the Hypothetico-Deductive (HD-) method of testing hypotheses, a Hypothetico-Probabilistic (HP-) method can be developed, with the HD-method as an extreme special case. It amounts to ‘deriving’ probabilistic test implications  $I$  from a hypothesis  $H$ , in the sense of (relatively) observational statements  $I$  satisfying ‘positive relevance’ for  $H$ , i.e.,  $p(I/H) > p(I)$ . When such probabilistic predictions come true,  $H$  is (said to be) probabilistically confirmed, otherwise  $H$  is probabilistically disconfirmed. In this paper I will first elaborate the HP-method of testing as a concretized version of the HD-method.

Elsewhere, I have argued that the comparative evaluation of two hypotheses, based on a repeated application of the HD-method, even when both have already been falsified, is functional for achieving empirical progress, with articulated perspectives on truth approximation. In this paper I argue along the same lines that the HP-method is functional for achieving (a concretized version of) empirical progress, at least as far as general probabilistic test implications of deterministic theories are concerned. Analysis of the perspectives on truth approximation has to be postponed to another occasion.

### 1. Introduction

By mid 1999 I completed a chain of argumentation (Kuipers 2000) for the claim that a sophisticated form of the Hypothetico-Deductive Method, the HD-method, was functional for achieving empirical progress and even truth approximation. The HD-method was put at work as a method of separate and comparative evaluation of hypotheses, rather than as merely a method of testing for their truth-value. In view of my earlier work on probabilistic confirmation, the natural question was whether this highly idealized HD-story could be ‘concretized’ to a HP-story, that is, a story in terms of something like a Hypothetico-Probabilistic Method, a HP-method. Since mid 2003 I have been working on it from time to time, with many problems, in particular with respect

to the final stages of the argument. However, up to describing HP-testing and separate and comparative HP-evaluation, the main lines were soon clear to me.

It was a great surprise to (re-)discover recently, just by seeing a reference in a manuscript of Ilkka Niiniluoto (*forthcoming*), that he has already introduced in the early 1970s the idea of a Hypothetico-Inductive Method, the HI-method. In fact, Niiniluoto and Tuomela (1973) is devoted to its crucial notion of a hypothetico-inductive inference. According to the *Preface*, the last chapter, entitled “Towards a non-inductivist logic of induction”, is written by Niiniluoto, apart from the 5<sup>th</sup>, and final, section ‘Conjectures’, being a joint piece of work. I quote from p. 209 (section 2): “Thus, one can develop a hypothetico-inductive method of testing hypotheses which is based, for example, upon some symmetrical explicate of inductive inference (such as positive relevance).” Here ‘positive relevance’ of evidence E for a hypothesis H refers to the probabilistic condition ‘ $p(H/E) > p(H)$ ’, which is almost<sup>1</sup> equivalent to ‘ $p(E/H) > p(E)$ ’, exemplifying the symmetry Niiniluoto is imposing for ‘inductive inference’ in general. In his (Niiniluoto 1999, p. 176) he briefly refers to that idea of 1973, and recently confirmed, in conversation, my impression on that basis that he had never elaborated the idea further. But the idea is very much present in 1973. I must have read that chapter in the late 1970s, at least diagonally, and I certainly must have read the brief reference in the 1999-book. But between mid 2003 and mid 2005 I took it to be my own idea to explicate, repeating the crucial phrases above, “a hypothetico-inductive method of testing hypotheses which is based, for example, upon (...) positive relevance”. That is, apart from my additional agenda about further evaluation, and except that I prefer to speak of a hypothetico-*probabilistic* method, for relatively marginal reasons that I will make clear soon. My crucial point of departure, explicating a probabilistic version of a test implication of a hypothesis along the lines of positive relevance, which occurred to me as a brain wave in 2003, may well be a pure case of cryptomnesia, that is, a case of unconscious plagiarism, I am afraid.

---

<sup>1</sup> That is, assuming  $p(E)$  and  $p(H)$  to be non-zero, for otherwise  $p(H/E)$  and  $p(E/H)$ , respectively, are not defined. That is, at least not by the standard definition of conditional probability, according to which, for example,  $p(E/H) = p(E \& H) / p(H)$ . Of course, there may be good reasons to add a specific definition in the case that  $p(H) = 0$ . In particular, when H logically entails E,  $p(E/H) =_{\text{df}} 1$  is highly plausible and will be assumed throughout.

I am nevertheless glad for the occasion to elaborate the method for testing and evaluation up to its functionality for achieving empirical progress in probabilistic terms. I still have to study its possible functionality for truth approximation more intensely, before that can be put on paper. But some provisional findings will be formulated. They lead to new challenges to whom it may concern and appeal, notably to myself and, I hope, to Niiniluoto.

After a methodological section (2) on concept explication by idealization and concretization, the basic concretization of a deductive consequence to a probabilistic consequence is introduced in section 3, followed by an intermezzo with some comparisons with related approaches and an impression of its formal properties in section 4. Section 5 presents the core transition of this paper, viz., from the HD-method of testing by deriving and checking deductive test implications to the HP-method of testing by deriving and checking probabilistic test implications, and section 6 the corresponding transition from deductive to probabilistic confirmation. After an intermezzo (section 7) about deductive- and probabilistic-nomological explanation and -prediction, some further challenges with respect to hypothesis evaluation are formulated in section 8. In section 9, HD- and HP-evaluation are presented as continuations of separate and comparative hypothesis evaluation even when HD- or HP-testing have resulted in their falsification. In section 10, I argue along similar lines as in the case of the HD-method that repeated application of the HP-method is functional for achieving empirical progress, at least as far as general probabilistic test implications of deterministic theories are concerned. In the concluding remarks (section 11) I note that it is more difficult to argue for the HP-method along the same lines as for the HD-method that empirical progress provides good reasons for having achieved truth approximation. Hence, presenting the additional perspectives of the HP-method on truth approximation relative to the HD-method, if they exist at all, has to be postponed to another occasion.

The title of Niiniluoto's 1973-chapter, recall "Towards a non-inductivist logic of induction", may sound strange in light of the title of the book of Niiniluoto and Tuomela, *Theoretical concepts and hypothetico-inductive inference*, and the suggestion (on p. 209) of the possibility of developing 'a hypothetico-inductive method of testing'. From section 1 ("Deductivism and Inductivism") it becomes already clear that he does not want to be an inductivist in any of the five senses of 'inductivism' he distinguishes, to be sure, nor a deductivist in any

of the three senses he describes. “Thus, the idea of hypothetico-inductive inference stands in contrast to both inductivism and deductivism” (p. 204). On p. 208 we find the crucial characterization of the type of inference: “Thus, [] the situations in which a hypothetical theory occurs among the premises of an inductive argument represent *hypothetico-inductive inference*. This notion is, therefore, an obvious generalization of the hypothetico-deductive inference.” Here an inductive argument can be any non-deductive argument, but it is clear that Niiniluoto prefers arguments in which ‘positive relevance’ is crucial. The main technical aim (and result) of the whole book is to “extend Hintikka’s system of inductive logic to apply to situations in which new concepts are introduced to the original language.” (*Preface*, p. IX). On p. 197 the authors say to do so in the spirit of Bunge: “Consequently, one can agree with Mario Bunge’s wish that a ‘non-inductivist logic of induction should be welcome’ (cf. Bunge 1963, p. 152). Our development of Hintikka’s system of inductive logic is a small but resolute step towards this goal.”

