IGOR DOUVEN AND WOUTER MEIJS

BOOTSTRAP CONFIRMATION MADE QUANTITATIVE

ABSTRACT. Glymour's theory of bootstrap confirmation is a purely qualitative account of confirmation; it allows us to say that the evidence confirms a given theory, but not that it confirms the theory to a certain degree. The present paper extends Glymour's theory to a quantitative account and investigates the resulting theory in some detail. It also considers the question how bootstrap confirmation relates to justification.

1. INTRODUCTION

It is widely acknowledged that empirical testing crucially involves the use of auxiliary theories and is thus in an uncontroversial sense always relative to some theory or theories. In his (1980a), Glymour presented an important, new confirmation theory according to which empirical theories can be confirmed in an absolute sense nonetheless. Glymour's theory of bootstrap confirmation – as he called it – is a purely qualitative confirmation theory; it allows us to say that the evidence confirms a given theory, but not that it confirms the theory to a certain degree. In the present paper we aim to take some first steps towards extending Glymour's theory to a quantitative account.

The plan of the paper is as follows. We start by outlining Glymour's theory of bootstrap confirmation (Section 2). We then in Section 3 formulate what we believe to be the main desiderata for a quantitative bootstrap theory. Guided by these desiderata, we define a measure of bootstrap confirmation and show how it can be put to use to formulate a quantitative version of Glymour's theory (Section 4). Finally, we consider the question how bootstrap confirmation relates to justification. In particular, we address a worry that bootstrap confirmation of some theory by the evidence does not really indicate *confirmation* of that theory (Section 5).

Synthese (2006) 149: 97–132
DOI 10.1007/s11229-004-6250-2

2. BOOTSTRAP CONFIRMATION

A confirmation theory is, roughly put, a theory that purports to specify, for any given evidence statement and any given hypothesis, whether or not the former *supports* the latter, or – in different terms that for present purposes can all be taken as equivalents – whether coming to know the evidence statement should *increase our confidence* in the hypothesis, whether the evidence *adds to the justificational status of the hypothesis,* whether it *gives reason to believe* the hypothesis. A confirmation theory may or may not also specify *to what extent* the evidence supports, or should affect our confidence in, a given hypothesis, *how much* it adds to the justificational status of a hypothesis, *how much* reason it gives to believe the hypothesis. If it does specify the extent of support, it is called a *quantitative* confirmation theory, if it does not, it is called a *qualitative* confirmation theory.

All these formulations suggest that confirmation is a two-place relation, viz., a relation between a body of evidence and a hypothesis. And indeed this is what philosophers for a long time believed. However, as Duhem (1906/1954) was the first to argue, and as Quine (1953) famously repeated in his assault on the logical empiricists' reductionist semantics, confirmation is a three- rather than a two-place relation: evidence generally accrues to a hypothesis only relative to one or more auxiliary hypotheses. It is no exaggeration to say that today this is something of a commonplace among analytic philosophers.[1]

While many have taken Duhem's thesis to imply that all confirmation must be 'relative' confirmation, according to Glymour (1980a), the indispensability of auxiliaries in the testing of *single* hypotheses is no impediment to absolute confirmation of *complexes* of such hypotheses or, as we may call them, theories. More specifically, Glymour presents a confirmation theory on which the piecemeal confirmation of the individual hypotheses comprised by a given theory relative to other hypotheses comprised by the same theory *may* add up to an unrelativized confirmation of that theory as a whole. Whether it does, depends on whether the separate tests of the various hypotheses *might* have turned out negative for these hypotheses. In a capsule formulation, Glymour's theory comes to the following definition:[2]

DEFINITION 2.1 (Bootstrap Confirmation). Let $T = \{H_1, \ldots, H_n\}$. Then evidence E bootstrap-confirms T exactly if $T \cup E \nvdash \bot$ and for each $i \in \{1, \ldots, n\}$ the following two conditions hold:

1. there is a $T' \subset T$ such that $H_i \notin T'$ and
   a. E confirms H_i with respect to T' ; and
   b. there is possible – but non-actual – evidence E' such that E' disconfirms H_i with respect to T';
2. there is no $T'' \subseteq T$ such that E disconfirms H_i with respect to T''.

We have eight comments on this definition:

(1) The reader should be warned that the definition reflects *our* understanding of Glymour's theory, and that Glymour's book leaves some room for interpretation.[3] Since our aim is not exegetical, we shall not argue for the correctness of our interpretation of Glymour's text. Moreover, we believe that, even if Definition 2.1 should fail to capture what Glymour 'really' had in mind, it defines a notion of confirmation that is well worth considering in its own right.

(2) As Glymour (1980a, 127; 1980b) emphasizes, bootstrapping is not to be thought of as being tied to a particular theory of (simple, non-bootstrap) confirmation. Accordingly, the terms 'confirms' and 'disconfirms' in the clauses of the definition can be cashed out in more than one way.[4] For example, they can be understood in terms of a Hempelian positive-instance account of confirmation (as Glymour does in his own presentation of bootstrapping), in hypothetico-deductive terms, or in probabilistic terms.

(3) Coupled with certain confirmation theories, in the face of subclause 1.a the second clause amounts to no more than the requirement that the theory at issue be consistent (and is thus redundant given the requirement that the theory be consistent with the evidence). Suppose for instance the notions of confirmation and disconfirmation are understood hypothetico-deductively. Without loss of generality, consider a theory consisting of three axioms, $T = \{H_1, H_2, H_3\}$, and suppose that E confirms H_1 with respect to H_2, and thus that (1) $\{H_1, H_2\} \vdash E$, but disconfirms H_1 with respect to H_3, and thus that (2) $\{H_1, H_3\} \vdash \neg E$. Then T must be inconsistent. For it follows from (1) that $\{\neg E\} \vdash \neg H_1 \vee \neg H_2$ and, similarly, it follows from (2) that $\{E\} \vdash \neg H_1 \vee \neg H_3$. And thus, by Constructive Dilemma, $\{E \vee \neg E\} \vdash (\neg H_1 \vee \neg H_2) \vee (\neg H_1 \vee \neg H_3)$, or, put differently, $\{E \vee \neg E\} \vdash \neg H_1 \vee \neg H_2 \vee \neg H_3$. From which it follows that $\vdash \neg(H_1 \wedge H_2 \wedge H_3)$, i.e., $\vdash \neg T$. But for those familiar with the common probabilistic understanding of confirmation and disconfirmation (see Section 3), it will be immediately clear that, if

some T and E satisfy clause 1 but not clause 2, that does not entail that T is inconsistent. So if the definition is to be neutral as regards theories of non-bootstrap confirmation, then the second clause cannot be dispensed with.

(4) Glymour originally allowed bootstrap testing in which evidence could confirm a hypothesis relative to itself – what later came to be called 'macho-bootstrapping'. However, under the pressure of criticism from, among others, Christensen (1983), Edidin (1983), and van Fraassen (1983a) he later restricted the auxiliaries admissible in a test to hypotheses other than the one under scrutiny in that test (Earman and Glymour 1988), as does clause 1 by requiring that the hypothesis under scrutiny not be in the set of hypotheses from which the auxiliaries in that hypothesis' test are taken. (Notice that a similar restriction in clause 2 would be superfluous. If a hypothesis is disconfirmed by the evidence, then this seems to be no less – but rather more – damaging to that hypothesis, and hence also to any theory that includes it, when the hypothesis served itself as an auxiliary in that test than when only other hypotheses did.)