This brings me to my reason for speaking of the hypothetico-probabilistic method<sup>2</sup>, instead of the hypothetico-inductive method. As is well known Hintikka has, building upon Carnap’s continuum of inductive methods, developed in the 1960s an impressive two-dimensional continuum of inductive methods, taking, in contrast to Carnap, universal generalizations into account.<sup>3</sup> As I have elaborated elsewhere (Kuipers 2001, and *forthcoming a*), Hintikka’s probability function is inductive even in two ways. The same applies to the functions presented in Niiniluoto and Tuomela (1973). Though impressive as such ‘double-inductive’ probability functions are, in the present project I would also like to take non-inductive probability functions into account.

---

<sup>2</sup> Creating in this way another confusion. On p. 8 Niiniluoto and Tuomela state that they want to deal with “nomological, non-probabilistic inductive systematization of general statements”. But here ‘non-probabilistic’ refers to the deterministic (as I would say) nature of the nomological statements involved, not to the relation between such statements and observation statements. The restriction to deterministic hypotheses will become relevant in section 9, and it is inherent in my qualitative theory of truth approximation, and so in my attempts at proving that the HP-method serves that purpose.

<sup>3</sup> Later, Hintikka and Niiniluoto presented an alternative continuum for essentially the same purpose. I greatly enjoyed the study in the 1970s of, among other things, the relation between the old and the new system (Kuipers 1978).

Unfortunately, I was unable to completely (re)read the book of Niiniluoto and Tuomela, but my partial reading strongly suggests that I, and all other formally inclined philosophers of science, should reread this book. It contains many other useful ideas and technical results. But this may seem an *ad hoc* expression of my admiration for the work of Niiniluoto (and of Raimo Tuomela, to be sure). So let me close this introduction by quoting from my *Foreword* to Kuipers (2001), in order to make clear that my admiration for the work of Ilkka Niiniluoto is not something invented for the present occasion:

I like to mention Ilkka Niiniluoto's *Critical Scientific Realism* (1999) as, as far as I know, the most learned recent exposition of some of the main themes in the philosophy of science in the form of an advanced debate-book, that is, a critical exposition and assessment of the recent literature, including his own major contribution, viz. *Truthlikeness* of 1987. Despite our major differences regarding the topic of truth approximation, I like to express my affinity to, in particular, his rare type of constructive-critical attitude in the philosophy of science.

## **2. Concept explication by idealization and concretization**

I will present this paper to some extent explicitly in terms of 'idealization and concretization'. In my view, elaborated in (Kuipers, *forthcoming b*), this is not only an important methodology in the empirical science, but also in philosophy, at least as far as philosophy is engaged in 'concept explication'. In concept explication one aims at the construction of a simple, precise and useful concept, which is, in addition, similar to a given informal concept, used in everyday life, science or philosophy. According to the standard strategy of concept explication one tries to derive from the informal concept to be explicated and relevant empirical findings, if any, conditions of adequacy that the explicated concept will have to satisfy, and evident examples and counter-examples that the explicated concept has to include and exclude, respectively.

As in the empirical case, it may be very useful to start with an idealized way of catching cases and conditions, in order to make it gradually more realistic. However, the concretization should not only be more realistic than the idealized point of departure, the latter should also appear as an extreme special case of the former. This I will call here the C(oncretization)-test. In a way, concretization is just generalization and so a way of aiming at some continuity.

However, speaking of ‘explication by idealization and concretization’ highlights in typical cases the fact that 1) the informal usages suggest both versions of the relevant concept, 2) the idealized version is an *extreme* (special) case of the concretized version, which 3) frequently has no real life applications, that is, at most toy applications. Although the HD-method is nowadays not very popular among philosophers of science, I would not dare to say the last thing about the HD-method, on the contrary. In my view, both informal usages of ‘test implication’ belong to ‘the language of scientific common sense’, i.e., a definite and a more or less probable consequence, and, hence, instead of merely being a generalization of the HD-method, the HP-method should also be an explication of an informal notion, with the HD-method as an extreme special case.

It should be noted that in this respect the Bayesian method, as standardly conceived in terms of the prior and posterior probabilities of hypotheses in the face of evidence, is not very adequate. Although it certainly appeals to informal ways of speaking and although it can deal with deductive relations, it is, first, not very plausible to see it as an *extreme* special case of the HD-method. And, second, it is not very realistic as long as it puts all falsified hypotheses on the scrapheap of hypotheses with zero posterior probability.

### 3. The basic concretization: from d-consequences to p-consequences

The idealized starting point is that a statement S is a deductive (d-)consequence of the hypothesis H; formally:

$$(1) \quad H \models S \quad \text{S is a d-consequence of H.}$$

There are at least two plausible ways to concretize this in a probabilistic way. The one is by imposing the condition of ‘(weak) positive relevance’; formally:

$$(2) \quad p(S/H) \geq p(S) \quad \text{S is a p-consequence of H.}$$

Hence, S is made at least as likely by H as it is *a priori*. The other way requires that S is made at least as likely as its negation; formally:

$$(2^*) \quad p(S/H) \geq p(\neg S/H) \quad \text{S is a p*-consequence of H.}$$

which is equivalent to:  
 $p(S/H) \geq 1/2.$

Both definitions satisfy the necessary, but by no means sufficient condition for being a concretization. According to the C-test, the concretized version should reduce to the idealized version as an extreme special case. If S is a d-consequence of H in the sense of (1) we have, or impose, for whatever probability function p, that  $p(S/H) = 1$ , which can't be lower than  $p(S)$ , let alone than  $p(\neg S/H)$ , which is 0. Hence, a d-consequence is an extreme special case of a p- as well as of a p\*-consequence.

The choice between (2) and (2\*) is not an easy one, both have attractive and problematic features. Whereas (2) may be intuitively more appealing, (2\*) has the advantage of not needing to bother about unconditional probabilities. We will meet other (dis-)advantages in due course. However this may be, as we will see, as soon as we introduce the comparative perspective of hypothesis evaluation the choice becomes irrelevant.

From the methodological perspective of hypothesis evaluation, we should in fact have started with the notion of a conditional deductive (cd-consequence), for a prediction usually needs at least one 'initial condition' for its derivation of a hypothesis. Many more types of conditions may be involved, for example, background knowledge. But this makes no difference for the general idea. Hence, we introduce just a condition C in the following concretizations of d- and p- and p\*-consequences.

- (1C)  $H \& C \vdash S$  S is a cd-consequence of H, assuming C.
  - (2C)  $p(S/H \& C) \geq p(S/C)$  S is a cp-consequence of H, assuming C.
  - (2C\*)  $p(S/H \& C) \geq p(\neg S/H \& C)$  S is a cp\*-consequence of H, assuming C.
- which is equivalent to:  
 $p(S/H \& C) \geq 1/2$ .

Note that all three 'conditional concretizations' reduce to the unconditional versions by taking a tautology as an extreme special case of C.

From a purely formal perspective it is interesting to study the class of p- or p\*-consequences of a hypothesis, and to see which interesting properties of the class of d-consequences remain intact, which properties get lost, and which new properties become interesting. However, our concern is primarily hypothesis evaluation. In the next section, by way of an optional digression, we will only give some indications of these formal concerns and refer to a selection of the literature devoted to several other approaches to 'consequences from a probabilistic perspective'.