(5) Bootstrap confirmation as presented here is defined for finitely axiomatizable theories only. It is not theoretically impossible to generalize Definition 2.1 to the infinite case, but it is hard to see how bootstrap confirmation of such theories could *practically* be achieved. Thus, in what follows, by 'theory' we shall mean finitely axiomatizable theories.[5]

(6) As in a bootstrap test of a theory T all the tests of the individual axioms of T rely on auxiliaries that also come from T, it might seem a questionable feature of this account that, provided both clauses of Definition 2.1 are satisfied, it allows us to conclude that the evidence confirms the theory, period, and not just that it confirms the theory with respect to itself (a conclusion – note – that would be barely significant unless one is already willing to accept the theory). More than questionable, in fact: a common response of those who first learn about bootstrap testing is to exclaim that the procedure is patently circular. After all – it is said – the very theory the truth of which is at stake in the test is presupposed in that test in the sense that it is allowed to supply the auxiliaries needed in the tests of the separate hypotheses comprised by the theory. At first sight, the situation may indeed seem analogous to one in which we (correctly) derive some proposition A using A itself as a premise and then present that as a proof of A (instead of just as a proof of $A \rightarrow A$, or

of $\{A\} \vdash A$). But it is not. Consider: if bootstrap testing were really circular, then how could any theory ever fail to be bootstrap-confirmed? And if the 'non-triviality subclause' 1.b is satisfied, a theory *can* fail to be bootstrap-confirmed. For what the subclause ensures is that adopting certain hypotheses as auxiliaries in testing some other hypothesis does not guard the latter against disconfirmation whatever the data. This, we believe, is as straightforward a way as any to see the non-circularity of bootstrap testing.[6]

(7) Definition 2.1 defines bootstrap confirmation. What about bootstrap disconfirmation? Glymour does not say, and it seems this notion can be defined in more than one plausible way. Like Glymour, we will mainly concern ourselves with bootstrap confirmation. Nevertheless, it will prove useful later on to have a definition of bootstrap disconfirmation at hand. As such we propose this:

DEFINITION 2.2 (Bootstrap Disconfirmation). Evidence E bootstrap-disconfirms theory $T = \{H_1, \ldots, H_n\}$ precisely if for at least one $i \in \{1, \ldots, n\}$ there is a $T' \subseteq T$ such that E disconfirms H_i with respect to T'.

(8) Our final comment has a heuristic intent. It may be helpful to think of Definition 2.1 as indicating some sort of coherence of the axioms of a theory both with one another and with the evidence. After all, a major intuition regarding coherence is that coherent propositions 'hang together' (cf. BonJour 1985, 93). And a positive bootstrap test from evidence E of a theory $T = \{H_1, \ldots, H_n\}$ is an indication of the hypotheses in T and E hanging together in a very clear sense: the hypotheses help each other to obtain support from the evidence. Of course the definition cannot quite be a definition of coherence. For a second generally held intuition regarding coherence is that coherence is a matter of degree – a set of propositions can hang together more or less tightly – and this cannot be expressed by a qualitative theory of bootstrap confirmation, which permits categorical judgements only. (The quantitative theory of bootstrap confirmation to be developed in this paper does make graded judgements possible. See Douven (2004) for an analysis of coherence explicitly in terms of that theory.[7])

The theory of bootstrap confirmation as just presented plainly is a qualitative confirmation theory: it allows us to say that a given theory is bootstrap-confirmed by the evidence, but not that it is bootstrap-confirmed by the evidence to a certain degree, nor that one theory is better bootstrap-confirmed by the evidence than another. A quantitative bootstrap theory is not available. In view of the fact that Glymour (1980a, 373, 375f) lists the development of such an account, specifically of that of a probabilistic theory of bootstrapping, among the projects for further research, this is surprising. One might think that there must have been attempts to provide a quantitative version of Glymour's theory, but just none that was successful. But the odd thing is that, to the best of our knowledge, no such attempt has been made at all. How come?

The following may at least partly answer this question. Glymour (1980, Ch. 3) presented what is now commonly known as the problem of old evidence. Basically the problem is that, on a Bayesian account, evidence that is already known at the time a theory is developed can never confirm that theory (for if it is known, it has unit probability, so that, for every theory T consistent with $E, p(T|E) = p(T)$), contrary to what we know from the practice of science. This problem had the Bayesians stupefied for some time, and once they had convinced themselves that it was not as damaging to their position as it had at first appeared,[8] and when they might have considered the question of how to place Glymour's account in a Bayesian framework, interest in bootstrapping had virtually faded[9] – for no good reason, we believe. Of course, there had been the criticisms related to Glymour's initial countenancing of macho-bootstrapping. As we already said, however, to dodge these it required no more than a minor adjustment of the original definition of bootstrap confirmation. There also had been some criticisms directed against Glymour's attempt to characterize, following Hempel, confirmation relations in a strictly syntactical fashion (cf. Christensen 1983, 1990). But, as van Fraassen already showed in his (1983a) by recasting bootstrap confirmation within a semantic approach to theories, Glymour's syntactic approach is not essential to the theory of bootstrap confirmation.

Whatever may be the exact cause of the loss of interest in Glymour's theory, in our opinion it deserves a second chance. By using Bayesian tools to formulate bootstrapping we not only circumvent the problems that beset the original theory due to its purely syntactical orientation but also make it relatively straightforward to extend the latter, qualitative theory to a quantitative one, thereby obtaining a richer theory that permits graded judgements concerning the confirmational status of scientific theories.[10]

3. DESIDERATA FOR A QUANTITATIVE THEORY OF BOOTSTRAP CONFIRMATION

Before we can begin to formulate a quantitative theory of bootstrap confirmation, we must clarify what relations between the evidence and the theory and/or between the hypotheses in the theory themselves intuitively matter to the degree of bootstrap confirmation (assuming that the evidence does bootstrap-confirm the theory). That is to say, we should start by asking what the desiderata are for a quantitative theory of bootstrapping (or, more exactly, what the *extra* desiderata are beyond those already canvassed in Glymour's book, i.e., beyond those that have to be met by any theory of bootstrapping, whether qualitative or quantitative).

The first desideratum is evident and concerns the degree to which, in the separate tests involved in a bootstrap-test of a theory as a whole, the individual hypotheses of the theory are confirmed by the evidence. So, given a measure of relative confirmation, i.e., of the confirmation supplied to a hypothesis relative to one or more auxiliaries, it should for instance hold that, if the hypotheses in some theory T are all 'relatively confirmed' by the evidence to a degree greater than that to which the evidence 'relatively confirms' the hypotheses in some other theory T', then, all else being equal, the evidence bootstrap-confirms T to a greater degree than it bootstrap-confirms T'.

A second desideratum is suggested by Glymour (1980a, 76f, 140). Here he stresses the importance of testing each hypothesis of a given theory in a variety of ways (if possible). It can of course never be excluded that, in testing one hypothesis with respect to another, an error in one compensates for an error in the other, thus leading to spurious confirmation of the hypothesis that is being tested. But the more ways in which we can test a hypothesis, the better we will be guarded against such spurious confirmation. Thus, while for Clause 1.a of Definition 2.1 to be satisfied it suffices if for each hypothesis of a given theory there is exactly one subset of the theory's axioms with respect to which the hypothesis is confirmed, it is nevertheless desirable that for each hypothesis there are more such subsets – and the more there are, the better it is.

However, for all the weight Glymour puts on variety of testing in the general discussion of bootstrap testing, it plays no role whatsoever, and *can* play no role whatsoever, in his formal theory. There is, for instance, no possibility to express in the framework of this theory that in testing T each of the hypotheses in T have been confirmed relative to a great variety of auxiliaries also in T whereas the hypotheses in T^* have each been confirmed only relative to one set of auxiliaries in T^*. And it is also difficult to see how this difference could be brought out by Glymour's account, given that it is only a qualitative account. Now assume, just for the moment, that the theory were quantitative, i.e., that it would allow us to express that a given theory is bootstrap-confirmed by the evidence to a degree of x. Then if that theory is to respect the intuition that variety of testing matters, it should in any case entail that, if evidence E bootstrap-confirms both theories T and T' but T is tested in a greater variety of ways than T', then, all else being equal, E bootstrap-confirms T to a greater extent than it bootstrap-confirms T'. This we will take to be a further important desideratum for any quantitative version of Glymour's theory of bootstrap testing.

A third desideratum is related to the non-triviality condition, the condition expressed by Clause 1.b of Definition 2.1. Recall that this condition demands that the auxiliaries do not shield the hypothesis under scrutiny from disconfirmation. Here, too, we have a condition that in Glymour's theory is of the yes-or-no type: either the auxiliaries shield the hypothesis from disconfirmation or they do not. But it seems intuitively plausible that auxiliaries can shield the hypothesis to a greater or lesser degree.

To render precise this intuition, it is easiest if we first recast Definition 2.1 in probabilistic terms. In those terms, confirmation of a hypothesis by the evidence is generally taken to mean that the hypothesis conditional on the evidence has a higher probability than taken on its own; disconfirmation on this account means that the hypothesis conditional on the evidence has a lower probability than taken on its own. If Duhem, Quine, Glymour, and many others are correct, and all confirmation requires the use of auxiliaries, then presumably the foregoing should be revised as follows: evidence E confirms hypothesis H relative to auxiliaries $A_1, \ldots, A_n$ exactly if $p(H \mid E \wedge A_1 \wedge \cdots \wedge A_n) > p(H \mid A_1 \wedge \cdots \wedge A_n)$; for disconfirmation, replace '$>$' by '$<$'.[11] This suggests a probabilistic version both of Clause 1.a and of Clause 2 of Definition 2.1.

How should Clause 1.b read in probabilistic terms? It turns out that, given the just-suggested probabilistic version of the first subclause, the second subclause is taken care of quite automatically. For it is straightforward that if $p(H \mid H' \wedge E) > p(H \mid H')$, then there is also evidence E' such that $p(H \mid H' \wedge E') < p(H \mid H')$, that is, in that case there is possible evidence disconfirming H relative to the same auxiliary as that relative to which E confirms it. After all, by the law of total probability, we have that

$$p(H \mid H') = p(H \mid H' \wedge E)p(E \mid H') + p(H \mid H' \wedge \neg E)p(\neg E \mid H').$$

Thus, since $p(E \mid H') = 1 - p(\neg E \mid H')$, the probability $p(H \mid H')$ is a mixture of $p(H \mid H' \wedge E)$ and $p(H \mid H' \wedge \neg E)$.[12] And so if $p(H \mid H' \wedge E) > p(H \mid H')$, it must be that $p(H \mid H' \wedge \neg E) < p(H \mid H')$.

Putting all this together now, we obtain the following probabilistic definition of bootstrap confirmation: [13, 14]

DEFINITION 3.1 (Probabilistic Bootstrap Confirmation). Evidence E probabilistically bootstrap-confirms theory $T = \{H_1, \ldots, H_n\}$ precisely if $p(T \wedge E) > 0$ and for each $i \in \{1, \ldots, n\}$ it holds that

1. there is a $T' \subset T$ such that $H_i \notin T'$ and $p(H_i \mid T' \wedge E) > p(H_i \mid T')$; and
2. there is no $T'' \subseteq T$ such that $p(H_i \mid T'' \wedge E) < p(H_i \mid T'')$.

As an obvious analogue of Definition 2.2 we have:[15]

DEFINITION 3.2 (Probabilistic Bootstrap Disconfirmation). Evidence E probabilistically bootstrap-disconfirms theory $T = \{H_1, \ldots, H_n\}$ if and only if for at least one $i \in \{1, \ldots, n\}$ there is a $T' \subseteq T$ such that $p(H_i \mid T' \wedge E) < p(H_i \mid T')$.

Appendix A charts some logical relationships between Definitions 3.1 and 3.2 and simple probabilistic confirmation.

It was just said that if in a test the hypothesis under scrutiny has unit probability given the auxiliary, that test can yield no disconfirmation of the hypothesis; this corresponds to a situation in which the non-triviality condition of Definition 2.1 is not met. But of course the probability of the hypothesis conditional on the auxiliary may assume any value in the interval (0, 1). And it seems intuitively clear that it makes some difference for the test whether this value is closer to one end of the interval than to the other. For instance, if the probability of the hypothesis conditional on the auxiliary is high, then although that does not completely trivialize the test in the sense that the auxiliary shields that hypothesis against a negative test result, something like shielding does occur in that case, in view of the fact that, *ceteris paribus,* the higher the probability of H conditional on H' is, the less impact evidence can have on H in a test relative to H'.[16] Accordingly, where $H_1, \ldots, H_m$ are the auxiliaries used in a particular test of H, we will say that $p(H \mid \bigwedge_{k=1}^{m} H_k)$ measures the degree of triviality of that test. As a third desideratum for a quantitative bootstrap theory, then, we propose that it take account of the degree of triviality of each of the tests passed by the various hypotheses of a theory.

While there may well be further desiderata for a quantitative theory of bootstrapping, we believe the above three to constitute the main ones, and our measure of bootstrap confirmation will be tailored to meet them (it may well meet others, of course, or be adaptable to others – if such there be).[17]

4. A QUANTITATIVE THEORY OF BOOTSTRAP CONFIRMATION

In this section we first define, and illustrate the use of, a quantitative measure of the bootstrap support a given theory receives from a certain body of evidence (Section 4.1). We then state some useful theorems concerning this measure (Section 4.2). Finally, we define a quantitative theory of bootstrap testing in terms of it (Section 4.3).

4.1. *A Measure of Bootstrap Confirmation*

If we let '*Th*' denote the class of all (finitely axiomatizable) theories that can be formulated in a given language and '*Sent*' the class of sentences of that language apt to report evidence (as was said earlier, this class is taken to be coextensive with the class of all sentences belonging to the language), then the measure we intend to define can be expressed as a function $B : Th \times Sent \to \mathbb{R}$, where the value indicates what might, slightly misleadingly (see below), be termed degree of bootstrap confirmation. We know from the discussion in the previous section the, or at least some major, desiderata for this function. Unfortunately, these desiderata fail to determine a *unique* measure of bootstrap confirmation. Here we are happy to provide one such measure, but refrain from claiming that it is the *true* one (if it makes sense at all to speak of a true measure).

One reason why we do not get a unique measure of bootstrap confirmation from our desiderata is that, for all anyone has said so far, there is no unique measure of confirmation for single hypotheses. We want our measure of bootstrap support for theories to depend on the degree of confirmation the evidence supplies to the various hypotheses comprised by the theories in the various tests the theories allow us to perform (that was the first desideratum).

Adapting the probability definition of confirmation to the needs of bootstrap testing, confirmation of a hypothesis, we said, comes down to the claim that the probability of the hypothesis conditional on the auxiliary or auxiliaries plus the evidence exceeds the hypothesis' probability conditional on the auxiliary/auxiliaries alone. While this seems to be unproblematic, it is unclear how we are to determine the degree to which the evidence confirms the hypothesis with respect to the auxiliary/auxiliaries. This problem already arises for 'plain' (i.e., non-bootstrap) confirmation. Several functions have been proposed as candidates for providing the 'true' measure of such confirmation, but none, so far, can count on any general acclaim. Nonetheless the following can be said to enjoy some popularity (for later purposes, we shall call these the *standard* measures of confirmation):[18]

–*the difference measure:* $d(H, E) =_{df} p(H \mid E) - p(H)$;
–*the* (*log*-)*ratio measure:* $r(H, E) =_{df} p(H \mid E)/p(H)$ (or, as some prefer, $r(H, E) = \log[p(H \mid E)/p(H)]$);
–*Carnap's relevance measure:* $\mathfrak{r}(H, E) =_{df} p(H \wedge E) - p(H)p(E)$;
–*the* (*log*-)*likelihood measure:* $l(H, E) =_{df} p(E \mid H)/p(E \mid \neg H)$ (or the logarithm of that ratio).