#### 4. Intermezzo. Some comparisons and formal properties

There is quite some literature on issues that are at least to some extent related to our notion of a p-consequence. One dominant issue is the question of the assertability or acceptability of a conditional in probabilistic terms. The proposals of Lewis (1976) and Jackson (1979) were shown by the latter (Jackson 1987) to amount to: ‘ $P \rightarrow Q$ ’ is assertable/acceptable if and only if  $p(Q/P)$  is high. Douven (*manuscript*), from whom I borrow this information, proposes to sharpen this by imposing an explicit threshold value (close but unequal to 1) for  $p(Q/P)$  and the condition of positive relevance,  $p(Q/P) > p(Q)$ . In a further refinement, he takes measures to escape from the lottery paradox on the basis of the background knowledge. Although ‘(weak) positive relevance’ is also crucial in my approach, a threshold in Douven’s style seems to me a rather arbitrary restriction when one ultimately aims at a general explication of a ‘probabilistic consequence’.

According to Adams (1966; 1975), the assertability/acceptability of ‘ $P \rightarrow Q$ ’ should just be measured by  $p(Q/P)$ , without some threshold. He adds a rather strong definition of ‘probabilistic validity’, viz., when, for all (rational!) probability functions, the uncertainty of the conclusion Q (i.e.,  $1-p(Q)$ ) does not exceed the sum of the uncertainties of the premises. Bradley and Swartz (1979, p. 196) define ‘Q follows probably from P’ if and only if most models of P are models of Q. In a note (7) to section 6 we will see when and how this definition fits in the definition of ‘p\*-consequence’, i.e., (2\*). Note, by the way, that the expression ‘follows probably from’ is also frequently used for informally characterizing an inductive argument in general.

Hailperin (1996) and Wagner (2004) have studied probabilistic versions of valid arguments, notably *modus ponens* and *modus tollens*. Assuming probabilities for the relevant premises leads to boundaries of the probability of the conclusion. Whereas Hailperin focused on the probabilities of a conditional premise, Wagner replaces this by conditional probabilities. For *modus ponens*, for example, the first and the second approach yield, respectively:

$$\begin{aligned} \text{If } p(Q \rightarrow P) = a \text{ and } p(Q) = b \text{ then } a+b-1 &\leq p(P) \leq a. \\ \text{If } p(P/Q) = a \text{ and } p(Q) = b \text{ then } ab &\leq p(P) \leq ab+1-b. \end{aligned}$$

For *modus tollens* the first approach is also easy, but not so the second. Interesting as this is as such, it does not seem to be directly relevant for our purposes.



Malinowski (*manuscript*) certainly is directly relevant for our purposes. Restricting himself to finite propositional languages, he defines ‘Q (probabilistically) supports P’ by the condition of weak positive relevance of Q for P,  $p(Q/P) \geq p(Q)$ , for all possible probability functions. A very strong extra condition indeed, but with some plausibility, for he interprets a probability function as a generalization of a valuation function. But he does not want to speak of ‘probabilistic entailment’ or, like Adams does for a related definition (see above), of ‘probabilistic validity’, because the defined notion is not truth preserving. He reports (Theorem 7) that Makinson (2005) has proved that the definition is equivalent to: “either P logically entails Q or vice versa”.

Let me first list some trivial properties of p-consequences (‘nd’ indicates that a property, or rule, does not hold for d-consequences):

(reflexivity)	P is p-consequence of P.
(or-elimination) <sup>nd</sup>	P is a p-consequence of $P \vee Q$ .
(and-introduction) <sup>nd</sup>	$P \& Q$ is a p-consequence of P.
(tautology)	if $p(Q) = 1$ , then Q is a p-consequence of any P.
(contradiction) <sup>nd</sup>	if $p(Q) = 0$ , then Q is a p-consequence of any P.

Following Kraus, Lehmann and Madigor (1990), Malinowski starts with listing (Theorem 5) Gentzen style rules of ‘weak positive relevance’, that is, of p-consequences:

(symmetry) <sup>nd</sup>	if Q is a p-consequence of P, then <i>vice versa</i> .
(left logical equivalence)	if P and Q are logically equivalent and R is a p-consequence of P then also of Q.
(right logical equivalence)	if P and Q are logically equivalent and P is a p-consequence of R then so is Q.
(classic logic)	if Q is a d-consequence of P then it is a p-consequence of P. (Our C-test condition!)
(contraposition)	if Q is a p-consequence of P then $\neg P$ is a p-consequence of $\neg Q$ .
(proof by cases)	if R is p-consequence of $P \& Q$ and of $P \& \neg Q$ then R is a p-consequence of P.

They are all easy to prove, even for any extension of a propositional language.

Malinowski continues by reporting several rules that the strong notion just can take over from ‘deductive entailment’, and which are easy to prove on the

basis of Theorem 7, see above, and shows that they fail to hold in general for an arbitrary probability function, that is, for p-consequences. I rephrase five of them as, partially related, ‘non-properties’ of p-consequences:

- (non-monotonicity) if R is a p-consequence of P, then not necessarily of P&Q.
- (non-and) if R is a p-consequence of P as well as of Q, then R need not be a p-consequence of P&Q.
- (non-transitivity) if Q is a p-consequence of P and R of Q, then R need not be a p-consequence of P.
- (non-equivalence) if Q is a p-consequence of P, and vice versa (trivially due to symmetry), and R is a p-consequence of P, then R need not be a p-consequence of Q.
- (non-right weakening) if  $P \rightarrow Q$  is a tautology and P a d-consequence of R, then Q need not be a p-consequence of R.

However, Malinowski also notes that three other ones have restricted validity for p-consequences (Theorem 11). I reformulate these properties in p-terms, where ‘r’ refers to their being restricted relative to the properties of the strong notion:

- (cut-r) if R is a p-consequence of P&Q and Q is a d-consequence of P then R is a p-consequence of P.
- (cautious monotonicity-r) if R is a p-consequence of P and Q is a d-consequence of P then R is a p-consequence of P&Q.
- (or-introduction-r) for incompatible P and Q, if R is a p-consequence of P as well as of Q then R is a p-consequence of  $P \vee Q$ .

These three rules can be transformed into equivalent versions by reversing (one or more of) the p-consequence relations (due to symmetry), e.g., (or-introduction-r) is equivalent to

for incompatible P and Q, if P and Q are both p-consequences of R then  $P \vee Q$  is p-consequence of R.

All proofs are easy, and it is again easy to check that they hold for any extension of the propositional language. One gets the corresponding unrestricted properties of the strong notion by 1) weakening in the first two cases the logical entailment premise to a p-consequence premise (i.e., weak positive relevance) and in the third case by skipping the incompatibility premise and 2) by adding to all resulting p-consequence claims ‘for all probability functions p’.

As suggested before, for our purposes, it would also be interesting to further study more specifically the formal properties of the class of p-consequences of a given hypothesis P,  $p\text{-Cn}(P) =_{\text{df}} \{Q \mid p(Q/P) \geq p(Q)\}$ , of course, being a superset of the class of its d-consequences,  $d\text{-Cn}(P) =_{\text{df}} \{Q \mid P \models Q\}$ . A crucial difference is that the latter is consistent, as soon as P is non-contradictory, whereas the former is, as a rule, inconsistent, as is easy to check.<sup>4</sup> For this and other reasons it may also be worthwhile to restrict the above classes to contingent hypotheses and consequences, and also to assume that they are (not weak but) positively relevant for each other and non-equivalent to each other, respectively.

## 5. From the HD-method of testing to the HP-method

The crucial idea of the HD-method of testing is the derivation of genuine test implications of a hypothesis. We define:

- (3) I is a d-test implication of H iff  
I is a non-tautological, observational d-consequence of H.

Hence, I must be phrased in ‘observational terms’ and it may not be the case that “ $\vdash I$ ” holds. Of course, observational terms are here understood in the sophisticated theory-relative sense, that is, I may not be laden by H, but may well be laden by other, ‘underlying’ theories.

The plausible concretizations of (3) are:

- (4) I is a p-test implication of H iff  
I is an observational p-consequence of H even such that  $p(I/H) > p(I)$ .

---

<sup>4</sup> For example, for a fair die, the hypothesis of an even outcome is positively relevant for all three mutually incompatible outcomes 4, 5 and 6.

- (4\*) I is a  $p^*$ -test implication of H iff  
 I is an observational  $p^*$ -consequence of H even such that  $p(I/H) > p(\neg I/H)$  (or, equivalently,  $p(I/H) > 1/2$ ).

Both have plausible ‘conditional concretizations’, where a condition only needs to be observational when it pertains to an ‘initial condition’.

For checking that (4) and (4\*) are concretizations of (3) in the minimal sense of the C-test, we have to assume that “ $H \models I$ ” and “not  $\models I$ ”. In this case, the condition in (4), viz.,  $1 = p(I/H) > p(I)$ , holds, except for an ‘almost tautological’ test implication, that is, a non-tautological test implication which has nevertheless probability 1. Although such implications do exist, they always have some infinite aspect, which we might want to exclude from being a genuine test implication. However, without further proviso, this would also exclude an observational universal generalization from being a test implication, which we would consider as too high a price. It is interesting to note that (4\*) does not have this problem, for “ $p(I/H) > 1/2$ ” straightforwardly holds for an almost tautological implication. Hence, we note a second technical advantage of (4\*). In sum, a d-test implication is (an extreme special case of) a  $p$ -/ $p^*$ -test implication, with a marginal exception for the unstarred case.

As a method of testing a hypothesis H, the HD-method can be said to aim at d(educative)-confirmation or falsification of H by checking d-test implications of H. That is, it amounts to deducing test implications and checking by (experiment followed by) observation whether they are (or become) true or false. If false, H has been falsified as well, if true, H has been (deductively) confirmed. Note that an HD-test never can lead to a neutral observation.

In light of ‘probabilistic testing’, the HD-method of testing is an idealized point of departure. Assuming that we can concretize (deductive) confirmation and falsification to some kind of ‘p(robabilistic)-confirmation and -falsification’,<sup>5</sup> or  $p^*$ -versions of them, it is plausible to submit that the Hypothetico-Probabilistic (HP-)method of testing aims at  $p$ -/ $p^*$ -confirmation or -falsification by checking  $p$ -/ $p^*$ -test implications. If this assumption is correct, we may conclude that the HP-method of testing is a concretization of the HD-

---

<sup>5</sup> The phrase ‘probabilistic disconfirmation’ is not only usually used instead of ‘probabilistic falsification’, it is also to be preferred for reasons which will be indicated below.

method of testing. That is, an explication of the title of this paper as far as testing is concerned.

## 6. From deductive to probabilistic confirmation

In Kuipers (2000, Part I) we have argued that there is a coherent landscape of confirmation notions, allowing different languages of (degrees of) confirmation. This has been presented more systematically in Kuipers (2001, Section 7.1.2, and *forthcoming a*). Here we just represent the main lines, but add the \*-versions.

Starting from the idealized notion of deductive (d-)confirmation of a hypothesis H by some evidence E:

$$(5) H \vdash E \quad E \text{ d-confirms } H,$$

the general form of the probabilistic concretization in line with that of a p-consequence focuses on the so-called likelihood of H in the light of E:  $p(E/H)$ , viz.,

$$(6) p(E/H) > p(E) \quad E \text{ p-confirms } H.$$

Although this deviates from the standard formulation,  $p(H/E) > p(H)$ , it is almost equivalent with the latter, viz., if both  $p(E)$  and  $p(H)$  are non-zero. The \*-version:

$$(6^*) p(E/H) > 1/2 \quad E \text{ p*confirms } H.$$

is rather unusual, but has some plausibility.

Analogously we get two concretizations of ‘falsification’, i.e., E (deductively) falsifies H iff

$$(7) H \vdash \neg E \quad E \text{ (d-)falsifies } H,$$

viz.,

$$(8) p(E/H) < p(E) \quad E \text{ p-disconfirms (p-falsifies)}^6 H,$$

---

<sup>6</sup> Speaking of ‘probabilistic falsification’ in case  $p(E/H) < p(E)$  may have some plausibility, but in view of the probabilistic ‘Confirmation Square’, building upon the ‘Deductive Confirmation Matrix’, it is much more plausible to speak of ‘probabilistic disconfirmation’ with deductive disconfirmation as extreme special case, i.e., the case that E is logically entailed by the negation of H. See Kuipers (2001, p. 22, p. 46).

(8\*)  $p(E/H) < 1/2$  E p\*-disconfirms (p\*-falsifies) H.

All these definitions have plausible conditional versions, straightforwardly in line with the definitions of conditional consequences.

Whatever definition of p-confirmation one prefers, they all presuppose that we choose a probability (p-)function. Now it is plausible to make a distinction between non-inductive and inductive p-functions. Inductive p-functions have some ampliative effect on the probability that “the future resembles the past”, to use Hume’s well known phrase. Assuming that the p-function can be decomposed over a partition of hypotheses and corresponding likelihoods, an ampliative effect can be generated in three ways, advocated by three famous philosophers, viz., by inductive likelihoods (Carnap), by inductive priors (Bayes), and by a combination of both (Hintikka). As a matter of fact most, if not all, probability functions presented in Niiniluoto and Tuomela (1973) are of this ‘double-inductive’ nature.

In order to get some more idea of the crucial types of inductive p-functions, that is, inductive likelihoods and inductive priors, we will specify the idea of non-inductive p-functions: whatever their precise form, they do not “learn from the past” by themselves. They nevertheless enable straightforward cases of confirmation, which I prefer to call ‘structural confirmation’, as opposed to ‘inductive confirmation’. So let me illustrate structural confirmation by my favorite example dealing with a fair die. Let E indicate the even (elementary) outcomes 2, 4, 6, and H the ‘high’ outcomes 4, 5, 6. Then (the evidence of) an even outcome confirms the hypothesis of a high outcome according to all three criteria (standard, (6), (6\*)), since  $p(E/H) = p(H/E) = 2/3 > 1/2 = p(H) = p(E)$ . By speaking of a fair die, we indicate that our p-function is an objective p - dealing with, or representing an objective probability process. However, by leaving out this information, and simply assuming a language of a family of 6 monadic predicates, with two defined disjunctive predicates, we would be able to generate the case of confirmation by consistently assuming the principle of indifference for the 6 possibilities. In this way we get an illustration of the ‘logical probability function’ or the ‘logical measure function’, introduced by Kemeny (1953), from now on indicated by ‘m’.

Structural confirmation is defined as confirmation on the basis of an objective probability function or on the basis of  $m$ .<sup>7</sup> Moreover, inductive confirmation is now defined as confirmation based on an inductive  $p$ -function, that is, restricted to monadic predicates, a  $p$ -function that has the property of ‘instantial confirmation’:

$$p(Fa/E \& Fb) > p(Fa/E) \quad \text{instantial confirmation,}$$

where ‘ $a$ ’ and ‘ $b$ ’ represent distinct individuals, ‘ $F$ ’ an arbitrary monadic property and ‘ $E$ ’ any kind of contingent evidence compatible with  $Fa$ . Note that this definition can be generalized to  $n$ -tuples and  $n$ -ary properties. As said, inductive (probability) functions can be obtained in two ways, which can now be specified, viz., by:

- ‘inductive priors’, i.e., positive prior  $p$ -values  $p(H)$  for  $m$ -zero hypotheses, and/or
- ‘inductive likelihoods’, i.e., likelihood functions  $p(E/H)$  having the property of instantial confirmation.