All these functions can be readily adapted to meet our concerns. Choosing any particular one must, at least at this point, lack a solid philosophical motivation. For simplicity, and because it seems to be slightly more popular than the others, we will adapt the difference measure for the purpose of measuring in a bootstrap test the confirmation of the separate hypotheses in the separate tests, as follows:[19]

DEFINITION 4.1.

$$d^*(H; H', E) =_{df} p(H \mid H' \wedge E) - p(H \mid H').$$

Of course, if one were to pick one of the others and adapt that, then that would result in a different measure of bootstrap confirmation. This is a major reason why the desiderata of Section 3 are (presently) incapable of determining a unique such measure. It is important to note, though, that all the theorems to be given in the remainder would go through if one were to replace the measure d^* by the (adapted) Carnap measure – i.e., the measure $\mathfrak{r}^*(H; H', E) =_{df} p(H \wedge H' \wedge E) - p(H \mid H')p(H' \wedge E)$ – in the measure of bootstrap confirmation to be defined below; and all theorems except Theorem 4.4 would go through if one were to replace d^* by the adapted (log-)ratio measure $r^*(H; H', E) =_{df} p(H \mid H' \wedge E)/p(H \mid H')$ (or $\log[p(H \mid H' \wedge E)/p(H \mid H')]$).[20] However, neither Theorem 4.2 nor Theorem 4.4 nor Theorem 5.1 holds for the adapted (log-)likelihood measure $l^*(H; H', E) =_{df} p(H' \wedge E \mid H)/p(H' \wedge E \mid \neg H)$ (or $\log [p(H' \wedge E \mid H)/p(H' \wedge E \mid \neg H)]$).[21]

The desideratum of variety of testing can now easily be realized: simply add up the support the evidence provides each of the hypotheses in a theory relative to each of the possible sets of auxiliaries taken from the same theory. Then, *ceteris paribus*, and provided a theory is bootstrap-confirmed, the more tests of each of its hypotheses it allows, the higher will be its degree of bootstrap support.

In order to satisfy our last desideratum we must make sure that the higher the 'degree of triviality' of a test of a particular hypothesis is, the less that test can contribute to the degree of bootstrap support the theory as a whole obtains. It was previously argued that the conditional probability of a hypothesis given one or more other hypotheses can be regarded as measuring the degree of triviality of a test of the former involving exactly the latter as auxiliaries. It appears that by letting our function d^* measure the degree of confirmation of single hypotheses and by measuring total bootstrap support by summing up the support each hypothesis receives from the evidence relative to each possible set of auxiliaries, we have already catered for this last desideratum: *ceteris paribus*, the higher the degree of triviality of a test, the smaller the value of d^* will be and thus, *ceteris paribus*, the smaller the total amount of bootstrap support of the theory will be. To see this, consider that, e.g., where H is the hypothesis being tested and $H_1, \ldots, H_m$ are the auxiliaries being used in the test, if $p(H \mid \bigwedge_{j=1}^{m} H_j) = 1$, then a test of H relative to $H_1, \ldots, H_m$ cannot possibly add anything to the overall bootstrap support the theory comprising $H, H_1, \ldots, H_m$ receives from E whatever E is. And if $p(H \mid \bigwedge_{j=1}^{m} H_j) \neq 1$ but still high, then a test of H relative to $H_1, \ldots, H_m$ can at best add a slight amount to the overall bootstrap support the theory receives from the evidence. But of course it holds quite generally that the lower $p(H \mid \bigwedge_{j=1}^{m} H_j)$ is, the more a test of H relative to $H_1, \ldots, H_m$ can add to the bootstrap support T receives from the evidence.

To put this in more formal terms, let $T = \{H_1, \ldots, H_n\}$. For each $H_i \in T$, there are exactly 2^{n-1} sets of auxiliary hypotheses also in T with respect to which it can be tested, namely, all the elements of the power set of T minus H_i, that is, $\wp(T \backslash \{H_i\})$.[22] Given some ordering $\langle H^T_{i_1}, \ldots, H^T_{i_{2^{n-1}}} \rangle$ of $\wp(T \backslash \{H_i\})$, let '$H^T_{i_j}$' denote the jth member of that ordering. Finally, let '$\bigwedge H_{i_j}^T$' denote the conjunction of the hypotheses in $H_{i_j}^T$. Then we can define our measure of bootstrap confirmation, B, as follows:[23]

DEFINITION 4.2 (Measure of Bootstrap Confirmation).

$$B(T, E) =_{df} \sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^T, E).$$

As already intimated, the name of this function may be slightly misleading in that it suggests that '$B(T, E) = x$' is to be generally interpreted as saying that E bootstrap-confirms T to a degree of x, an interpretation that does *not* seem appropriate in all cases. Indeed, it only seems natural to say that B measures bootstrap confirmation if B takes as its arguments a theory and a collection of data such that the latter bootstrap-confirms the former in the qualitative sense defined earlier. Nevertheless, for want of a better name, we shall stick to this one.

To get a feel for Definition 4.2, it may be helpful to see an actual application of it. Consider the following example in which the bootstrap support the evidence supplies to a theory consisting of four axioms is calculated:

EXAMPLE 4.1. Theory T has as axioms hypotheses H_1, H_2, H_3, and H_4. Each of these hypotheses has a prior probability of .25, and they are mutually probabilistically independent.[24] Evidence E has a prior probability of .5. Further we have the following:[25]

- $p(E \wedge H_i) = .125$, for all $i \in \{1, \ldots, 4\}$;
- $p(E \wedge H_i \wedge H_j) = .05$, for all $i, j \in \{1, \ldots, 4\}$ such that $i \neq j$;
- $p(E \wedge H_i \wedge H_j \wedge H_k) = .015$, for all $i, j, k \in \{1, \ldots, 4\}$ such that $i \neq j \neq k$;[26]
- $p(E \wedge H_1 \wedge \cdots \wedge H_4) = .0038$.

It can easily be verified that, given the previous assumptions, E bootstrap-confirms T in the sense of Definition 3.1. To calculate the *degree* of support E provides for T, first derive for all $i \in \{1, \ldots, 4\}$ that

- $d^*(H_i; E) = (p(H_i \wedge E)/p(E)) - p(H_i) = (.125/.5) - .25 = 0;$
- for all $j \in \{1, \ldots, 4\}$ such that $i \neq j$:

$$d^*(H_i; H_j, E) = \frac{p(H_i \wedge H_j \wedge E)}{p(H_j \wedge E)} - \frac{p(H_i \wedge H_j)}{p(H_j)} = \frac{.05}{.125} - \frac{.0625}{.25} = .15;$$

- for all $j, k \in \{1, \ldots, 4\}$ such that $i \neq j \neq k$:

$$d^*(H_i;H_j \wedge H_k, E) = \frac{p(H_i \wedge H_j \wedge H_k \wedge E)}{p(H_j \wedge H_k \wedge E)} - \frac{p(H_i \wedge H_j \wedge H_k)}{p(H_j \wedge H_k)}$$
$$= \frac{.015}{.05} - \frac{.015625}{.0625} = .05;$$

- for all $j, k, l \in \{1, \ldots, 4\}$ such that $i \neq j \neq k \neq l$:

$$d^*(H_i; H_j \wedge H_k \wedge H_l, E) = \frac{p(H_i \wedge H_j \wedge H_k \wedge H_l \wedge E)}{p(H_j \wedge H_k \wedge H_l \wedge E)} - \frac{p(H_i \wedge H_j \wedge H_k \wedge H_l)}{p(H_j \wedge H_k \wedge H_l)}$$
$$= \frac{.0038}{.015} - \frac{.00390625}{.015625} \approx .003.$$

Since for each H_i there is exactly one way in which it can be tested relative to no auxiliary hypotheses, three different ways in which it can be tested relative to one auxiliary hypothesis, three different ways in which it can be tested relative to two auxiliary hypotheses, and one way in which it can be tested relative to three auxiliary hypotheses, the total bootstrap support each of the H_i *individually* receives from E equals (approximately): $(1)(0) + (3)(.15) + (3)(.05) + (1)(.003) = .603$. Since the bootstrap support T receives is just the sum of the bootstrap supports each of its axioms receives, $B(T, E) \approx (4)(.603) = 2.412$.[27]

4.2. *Some Theorems*

Before stating our quantitative account of bootstrapping, we want to point to some important facts concerning B.

First, the standard measures of confirmation given earlier apply to theories (that is, to sets of hypotheses) no less than they do to single hypotheses. Our measure of bootstrap support would of course be entirely superfluous if it coincided with one of these standard measures or indeed with any of the other known measures of confirmation (see Note 18), or if the degree of bootstrap support provided by some piece of evidence to a theory were just the degree of confirmation of the theory by the evidence given some of those measures of confirmation modulo some scale transformation. This is not the case, however. Note that each of the known measures of confirmation makes the degree of confirmation a theory T receives from evidence E a function of some subset of $\{p(T), p(E), p(T \mid E),$

$p(T \mid \neg E), p(E \mid T), p(E \mid \neg T)\}$. More generally, call any measure, whether or not actually proposed, that is a function of any such subset a non-bootstrap measure of confirmation. Then we have the following theorem (see Appendix B for proofs of this and of Theorems 4.2–4.4):

THEOREM 4.1. There is no function f such that, for all T and E, $B(T, E) = f \circ m(T, E)$, with m any non-bootstrap measure of confirmation.