Note first that standard confirmation of  $m$ -zero hypotheses requires inductive priors, whereas confirmation of such hypotheses according to (6) and (6\*) is always possible, assuming that  $p(E/H)$  can be interpreted.<sup>8</sup>

Popper rejected both kinds of inductive confirmation, for roughly three reasons: two problematic ones and a defensible one. The first problematic one (Popper 1934/1959) is that he tried to argue, not convincingly (see e.g., Earman 1992; Howson and Urbach 1989; Kuipers 1978), that  $p(H)$  could not be positive for universal hypotheses. The second one is that any probability function has the property “ $p(E \rightarrow H/E) < p(E \rightarrow H)$ ” (Popper and Miller 1983). Although the claimed property is undisputed, the argument that a proper inductive probability function should have the reverse property, since ‘ $E \rightarrow H$ ’

---

<sup>7</sup> Note that if ‘most’ in the definition of Bradley and Swartz (1979, p. 196) (recall from section 4: ‘ $Q$  follows probably from  $P$ ’ if and only if most models of  $P$  are models of  $Q$ ) is interpreted as ‘more than not’, this definition amounts to (2\*), (4\*) and (6\*), i.e., the \*-definitions, specifically based on the logical probability function.

<sup>8</sup> Confirmation according to an inductive probability function is, as a rule, a mixture of structural and inductive confirmation. A precise definition of (degree of) inductive confirmation abstracts from the structural component. See (Kuipers 2001, *forthcoming* a).

is the ‘inductive conjunct’ in the equivalence “ $H \Leftrightarrow (E \vee H) \& (E \rightarrow H)$ ”, is not convincing. The indicated reverse property may well be conceived as an unlucky first attempt to explicate the core of (probabilistic) inductive intuitions, which should be replaced by the property of instancial confirmation.

The defensible reason is that the property of instancial confirmation merely reflects a subjective attitude and, usually, not an objective feature of the underlying probability process, if there is such a process at all.

However this may be, the logical measure function perfectly permits structural confirmation, in particular in the form of (6) or (6\*), for then the m-zero character of many interesting hypotheses need not prevent confirmation, assuming the likelihoods are well-defined. Hence, taking Popper’s reserves regarding inductive probabilities into account, and assuming that he would like the idea of a non-inductive probabilistic version of the hypothetical method, I prefer to speak of the HP-method, instead of speaking of the Hypothetico-Inductive (HI-) method, as Niiniluoto did in (1973) and (1999).

One argument against the very idea of speaking of confirmation, albeit structural confirmation’, in view of Popperian considerations, is that Popper did not like the term ‘confirmation’. This is for at least two reasons. First, ‘confirmation’ suggests judging the ‘future performance’ of a hypothesis, whereas hypotheses should be evaluated on the basis of their ‘past performance’. Second, aiming at confirmation, whether deductive or probabilistic, becomes an easy task of searching for probable confirmation. However, the first argument may apply to the standard definition of p-confirmation for p-non-zero hypotheses<sup>9</sup>, it does not apply to (6) and (6\*) for p-zero hypotheses, assuming the likelihoods are well-defined. Moreover, explicitly according to (4), (6) and (8), and implicitly according to (4\*), (6\*) and (8\*), the lower the prior probability of a test implication, the more one expects disconfirmation and hence the more surprising it is when confirmation follows. Hence, these definitions leave a lot of room for Popperian motives.

Which of the four kinds of p-functions to choose, is no fact of the matter. Hence, as soon as one uses the probability calculus, it does not matter very much which ‘confirmation language’ one chooses, for that calculus provides the crucial means for updating the likelihood of a hypothesis in the light of

---

<sup>9</sup> But even then it is a matter of probabilistic consistency, for then, assuming p-non-zero evidence, it is just equivalent to (6) or (6\*), which evaluate past performance.



evidence. Hence, the only important point which then remains is always to make it clear which confirmation language one has chosen.

In all cases, it is practical to have a degree of confirmation (or degree of success). For various reasons, explained in (Kuipers 2000), I prefer the ratio measure.

$$p(E/H)/p(E)$$

It is not only directly in line with the unstarred versions (4), (6) and (8), but it also explicitly accounts for the degree of surprise involved: the lower  $p(E)$  and the higher  $p(E/H)$ , the more this ratio exceeds 1, and the more happily surprised we will be with this performance of H. Conversely, the more the ratio is below 1, the more disappointed we will be with its performance. Since the ratio degree of confirmation is not so obvious starting from the starred versions (4\*), (6\*) and (8\*), I prefer the unstarred versions, and will assume them from now on.

Anticipating later sections, note already now that as soon as we come to compare the performance of two hypotheses, H1 and H2, relative to the same evidence, it becomes highly plausible to compare the respective likelihoods and the plausible comparison measure becomes the so-called 'likelihood ratio',  $p(E/H2)/p(E/H1)$ , in which a role for  $p(E)$  disappears. Hence, by the way, the comparative perspective relativizes all reasons to prefer the starred or the unstarred versions.

## **7. Intermezzo. Deductive- and Probabilistic-Nomological explanation and -prediction**

Not only probabilistic confirmation of hypotheses by evidence has been studied intensely. Also several forms of probabilistic explanation and prediction as concretization of their deductive forms have been studied. Here we will just indicate the main points of Probabilistic-Nomological (PN-) explanation and -prediction as concretization of their Deductive-Nomological forms. They nicely fit in the basic probabilistic concretization of a deductive consequence. The basic idea of DN-explanation of some evidence E by a well established nomological statement N, assuming some further well established conditions C, is

$N \& C \vDash E$ , and not  $C \vDash E$ ,

with the probabilistic form, PN-explanation, satisfying the C-test:

$$p(E/N \& C) > p(E/C).$$

Moreover, it is now plausible to say that, assuming C, N2 explains E better than N1 iff

$$p(E/N2 \& C) > p(E/N1 \& C).$$

Note that this comparative claim has no straightforward deductive version, except when one brings non-empirical differences between N1 and N2 into account.

As a matter of fact, the whole idea of DN- and PN-*prediction* is even more analogous to deriving HD- and HP-test implications. The only difference is that the implications no longer have the status of *test* implications, for, as in the case of DN- and PN-explanation, N as well as C are again supposed to be well established. Hence, to obtain DN- and PN-prediction, 'E' can simply be re-interpreted as 'predicted evidence' in the first and the second formula above, respectively.

Analogous to the case of explanation, we now also have a comparative version of saying that, assuming C, E2 is a more surprising, or a more risky, prediction of N than E1 iff

$$p(E2/C) < p(E1/C) \quad \text{and} \quad p(E2/N \& C) \geq p(E1/N \& C)$$

Here there is a deductive version, assuming that E1 and E2 are logically entailed by N&C and that there is some qualitative way of arguing that E2 is less plausible than E1. Of course, speaking of 'more risky' may suggest that we are back at the topic of hypothesis testing and evaluation, to which we return now.

## 8. Further challenges

In Kuipers (2000) we have given a coherent explication of the HD-method, not only as a method of testing hypotheses, in search for an answer to 'the truth question', but also as a method of separate and comparative empirical success evaluation, and, finally, as a method of evaluating truth approximation claims. Whereas testing becomes superfluous after identifying a (convincing) counter-

example, i.e., after falsifying a hypothesis, the indicated types of evaluation of hypotheses remain highly meaningful.

The general challenge of this project may now be described as the task of giving a coherent probabilistic concretization of the HD-method for testing and for the various types of evaluation. So far we have described how the first task can be fulfilled, heavily leaning on (Kuipers 2000 and 2001). We will now describe for the first time how the probabilistic concretization of separate and comparative success evaluation naturally emerges from their deductive point of departure.

## **9. HD- and HP-testing and -evaluation**

Following the HD-method, the result of checking a deductive test implication  $I$  of  $H$  is either a case of d-confirmation or a case of falsification. Since ‘confirmation’ has the connotation of ‘not yet falsified’, we jump for describing the results of continued (separate) evaluation of  $H$  by checking new deductive test implications to the terminology of positive and negative evaluation d-results. This naturally leads to an ‘evaluation report’ of  $H$ , listing the positive d-results at one side and the negative d-results at the other.

Analogously we have that the result of checking a probabilistic test implication  $I$  of  $H$  is either a case of p-confirmation or p-disconfirmation. And again we may better describe such results for the general evaluation of  $H$  as positive and negative p-results, leading to an evaluation report of  $H$  listing these results.

Note that the terminology of positive and negative (d- and p-)results is a variant of Larry Laudan’s instrumentalist terminology of (empirical) successes and problems.

For the comparative evaluation of two hypotheses  $H_1$  and  $H_2$  we of course compare the two evaluation reports. In the deductive case we now have to take into account that a deductive test implication of one hypothesis, nor its negation, need to be a test implication of the other, in which case the result of the test becomes a ‘neutral d-result’ for the latter. Hence, we get in total 9 possible combinations of being a positive, negative or neutral d-result. Of course, we say that a test result favors  $H_2$  over  $H_1$  if it is a positive d-result of  $H_2$  and a neutral or even negative d-result for  $H_1$  or if it is a neutral d-result for  $H_2$  and a negative one for  $H_1$ . A test result is indifferent between  $H_1$  and  $H_2$

when it is of the same kind for both. Note that a result neutral for both can only come into play due to a third hypothesis.

For the probabilistic concretization it is highly plausible to generalize the definition of a test result  $E$  favoring  $H_2$  over  $H_1$  by comparing the likelihoods, as follows,

$$E \text{ favors } H_2 \text{ over } H_1 \text{ iff } p(E/H_2) > p(E/H_1),$$

that is, if and only if the likelihood ratio of  $H_2$  relative to  $H_1$  exceeds 1. It is easy to check that this definition covers the three cases constituting the deductive definition of “ $E$  favors  $H_2$  over  $H_1$ ” as extreme special cases. Hence, the definition can be seen as a probabilistic concretization.

A further differentiation is plausible. So far we included in our definition of, for example, a positive  $p$ -result a positive  $d$ -result as an extreme special case. We may speak of a proper positive  $p$ -result, or a positive  $pp$ -result, when it is a positive  $p$ -result but not a positive  $d$ -result. Note that this classifies as positive  $pp$ -result ‘almost positive  $d$ -results’, that is, cases where  $p(E/H) = 1$ , but  $H$  does not logically entail  $E$ . A similar differentiation for negative  $pp$ -results can be made. Finally, a neutral  $p$ -result is defined by  $p(E/H) = p(E)$ , and hence this is a very special case of a neutral  $d$ -result, that is, neither  $E$  nor  $\neg E$  is entailed by  $H$ . Of course, a neutral  $d$ -result is either a positive or a negative  $pp$ -result, or a neutral  $p$ -result.

In total this generates 25 possible combinations of comparative results. Table 1 gives a survey of the basic types of the results of HD- and HP-testing and -evaluation.

Test implications	Test results	Evaluation results	
		separate	comparative
Deductive $H \vdash I$	$d$ -confirmation or falsification	positive or negative $d$ -results	$d$ -results favoring $H_2$ over $H_1$ , or $H_1$ over $H_2$ , or being indifferent (9 possible kinds)
Probabilistic $p(I/H) \geq p(I)$	$p$ -confirmation or $p$ -disconfirmation	positive or negative $p$ -results	$p$ -results favoring $H_2$ over $H_1$ , or $H_1$ over $H_2$ , or being indifferent (25 possible kinds)

Table 1. Basic types of results of HD- and HP-testing and -evaluation.

Table 2 differentiates the comparative results into the indicated 9 and 25 possible combinatory results for HD- and HP-comparative evaluation. In line with (Kuipers 2000, p. 117; 2001, p. 235), in which I called the table restricted to deductive results the ‘deductive evaluation matrix’, I like to call the probabilistic extension, the probabilistic evaluation matrix.<sup>10</sup> In Table 2 I have indicated the 9 ‘deductive (elementary or unified) boxes’ DB1-9 that result by interpreting p-results as d-results. Four DB’s just coincide with the ‘probabilistic boxes’ (PB’s) on the four corners (DB1, 4, 6, 9), four DB’s are the union of three PB’s (DB2, 3, 7, 8) and one is the disjunction of the remaining 9 PB’s (DB5). Hence, the deductive version of Table 2 is a coarse version of its probabilistic version, with the consequence that the deductive second row of Table 1 is a coarse version of the probabilistic third row. This shows that the C-test of both tables amounts to the conclusion that HP-testing and -evaluation form a concretization of their HD-variants in the form of a refinement.

		H1				
		negative d-result	negative pp-result	neutral p-result	positive pp-result	positive d-result
H2	negative d-result	DB4 0	DB2 –	DB2 –	DB2 –	DB1 –
	negative pp-result	DB8 +	DB5 0	DB5 –	DB5 –	DB3 –
	neutral p-result	DB8 +	DB5 +	DB5 0	DB5 –	DB3 –
	positive pp-result	DB8 +	DB5 +	DB5 +	DB5 0	DB3 –
	positive d-result	DB9 +	DB7 +	DB7 +	DB7 +	DB6 0

Table 2. The deductive and probabilistic evaluation matrix.

It is also easy to check that for HP- and HD-notions alike, separate evaluation is a concretization of testing, and comparative evaluation is a concretization of separate evaluation. For the first C-test, assume the extreme (!)

---

<sup>10</sup> I also introduced a quantitative version (l.c. p. 119; p. 237, respectively). From the present point of view, just inserting the likelihood ratios seems at first sight the most plausible thing to do, but I did not investigate this in detail.

special case that H has not yet been falsified. For the second C-test, assume that H1 is a tautology.

## 10. Empirical progress

The ultimate aim of comparative HD-evaluation in terms of positive and negative d-results is to come to conclusions about whether one hypothesis is in the light of the obtained empirical evidence really better than the other. The main line of reasoning developed in Kuipers (2000, Part II, reproduced in 2001, Part IV) is as follows. First we need a definition of “more successfulness”, which amounts in the present formulation to:

More Successful (MSF)

H2 is more successful than H1 in light of the available evidence iff

- (1) all positive d-results of H1 are positive d-results of H2
- (2) all negative d-results of H2 are negative d-results of H1
- (3) H2 has some more positive d-results or some fewer negative d-results than H1.

Note that this definition does not refer to neutral results. Assuming three possible results, they are handled implicitly. For one might expect, for example, “all neutral d-results of H1 are neutral or positive d-results of H2”. But a neutral d-result of H1 being a negative d-result of H2 is excluded by clause (2). Note also that MSF is rather strict; more liberal versions are easy to imagine, but for our purposes we want to focus on clear-cut cases. They are sufficient to tell the principal story.