Second, we have some theorems concerning relations between the measure B and qualitative bootstrap confirmation and disconfirmation as defined by Definitions 3.1 and 3.2, respectively:

THEOREM 4.2. For all T and E, if $B(T, E) \not> 0$, then E does not bootstrap-confirm T; if in addition $B(T, E) < 0$, then E bootstrap-disconfirms T.

THEOREM 4.3. There is no $a \in \mathbb{R}$ such that, for all T and E, if $B(T, E) > a$, then E bootstrap-confirms T, nor is there a $b \in \mathbb{R}$ such that, for all T and E, if E bootstrap-disconfirms T, then $B(T, E) < b$.

Note that the first conjunct of Theorem 4.3 only indicates that there is no *general* numerical threshold value for qualitative bootstrap confirmation in the sense that, for any T and E, if we are informed that $B(T, E)$ has a value above that threshold, we can immediately infer that E qualitatively bootstrap-confirms T. This leaves open the possibility that with every *particular* theory T some value a can be associated such that, if $B(T, E) > a$ for some E, then E qualitatively bootstrap-confirms T. In fact, as the following theorem shows, something stronger holds: for each *class* of theories that have the same number of axioms there is a threshold value such that for every theory T within the class, if $B(T, E)$ is greater than or equal to the threshold value, E bootstrap-confirms T:

THEOREM 4.4. For all $n \in \mathbb{N}$, T, and E, if $T = \{H_1, \ldots, H_n\}$ and $B(T, E) \geqslant (n)(2^{n-1}) - 1$, then E bootstrap-confirms T.

This theorem is of limited interest, though. As the proof clearly shows, a value equal to or greater than the threshold is only reached in the special case in which *each* axiom of a theory is confirmed by the evidence relative to *each* possible set of auxiliaries also from that theory (recall that Definition 3.1 only requires that each axiom is confirmed relative to *some* possible set of auxiliaries and not disconfirmed relative to any of the others).[28]

By the same reasoning as is utilized in the proof of Theorem 4.4, we can associate a numerical threshold – to wit, also $(n)(2^{n-1}) - 1$ – with every class of theories axiomatized by n axioms ($n \in \mathbb{N}$) such that, if some E bootstrap-disconfirms a theory T within that class, then $B(T, E)$ is below that threshold. Here, too, we do not have a contradiction with Theorem 4.3, whose second conjunct only denies the stronger claim that there is a *general* numerical threshold such that all theories that are bootstrap-disconfirmed have a B-value below that threshold.

4.3. *Quantitative Bootstrap Confirmation Defined*

Turning now to the task of defining quantitative bootstrap confirmation, first notice that Theorem 4.3, which says that no value of B necessarily indicates bootstrap confirmation, implicates that (unless we are willing to say that E bootstrap-confirms T to a degree of x even in a case in which E does not bootstrap-confirm T at all according to Definition 3.1) we cannot simply have a quantitative theory of bootstrapping according to which E bootstrap-confirms T to a degree of x just in case $B(T,\ E) = x > a$, for some $a \in \mathbb{R}$. Second, while a value of $B(T,\ E)$ of 0 or below indicates that the evidence does not bootstrap-confirm the theory (and a value below 0 even that the evidence bootstrap-disconfirms the theory), and a value of $B(T,\ E)$ of $(n)(2^{n-1}) - 1$ or higher indicates, for a theory with n axioms, that it is bootstrap-confirmed by the evidence, nevertheless if $0 < B(T,\ E) < (n)(2^{n-1}) - 1$, no similar conclusion can be drawn: all values within that range are compatible with E bootstrap-confirming, bootstrap-disconfirming, and being bootstrap-irrelevant to, T. Hence, nor can we have a quantitative theory of bootstrapping according to which E bootstrap-confirms T to a degree of x just in case $B(T, E) = x$ and x is above some threshold that is a function of the number of T's axioms. The following quantitative bootstrap theory is hardly more involved, however:

DEFINITION 4.3 (Quantitative Bootstrap Confirmation). Evidence E bootstrap-confirms theory $T = \{H_1, \ldots, H_n\}$ to a degree of x exactly if

1. clauses 1 and 2 of Definition 3.1 are satisfied; and
2. $B(T, E) = x$.

What this definition basically says is that Definition 3.1 determines *when* a theory is bootstrap-confirmed by the evidence, and that thereupon the measure of bootstrap-confirmation as specified by Definition 4.2 determines *to what degree* the theory is bootstrap-confirmed by the evidence. Quantitative bootstrap disconfirmation can be defined in a parallel fashion.

The account of bootstrap confirmation we have hereby obtained enables us to express such things as that, if evidence E bootstrap-(dis)confirms a given theory T, then it does this the stronger the higher (lower) the value of $B(T, E)$ is, and also that E bootstrap-(dis)confirms theory T to a greater extent than some other theory T' if E bootstrap-(dis)confirms both T and T' and $B(T, E) > B(T', E)$, or $B(T, E) < B(T', E)$, respectively.

5. A PUZZLE ABOUT BOOTSTRAP CONFIRMATION

What is the connection between bootstrap support and justified belief? We here consider a puzzle raised in van Fraassen (1983b), which seems to show that qualitative bootstrap confirmation of some theory does not give reason to believe that theory. As will be seen, our quantitative theory may give rise to basically the same puzzle. But we suggest that the puzzle does not necessarily show that bootstrapping is not really a theory of confirmation at all.

Consider a simple, comparative definition of justification in terms of bootstrap confirmation

(B1) If E bootstrap-confirms T but not T', then, if E is our total evidence, belief in T is more justified than belief in T'.
(B2) If E bootstrap-confirms T to a higher degree than T', then, if E is our total evidence, belief in T is more justified than belief in T'.

Though appealingly simple, this answer to our initial question appears to conflict with the following principle:

(P) If $p(T \mid E) > p(T' \mid E)$, then, if E is our total evidence, belief in T' cannot be more justified than belief in T,

a principle that van Fraassen (1983b) presents as a truism.

Van Fraassen's argument that (B1) and (P) conflict goes roughly as follows: Suppose $T = \{H_1, H_2, H_3\}$ is bootstrap-confirmed by E (according to Definition 3.1). Suppose in particular that E confirms H_1 relative to H_2 (and perhaps also relative to H_3 and to $H_2 \wedge H_3$), E confirms H_2 relative to H_3 (and perhaps also relative to $H_1 \wedge H_3$, but not relative to H_1), and E confirms H_3 relative to H_1 (and perhaps also relative to $H_1 \wedge H_2$). Now the subset $T' \subset T$, containing only H_1 and H_2, is *not* bootstrap-confirmed by E (for E does not confirm H_2 relative to H_1). Without loss of generality, assume that $\{H_1, H_2, E\} \nvdash H_3$. Then it is an elementary truth of probability theory that $p(T' \mid E) > p(T \mid E)$. So, according to (B1), we are more justified in believing T than in believing T', but according to (P), we are not.

It must be immediately clear that although (B1) always conflicts with (P), this does not hold for (B2). In quantitative bootstrapping there are various ways in which a theory can become more bootstrap-confirmed. Many of these will simply raise the probability of the theory given the evidence, and in such cases (B2) and (P) are in perfect accordance with each other. This said, it is not hard to see that in other cases (B2) and (P) will still conflict.

For example, consider subset T' of T of Example 4.1, consisting of hypotheses H_1, H_2, H_3. This theory has a probability of .03 given evidence E, which clearly exceeds that of T given E ($= .0076$). However, it is bootstrap-supported by E to a much lower degree than T, namely $B(T', E) = 1.05$ ($B(T, E) \approx 2.412$). Thus, according to (B2), we are more justified in believing T than in believing T', but according to (P) we are not.