Of course, H2 may be more successful than H1 by sheer luck or prejudice regarding the choice of test implications so far. Hence, this situation only suggests the

Comparative Success Hypothesis (CSH)

H2 (is and) remains more successful than H1.

CSH is an interesting hypothesis, even if H2 has already been falsified (by negative d-results). It can be tested by a comparative application of the HD-method of testing: regarding clause (1), derive and test potential positive d-results of H1 that cannot be derived from H2; and similarly for clause (2).

Only after various tests of CSH may we come to the conclusion, for the time being, that H2 is really better than H1. We like to state this in the form of the

Instrumentalist Rule of Success (IRS)

When H2 has so far proven to be more successful than H1, i.e., when CSH has been ‘sufficiently confirmed’ to be accepted as true, discard H1 in favor of H2, at least for the time being.

Of course, the phrase ‘sufficiently confirmed’ hides a lot of problems dealing with inductive generalization. However this may be, the acceptance of CSH and the consequent application of IRS I would claim to be the core idea of ‘empirical progress’, a new hypothesis that is better than an old one. IRS may even be considered the (fallible) criterion and hallmark of scientific rationality, acceptable for the empiricist as well as for the realist. Of course, IRS is called instrumentalistic because it remains to take hypotheses seriously when they have been falsified. By IRS such a hypothesis is only discarded or eliminated, for the time being, when there is a really better, but perhaps also already falsified, hypothesis.

Finally, it is possible to show (Kuipers 2000, Part III and IV) that IRS is functional (or instrumental) for truth approximation, in precise senses depending on the particular type of truth approximation (basic or refined, and observational, referential, or theoretical). For the moment, however, I restrict myself to empirical progress by the HD-method as specified along the lines of d-MFS, d-CSH and d-IRS, where I introduce the prefix ‘d-’, to distinguish it from their possible p-versions. The question is whether this story can be concretized in probabilistic terms.

The crucial point of departure is the concretization of d-MFS. Of course, we just might replace ‘d-results’ by ‘p-results’, only guaranteeing that the scores (positive, neutral, negative) in the report of separate evaluation for H2 are never ‘lower’ than for H1 and sometimes ‘higher’. We may even differentiate this in terms of (positive and negative) d- and proper p-results (pp-results). This would only amount to a translation of the deductive story into probabilistic terms. A genuine form of more successfulness in probabilistic terms is, of course, in terms of ‘p-results favoring one hypothesis over another’. Recall:

E favors H2 over H1 iff  $p(E/H2) > p(E/H1)$ .

This suggests:

Probabilistically More Successful (p-MSF)

H2 is probabilistically more successful than H1 in light of the available evidence iff

- (1) no p-results favor H1 over H2
- (2) some p-results favor H2 over H1

Note that p-MSF confronts each p-result with H1 and H2, independently of the other results. This could of course be done otherwise, notably with successive concatenation. This would not only be more complicated but would also lead to a less strict version of ‘more successfulness’, though perhaps a no less interesting one, in particular when one favors an inductive rather than a structural probability function.

Before we apply the C-test, to check whether d-MSF is indeed an extreme special case of p-MSF, we will discuss the fact that the present form of p-MSF may be criticized for being not very realistic. Not in the sense of being too strict to have only few, if any, real life applications, that is, the same reason why d-MSF can also be said to be very strict. In both cases this is what we may expect for a genuinely superior hypothesis H2 relative to H1, that is, being an exception rather than a rule in actual science.

But even if H2 is superior to H1 in any intuitive sense, there is something strange to the definition of p-MSF. For even in this case, a particular p-result may well be the negation of the result ‘predicted’ by a p-test implication of H2, providing a higher likelihood of H1, just by bad luck. This is particularly the case for so-called ‘individual test implication’, that is, test implications that probabilistically predict an outcome of an individual experiment in the face of the hypothesis and some initial conditions. However, if we look at ‘general test implications’,<sup>11</sup> the definition of p-MSF becomes more plausible. Suppose that for some general statement I holds  $p(I/H2) > p(I)$  and  $p(I/H2) > p(I/H1)$ . Hence it is a more probable p-test implication of H2 than of H1, if it is a p-test implication of H1 at all, that is, if  $p(I/H1) > p(I)$ , which need not be the case. If such an I becomes established as evidence, no sheer luck can be involved and it should of course be counted as a p-success favoring H2 over H1, as

---

<sup>11</sup> For details about the distinction between individual and general test implications, see Kuipers (2000; 2001).



prescribed by clause (2). On the other hand, if H1 scores such a success relative to H2, clause (1) is rightly violated, for H2 apparently is not overall superior to H1. Hence we conclude that the definition of p-MSF is adequate, provided we restrict the attention to general test implications, which we will indicate by pg-MSF, etc.

Let us now turn to the C-test. In Kuipers (2000, Section 5.1.5) we have argued that there are in fact three interesting models of HD-evaluation, viz. in terms of ‘individual successes’ and ‘individual problems (counterexamples)’, or in terms of ‘general successes’ and ‘individual problems’, or in terms of ‘general successes’ and ‘general problems (falsifying general facts)’. The definition of d-MSF can then be given in one of these combinations. The second, ‘asymmetric’ model follows most naturally when we focus HD-evaluation on general test implications, either leading to general successes or to individual problems. In this model,<sup>12</sup> d-MSf reads:

dg-More Successful (dg-MSF)

H2 is more successful than H1 in light of the available evidence iff

- (1) all general successes of H1 are general successes of H2
- (2) all individual problems of H2 are individual problems of H1
- (3) H2 has some extra general successes or some fewer individual problems than H1.

For the C-test of dg-MSF relative to pg-MSF we now first assume that we have a general success of H1, that is, some (general) I logically entailed by H1, and established as correct. According to dg-MSF-(1), I is also entailed by H2. In probabilistic terms it follows that  $p(I/H1) = 1$  and if I is established  $p(I/H2)$  should, according to pg-MSF-(1), not be lower than  $p(I/H1)$ , hence it also equals 1. Hence, pg-MSF-(1) covers dg-MSF-(1) as an extreme special case. Let us now, second, assume that we have an individual problem of H2, that is, some (general) I logically entailed by H2 was established as incorrect, by a counterexample of I and hence of H2. According to dg-MSF-(2) it should also be a counterexample to H1. In probabilistic terms we have that  $p(E/H2) = 0$ .

---

<sup>12</sup> As suggested, the first, symmetric model makes no sense in probabilistic terms, due to the possibility of bad luck. Of course, also the possibility of sheer luck interferes. However, the third, also symmetric model may well make sense in probabilistic terms, but I did not yet investigate this in enough detail.

Again according to pg-MSF-(1),  $p(E/H1)$  may not be higher than 0, hence  $p(E/H1) = 0$  (and E amounts to a counterexample of H1 in probabilistic terms). Hence, pg-MSF-(1) also covers dg-MSF-(2) as an extreme special case. By two similar arguments it is easy to show that pg-MSF-(2) covers both alternatives in dg-MSF-(3). Q.e.d.

Having restricted the notion of being probabilistically more successful to general test implications, the rest of the concretization follows easily. The probabilistic versions of the Comparative Success Hypothesis (pg-CSH) and the Instrumentalist Rule of Success (pg-IRS) result from simply substituting pg-MSF for ‘more successful’, instead of dg-MSF. Finally, again the acceptance of pg-CSH and consequent application of pg-IRS is the core idea of ‘empirical progress’ in probabilistic terms, a new hypothesis that is better than an old one, now also taking probabilistic general test implications into account.