To generalize the problem, note that, if $T \subset T'$, then, for all E, $p(T \mid E) \geqslant p(T' \mid E)$. And unless $T \cup \{E\} \vdash T'$, $p(T \mid E)$ will even exceed $p(T' \mid E)$. On the other hand, we have the following theorem (see Appendix C for a proof):

THEOREM 5.1. For all T, T', and E, if $T \subset T'$ and E does not bootstrap-disconfirm T', then $B(T', E) \geqslant B(T, E)$; if in addition E bootstrap-confirms T', then $B(T', E) > B(T, E)$.

Hence, degree of bootstrap confirmation and probability may pull in opposing directions; they cannot always be jointly maximized. But then how can bootstrap confirmation be related to justification?

Van Fraassen's conclusion is that it cannot. This is not to say that van Fraassen believes a positive bootstrap test is insignificant. Quite the contrary – he believes it gives reason to *accept* a theory, where the notion of acceptance is considerably weaker than that of belief (van Fraassen 1983a, 1983b). More specifically, acceptance of a theory involves the belief that it is empirically adequate (roughly, true of the observable part of the world) as well as a commitment to use the theory's conceptual apparatus in describing future phenomena (van Fraassen 1980, Ch. 1). Should his view on bootstrap testing be correct, then that hardly detracts from the importance of having a quantitative account of bootstrap confirmation: surely it makes sense to say that one bootstrap test provides stronger reason to accept a given theory than another bootstrap test, and it seems that only a quantitative bootstrap theory is capable of capturing that intuition. However, van Fraassen's conclusion may not be inescapable.

Plausible though it may appear, principle (P) has been denied by, among others, Levi (1967), Kaplan (1981a), (1981b), Lehrer (1990), and Maher (1993), who have argued that – loosely – justification has the structure of a decision-making problem. In such a problem, one heeds not only the probabilities of the possible outcomes of a certain decision, but also their utilities. More exactly, in decision making one chooses the option that has greatest expected utility of the available alternatives, where an option's expected utility is just the sum of the utilities of its various possible outcomes weighted by the probabilities of those outcomes.[29] According to the aforementioned authors, there is nothing in the way decision theory is set up that would prevent applying it to matters epistemological; we can perfectly well assign cognitive or epistemic utilities to the 'acts' of accepting, rejecting, and suspending judgement on particular hypotheses or theories under particular circumstances, and then apply the decision-theoretic apparatus to these acts in the normal manner in order to determine what the agent is justified to do. So, on this approach to justification, a person may well be more justified in believing one theory than she is in believing a second even if the former is less likely to her than the latter, because, given her probabilities and utilities, the former may well have a greater expected cognitive utility than the latter.[30]

Now the notion of utility is anything but crystal-clear.[31] The notion of cognitive utility appears even more problematic. We are told that the cognitive utility of accepting some hypothesis depends on the informativeness of that hypothesis (cf. e.g., Lehrer 1990; Maher 1993). But the notion of informativeness itself is still very much in need of clarification. Maher thinks this notion is to be cashed out in terms of verisimilitude. However, given that there is still widespread disagreement over the nature of verisimilitude (cf. e.g., Niiniluoto 1998), this suggestion seems rather unhelpful.[32] One way in which our quantitative bootstrap account could be positively related to justification is by replacing, in cognitive decision theory, the ill-defined notion of cognitive utility by the clearly defined notion of degree of bootstrap support. In order to determine the justificational status of a hypothesis or theory we would thus have to weigh not probability and utility but probability and degree of bootstrap support against one another. This is only a rough proposal that can be filled out in quite diverse ways. We will not explore the possibilities here. Our aim in this section merely was to point out that there may still be a positive role for degree of bootstrap confirmation even though a higher degree of bootstrap support does not generally indicate a higher probability.[33]

6. CONCLUDING REMARKS

In our view, it is hard to overestimate the philosophical significance of Glymour's work on bootstrap confirmation. Still, the work was in an important respect left unfinished: Glymour provided a qualitative theory only. In this paper, we have begun the formulation of a quantitative account of bootstrapping. We started by laying down a number of desiderata for such an account, and then defined, guided by these desiderata, a quantitative measure of bootstrap support. Quantitative bootstrap confirmation could then be rather straightforwardly defined by means of that measure.

As was indicated at various junctures, the theory offered here is far from meant to be the final word on quantitative bootstrap confirmation, but rather the beginning of a larger project. Two avenues for future research deserve special mentioning. First, it will be remembered that we left open the possibility that there are other desiderata for a quantitative theory of bootstrap confirmation beyond those we identified. More research is needed to see whether there indeed are and, if so, what (if any) changes our measure of bootstrap confirmation will need to undergo. Secondly, still relatively little is known about the mathematical properties of this measure and, especially, of the variant measures that are obtained if, instead of on the difference measure, the measure of bootstrap confirmation is built upon another measure of probabilistic confirmation. We in fact hope that a comparison of the mathematical properties of the various measures may help us settle on a particular measure in a more motivated way than was done in the present paper.

APPENDIX A: BOOTSTRAP CONFIRMATION VERSUS PROBABILISTIC CONFIRMATION

In this appendix we prove some facts concerning the logical relations between bootstrap confirmation and probabilistic confirmation.

'Bootstrap confirmation' and 'bootstrap-disconfirmation' – and derived terms – are here as well as in the following appendices understood as defined by Definitions 3.1 and 3.2; instead of 'probabilistically-(dis)confirms' we will throughout simply write '(dis)confirms'.

THEOREM A.1. For all T and E, if E bootstrap-confirms T, then E also confirms T.

Proof. Let $T = \{H_1, \ldots, H_n\}$ and suppose E bootstrap-confirms T. Then, by Clause 1 of Definition 3.1, there must for every $H_i \in T$ be at least one $T' \subset T \setminus \{H_i\}$ such that $p(H_i \mid E \wedge T') > p(H_i \mid T')$. Thus in particular there must for H_1 be a subset T^* of $T \setminus \{H_1\}$ such that

$$p(H_1 \mid E \wedge T^*) > p(H_1 \mid T^*). \tag{1}$$

Now let $\pi_1, \ldots, \pi_{n!}$ denote the permutations on $1, \ldots, n$. Clearly

$$\begin{aligned} p(E \wedge T) &= p(E)p(H_{\pi_i(1)} \mid E) \cdots p\big(H_{\pi_i(n)} \mid E \wedge H_{\pi_i(1)} \wedge \cdots \wedge H_{\pi_i(n-1)}\big) \\ &= p(E)p(H_{\pi_j(1)} \mid E) \cdots p\Big(H_{\pi_j(n)} \mid E \wedge H_{\pi_j(1)} \wedge \cdots \wedge H_{\pi_j(n-1)}\Big) \end{aligned}$$

for all $i, j \in \{1, \ldots, n!\}$. Observe that for some $k \in \{1, \ldots, n!\}$, $p(H_1 \mid E \wedge T^*)$ must occur as a factor in $p(E)p(H_{\pi_k(1)} \mid E) \cdots p(H_{\pi_k(n)} \mid E \wedge H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(n-1)})$. Next suppose, towards a reductio, that the consequent of the theorem does not hold, i.e., E does not confirm T. Then $p(T \mid E) \leqslant p(T)$, or $[p(T \wedge E)/p(E)] \leqslant p(T)$, or again written differently

$$\frac{p(E \wedge H_1 \wedge \cdots \wedge H_n)}{p(E)} \leqslant p(H_1 \wedge \cdots \wedge H_n).$$

By the general multiplication rule and after cancelling $p(E)$ in the left-hand expression, this is equivalent to

$$\begin{aligned} &p(H_1 \mid E) \cdots p(H_n \mid E \wedge H_1 \wedge \cdots \wedge H_{n-1}) \\ &\quad \leqslant p(H_1)p(H_2 \mid H_1) \cdots p(H_n \mid H_1 \wedge \cdots \wedge H_{n-1}). \end{aligned} \tag{2}$$

Given (2), the following must also hold:

$$\begin{aligned} &p(H_{\pi_k(1)} \mid E) \cdots p(H_{\pi_k(n)} \mid E \wedge H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(n-1)}) \\ &\quad \leqslant p(H_{\pi_k(1)})p(H_{\pi_k(2)} \mid H_{\pi_k(1)}) \cdots p(H_{\pi_k(n)} \mid H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(n-1)}). \end{aligned}$$

From inequality (1) we know that for one i with $1 \leqslant i \leqslant n$, it must hold that

$$\begin{aligned} &p(H_{\pi_k(i)} \mid E \wedge H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(i-1)}) \\ &\quad > p(H_{\pi_k(i)} \mid H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(i-1)}). \end{aligned}$$

Combining this with (3), we get that for at least one j with $1 \leqslant j \leqslant n$

$$p(H_{\pi_k(j)} \mid E \wedge H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(j-1)}) < p(H_{\pi_k(j)} \mid H_{\pi_k(1)} \wedge \cdots \wedge H_{\pi_k(j-1)})$$

for else the left-hand side of (3) will be larger than the right-hand side. It follows that there is a $H_i \in T$ and a $T' \subset T$ such that $p(H_i \mid E \wedge T') < p(H_i \mid T')$. But this violates Clause 2 of Definition 3.1 and hence our assumption that E bootstrap-confirms T. Thus the assumption that E does not confirm T is false. □

THEOREM A.2. For all T and E, if E disconfirms T, then E also bootstrap-disconfirms T.

Proof. Assume the antecedent, i.e., $p(T \mid E) < p(T)$, where $T = \{H_1, \ldots, H_n\}$. Then

$$\frac{p(T \wedge E)}{p(E)} < p(T)$$

or, with T written out,

$$\frac{p(E \wedge H_1 \wedge \cdots \wedge H_n)}{p(E)} < p(H_1 \wedge \cdots \wedge H_n).$$

Multiplying both sides by $p(E)$ yields

$$p(E \wedge H_1 \wedge \cdots \wedge H_n) < p(E)p(H_1 \wedge \cdots \wedge H_n).$$

Using the general multiplication rule for both sides, we obtain

$$p(E)p(H_1 \mid E) \cdots p(H_n \mid E \wedge H_1 \wedge \cdots \wedge H_{n-1}) < p(E)p(H_1) \cdots p(H_n \mid H_1 \wedge \cdots \wedge H_{n-1}),$$

which can only be the case if

$$[p(H_1 \mid E) < p(H_1)] \vee \cdots \vee [p(H_n \mid E \wedge H_1 \wedge \cdots \wedge H_{n-1}) < p(H_n \mid H_1 \wedge \cdots \wedge H_{n-1})].$$

Thus there is at least one $H_i \in T$ and at least one $T' \subset T \backslash \{H_i\}$ such that $p(H_i \mid E \wedge T') < p(H_i \mid T')$ and hence, by Definition 3.2, E bootstrap-disconfirms T. □

Theorems A.3 and A.4 show that neither Theorem A.1 nor Theorem A.2 can be strengthened to a bi-implication.

THEOREM A.3. It is not the case that, for all T and E, if E confirms T, then also E bootstrap-confirms T.

Proof. Let theory T have as axioms hypotheses H_1, H_2, and H_3, each of which has a prior probability of .1; let E have a prior probability
of .5. Further assume the following:

- $p(H_i \wedge H_j) = .015$ for all $i,j \in \{1,2,3\}$ such that $i \neq j$;
- $p(H_1 \wedge H_2 \wedge H_3) = .005$;
- $p(E \wedge H_i) = .05$ for all $i \in \{1,2,3\}$;
- $p(E \wedge H_i \wedge H_j) = .004$ for all $i, j \in \{1,2,3\}$ such that $i \neq j$;
- $p(E \wedge H_1 \wedge H_2 \wedge H_3) = .003$.

Then E does not bootstrap-confirm T, since for all i, $j \in \{1,2,3\}$,

$$p(H_i \mid E \wedge H_j) = \frac{p(E \wedge H_i \wedge H_j)}{p(E \wedge H_j)} = \frac{.004}{.05} = .08,$$

which is smaller than $.15 (= p(H_i \mid H_j))$. A fortiori, this model violates Clause 2 of Definition 3.1.
However, E does confirm T, for

$$\begin{aligned} p(T \mid E) &= \frac{p(E \wedge H_1 \wedge H_2 \wedge H_3)}{p(E)} \\ &= \frac{.003}{.5} = .006 > .005 = p(T) \end{aligned}$$ □

THEOREM A.4. It is not the case that, for all T and E, if E bootstrap-disconfirms T, then also E disconfirms T.

Proof. From the probability model constructed in the proof of Theorem A.3. □

APPENDIX B: PROOFS OF THEOREMS 4.1–4.4

THEOREM 4.1. There is no function f such that, for all T and E, $B(T,E) = f \circ m(T,E)$, with m any non-bootstrap measure of confirmation.

Proof. We prove this theorem by specifying a probability model involving a theory T^* and evidence E^* for which the following hold (T and E are as in Example 4.1): (i) $p(T^*) = p(T)$, (ii) $p(E^*) = p(E)$, (iii) $p(T^* \mid E^*) = p(T \mid E)$, (and thus also) (iv) $p(E^* \mid T^*) = p(E \mid T)$, (v) $p(T^* \mid \neg E^*) = p(T \mid \neg E)$, and (vi) $p(E^* \mid \neg T) = p(E \mid \neg T)$, but (vii) $B(T^*, E^*) \neq B(T, E)$. It can readily be seen that, given (i)–(vi), and given how we defined the notion of a non-bootstrap measure of confirmation, there can be no function f such that $f \circ m(T, E) = B(T, E)$, where m is such a non-bootstrap measure of confirmation.

Like T, the theory T^* consists of four axioms, $H_1^*, \ldots, H_4^*$. Like the hypotheses in T, the H_i^* all have a prior probability of .25, and are all mutually probabilistically independent; we thus see immediately that (i) holds. Evidence E^* has a prior probability of .5 (like E in Example 4.1; so (ii) holds). Further we have the following:

- $p(E^* \wedge H_i^*) = .125$, for all $i \in \{1, \ldots, 4\}$;
- $p(E^* \wedge H_i^* \wedge H_j^*) = .045$, for all $i, j \in \{1, \ldots, 4\}$ such that $i \neq j$;
- $p(E^* \wedge H_i^* \wedge H_j^* \wedge H_k^*) = .0125$, for all $i, j, k \in \{1, \ldots, 4\}$ such that $i \neq j \neq k$;
- $p(E^* \wedge H_1^* \wedge \cdots \wedge H_4^*) = .0038$.

From the fact that $p(E^* \wedge H_1^* \wedge \cdots \wedge H_4^*) = p(E \wedge H_1 \wedge \cdots \wedge H_4)$ it follows that $p(T^* \mid E^*) = p(T \mid E)$ (so (iii) holds; and given (i)–(iii), (iv) and, by the law of total probability, (v) and (vi) must hold as well). We now calculate the bootstrap support for T^* from the following values, which hold for all $i \in \{1, \ldots, 4\}$:

- $d^*(H_i^*;\, E^*) = (.125/.5) - .25 = 0$;
- $d^*(H_i^*;\, H_j^*\, E^*) = (.045/.125) - .25 = .11$, for all $j \in \{1, \ldots, 4\}$ such that $i \neq j$;
- $d^*(H_i^*;\, H_j^* \wedge H_k^*,\, E^*) = (.125/.045) - .25 \approx .0278$, for all $j, k \in \{1, \ldots, 4\}$ such that $i \neq j \neq k$;
- $d^*(H_i^*;\, H_j^* \wedge H_k^* \wedge H_l^*,\, E^*) = (.0038/.0125) - .25 = .054$, for all $j, k, l \in \{1, \ldots, 4\}$ such that $i \neq j \neq k \neq l$.

So, the bootstrap support each of the H_i^* gets from E^* totals (approximately): (1)(0) + (3)(.11) + (3)(.0278) + (1)(.054) = .4674. And thus $B(T^*, E^*) \approx (4)(.4674) = 1.8696$. This is unequal to the bootstrap support T was seen to get from E, namely 2.412, despite the fact that, as we saw, (i)–(vi) hold, and thus on any non-bootstrap measure of confirmation m, we have $m(T^*, E^*) = m(T, E)$.