It is important to note that the expression ‘probabilistic general test implications’ should not be taken in the sense of an implication being probabilistic in nature. In that expression, ‘probabilistic’ only refers to the non-deductive nature of the relation between the relevant hypothesis and the implication. Hence, the restriction to general test implications of the HP-method is a restriction to deterministic, non-probabilistic hypotheses. For the HD-method this restriction is even presupposed for individual test implications. However, I would not exclude the possibility to extend the whole analysis of the HD- and the HP-method to probabilistic hypotheses.

## **11. Concluding remarks**

We may conclude that there is an HP-method of testing and separate and comparative evaluation, which is functional for empirical progress in probabilistic terms and which is a straightforward concretization of the HD-method. The relevant probability function and corresponding ‘confirmation language’ may be of a Popperian, Carnapian, Bayesian or Hintikkian nature. DN-explanation and -prediction can similarly be concretized to PN-explanation and -prediction. All relevant C-tests are successfully passed, where 1 and 0 represent the extreme special cases of  $p(I/H)$  or  $p(E/H)$ . Recall also that for HP- and HD-notions alike, separate evaluation is a concretization of testing, and comparative evaluation is a concretization of separate evaluation.

In sum, a coherent story can be told up to and including the functionality of the HD- and the HP-method for achieving empirical progress in deductive and probabilistic terms, respectively. The remaining challenge is the question whether the HP-method is, like the HD-method, functional for truth approximation.

In my *From Instrumentalism to Constructive Realism. On some relations between confirmation, empirical progress and truth approximation*, to quote the title of Kuipers (2000) in full, I have argued that the comparative evaluation of two hypotheses, based on repeated application of the HD-method, to be continued if both have already been falsified, not only results in a plausible explication of the notion of empirical progress, but also leads to articulated perspectives on (observational, referential and theoretical) truth approximation.

Hence, it is plausible to try to argue along similar lines that the comparative evaluation of theories based on the repeated application of the HP-method, also results in positive perspectives on truth approximation. However, although the explication part of the project has been successfully completed, that is, the HP-method also suggests a probabilistic concretization of the (qualitative, ‘deductive’) explication of ‘truth approximation’, so far the rest of the project only turned into restricted but very surprising grist to the mill of dogmatic Popperians, or so it seems. For, as far as equally probable theories are concerned, it seems provable that the HP-method is, relative to the HD-method, redundant: empirical progress and truth approximation by the HD-method, regarding deterministic hypotheses, seem to entail these respective objectives of the HP-notions in probabilistic terms, hence including an important relativization of the analysis of probabilistic empirical progress in the last section.

Whether or not this project can nevertheless be finished successfully in both cases it is also a challenge for Ilkka Niiniluoto to investigate to what extent his quantitative theory of truth approximation can be adapted to, the above or his own future explication of, the HP-method such that the latter is functional for quantitative probabilistic empirical progress and truth approximation, at least as far as deterministic theories are concerned.

## **12. Acknowledgements**

I would like to thank for all the comments I got after presenting the above ideas in Bielefeld (2003), Rotterdam (2004), Konstanz (2004), and in the PCCP

research group in Groningen (2004). In particular I want to thank Igor Douven, Dale Jacquette, Jacek Malinowski and Elliott Sober for their detailed comments in various stages, and Malinowski also for even writing a useful paper on the formal properties of (a strong version of) probabilistic consequence. I like to thank the Netherlands Institute of Advanced Study (NIAS) in Wassenaar for the privilege of allowing me to return for two weeks to this paradise for thinking and writing. Finally, I like to thank Ilkka Niiniluoto for his stimulating friendship, starting in June 1974 in Warsaw on a conference on Formal Methods in the Methodology of Empirical Sciences. I hope he will forgive me my, not unlikely, unconscious plagiarism with respect to the basic idea of this paper.

## References

- Adams, E. W. 1966: "Probability and the Logic of Conditionals", in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*, North-Holland, Amsterdam, pp. 265–316.
- Adams, E. W. 1975: *The Logic of Conditionals*, Reidel, Dordrecht.
- Bradley, R. and N. Swartz 1979: *Possible Worlds*, Basil Blackwell, Oxford.
- Bunge, M. 1963: *The Myth of Simplicity*, Prentice Hall, Englewood Cliffs.
- Douven, I. *manuscript*: *The Evidential Support Theory of Conditionals*.
- Earman, J. 1992: *Bayes or Bust. A Critical Examination of Bayesian Confirmation Theory*, MIT-press, Cambridge.
- Hailperin, T. 1996: *Sentential Probability Logic*, Lehigh University Press, Bethlehem.
- Howson, C. and P. Urbach 1989: *Scientific Reasoning: the Bayesian Approach*, Open Court, La Salle.
- Jackson, F. 1979: "On Assertion and Indicative Conditionals", *Philosophical Review* 88, 565–589.
- Jackson, F. 1987: *Conditionals*, Blackwell, Oxford.
- Kemeny, J. 1953: "A Logical Measure Function", *The Journal of Symbolic Logic* 18.4, 289–308.
- Kraus, S., D. Lehmann and M. Magidor 1990: "Nonmonotonic Reasoning, Preferential Models and Cumulative Logics", *Artificial Intelligence* 44, 167–207.
- Kuipers, T. 1978: *Studies in Inductive Probability and Rational Expectation*, Synthese Library 123, Reidel, Dordrecht.
- Kuipers, T. 2000: *From Instrumentalism to Constructive Realism*, Synthese Library 287, Kluwer Academic Publishers, Dordrecht.
- Kuipers, T. 2001: *Structures in Science*, Synthese Library 301, Kluwer Academic Publishers, Dordrecht.
- Kuipers, T. *forthcoming* a: "Inductive Aspects of Confirmation, Information, and Content", forthcoming in R. E. Auxier and L. E. Hahn (eds.), *The Philosophy of Jaakko Hintikka*, Library of Living Philosophers, Vol. 30, Open Court, LaSalle.

- Kuipers, T. *forthcoming* b: “Empirical and Conceptual Idealization and Concretization. The Case of Truth Approximation”, *forthcoming* in (eds.), *Liber Amicorum* for Leszek Nowak, Rodopi, Amsterdam.
- Lewis, D. K. 1976: “Probabilities of Conditionals and Conditional Probabilities”, *Philosophical Review* 85, 297–315.
- Makinson, D. 2005: *Bridges from Classical to Nonmonotonic Logic, Texts in Computing*, Kings College, London.
- Malinowski, J. *manuscript*, Bayesian Propositional Logic.
- Niiniluoto, I. 1987: *Truthlikeness*, Reidel, Dordrecht.
- Niiniluoto, I. 1999: *Critical Scientific Realism*, Oxford University Press, Oxford.
- Niiniluoto, I. and R. Tuomela 1973: *Theoretical Concepts and Hypothetico-Inductive Inference*, Synthese Library, Vol. 53, Reidel, Dordrecht.
- Niiniluoto, I. *forthcoming*: “Evaluation of Theories”, *forthcoming* in T. Kuipers (ed.), *Handbook of the Philosophy of Science*, Vol. 1, Focal Issues, Elsevier, Amsterdam.
- Popper, K. 1934/1959: *Logik der Forschung*, Vienna, 1934; translated as *The Logic of Scientific Discovery*, Hutchinson, London, 1959.
- Popper, K. and D. Miller 1983: “A Proof of the Impossibility of Inductive Probability”, *Nature* 302, 687–688.
- Wagner, C. G. 2004: “*Modus Tollens* Probabilized”, *The British Journal for the Philosophy of Science* 55, 747–753.