THEOREM 4.2. For all T and E, if $B(T,E) \not> 0$, then E does not bootstrap-confirm T; if in addition $B(T,E) < 0$, then E bootstrap-disconfirms T.

Proof. Let $T = \{H_1, \ldots, H_n\}$ and assume that E bootstrap-confirms T. It follows from the second clause of Definition 3.1 together with Definition 4.1 that, for all $H_i \in T$ and all $H_{i_j}^T \in \wp(T \backslash \{H_i\})$, we have $d^*(H_i; \bigwedge H_{i_j}^T, E) \geqslant 0$. From the first clause of Definition 3.1 together with the definition of d^* it follows that for all $H_i \in T$ there is at least one $H_{i_j}^T \in \wp(T \backslash \{H_i\})$ such that $d^*(H_i; \bigwedge H_{i_j}^T, E) > 0$. Since B just tallies the bootstrap support each hypothesis receives relative to each set of possible auxiliaries, it must be that $B(T,E) > 0$.

To see that for all T and E, if $B(T,E) < 0$, then E bootstrap-disconfirms T, let again $T = \{H_1, \ldots, H_n\}$. Then if $B(T,E) < 0$, there must by Definition 4.2 be at least one $H_i \in T$ such that, for at least one $H_{i_j}^T \in \wp(T \backslash \{H_i\})$, we have $d^*(H_i; \wedge H_{i_j}^T, E) < 0$. Hence, for at least one H_i and one $H_{i_j}^T \in \wp(T \backslash \{H_i\})$, we have $p(H_i | \bigwedge H_{i_j}^T \wedge E) < p(H_i \mid \bigwedge H_{i_j}^T)$. And thus, by Definition 3.2, E bootstrap-disconfirms T. □

THEOREM 4.3. There is no $a \in \mathbb{R}$ such that, for all T and E, if $B(T,E) > a$, then E bootstrap-confirms T, nor is there some $b \in \mathbb{R}$ such that, for all T and E, if E bootstrap-disconfirms T, then $B(T,E) < b$.

Proof. We first show that there is no $a \in \mathbb{R}$ such that, for all T and E, if $B(T,E) > a$, then E bootstrap-confirms T. Toward a reductio, suppose that, for all T, E, if $B(T,E) > c$ for some particular $c \in \mathbb{R}$, then E bootstrap-confirms T. Then let $T' = \{H_1, \ldots, H_n\}$ and furthermore let it be the case that $B(T', E') > c$ for some E'. Now let $T'' = \{H_1, \ldots, H_n, H_{n+1}\}$ with H_{n+1} any hypothesis that is probabilistically independent of any subset of $\{H_1, \ldots, H_n, E'\}$ (we can, without loss of generality, assume that such an H_{n+1} exists). Thus in particular the following facts hold:

(i) $p(H_i \mid \bigwedge H_{i_j}^{T'} \wedge H_{n+1}) = p(H_i \mid \bigwedge H_{i_j}^{T'})$ for all $H_i \in T'$ and all $H_{i_j}^{T'} \in \wp(T' \backslash \{H_i\})$;
(ii) $p(H_i \mid \bigwedge H_{i_j}^{T'} \wedge H_{n+1} \wedge E') = p(H_i \mid \bigwedge H_{ij}^{T'} \wedge E')$ for all $H_i \in T'$ and all $H_{i_j}^{T'} \in \wp(T' \backslash \{H_i\})$;
(iii) $p(H_{n+1} \mid \bigwedge H_{n+1_k}^{T''} \wedge E') = p(H_{n+1} \mid \bigwedge H_{n+1_k}^{T''}) = p(H_{n+1})$ for all $H_{n+1_k}^{T''} \in \wp(T'' \backslash \{H_{n+1}\})$.

Dividing into three parts the sum which, by Definition 4.2, gives the value of $B(T'', E')$, we have

$$B(T'', E') = \sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'}, E') + \sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'} \wedge H_{n+1}, E') + \sum_{k=1}^{2^n} d^*(H_{n+1}; \bigwedge H_{n+1_k}^{T''}, E'). \quad (4)$$

Given (i) and (ii), we have, for all i with $1 \leqslant i \leqslant n$ and all $H_{i_j}^{T'}$ such that $1 \leqslant j \leqslant 2^{n-1}$, that $p(H_i \mid \bigwedge H_{i_j}^{T'} \wedge E') - p(H_i \mid \bigwedge H_{i_j}^{T'}) = p(H_i \mid \bigwedge H_{i_j}^{T'} \wedge H_{n+1} \wedge E') - p(H_i \mid \bigwedge H_{i_j}^{T'} \wedge H_{n+1})$, and thus also that $d^*(H_i; \bigwedge H_{i_j}, E') = d^*(H_i; \bigwedge H_{i_j} \wedge H_{n+1}, E')$. And from this it follows that the first two summands in (4) are equal, that is,

$$\sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'}, E') = \sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'} \wedge H_{n+1}, E'). \quad (5)$$

Furthermore, from (iii) it can be immediately seen to follow that

$$\sum_{k=1}^{2^n} d^*(H_{n+1}; \bigwedge H_{n+1_k}^{T''}, E') = 0. \quad (6)$$

Since the first of the summands in (4) equals $B(T', E')$, it follows from(5) and (6) that $B(T'', E') = 2B(T', E')$. Since, furthermore, it follows from Theorem 4.2 that $B(T', E') > 0$, it must be that $B(T'', E') > c$. Thus, by our hypothesis, E' bootstrap-confirms T''. However, it follows from (iii) above that Clause 1 of Definition 3.1 is not satisfied, so that E' does not bootstrap–confirm T''. Hence, the assumption that there is a numerical threshold for bootstrap confirmation leads to contradiction. Hence, there is no such numerical threshold.

To show that neither is there a $b \in \mathbb{R}$ such that, for all T and E, if E bootstrap–disconfirms T, then $B(T, E) < b$, we first note that, if E bootstrap–disconfirms T, then that does not exclude that $B(T, E) > 0$. To see this, we only need to slightly change the model given in Example 4.1. Let in that model $p(E \wedge H_i \wedge H_j \wedge H_k) = .01$ (instead of .015), for all $i, j \in \{1, \ldots, 4\}$ such that $i \neq j$. Then E bootstrap–disconfirms T (for E disconfirms H_i for every $i \in \{1, \ldots, 4\}$ relative to the conjunction of every $H_j, H_k \in T$ such that $i \neq j \neq k : p(H_i \mid H_j \wedge$

$H_k \wedge E) = .01/.05 = .2 < .25 = p(H_i \mid H_j \wedge H_k))$. Still, as an easy calculation shows, $B(T, E) = 1.212$ and is thus positive. Second, we saw in the first part of this proof that if we add an hypothesis H to any theory T that is probabilistically independent of that theory together with the evidence E (in the precise sense specified above), then $B(T', E) = 2B(T, E)$for $T' = T \cup \{H\}$. So let then T be bootstrap-disconfirmed by E and such that $B(T, E) = c > 0$. Adding a probabilistically independent hypothesis to T will result in a theory T' that is also bootstrap-disconfirmed by E but for which $B(T', E) = 2c > c$. Since this procedure can be repeated as often as one likes, there can be no $b \in \mathbb{R}$ such that $B(T, E) \geqslant b$ indicates that T is not bootstrap-disconfirmed by E. □

THEOREM 4.4. For all $n \in \mathbb{N}, T$, and E, if $T = \{H_1, \ldots, H_n\}$ and $B(T, E) \geqslant (n)(2^{n-1}) - 1$, then E bootstrap-confirms T.

Proof. Suppose $B(T, E) \geqslant (n)(2^{n-1}) - 1$ for some E, and $T = \{H_1, \ldots, H_n\}$. Then it follows from Definition 4.2 that for no $H_i \in T$ can it be the case that there is a $H^T_{i_j} \in \wp(T \backslash \{H_i\})$ such that $d^*(H_i; \bigwedge H^T_{i_j}, E) \leqslant 0$. Hence for all $H_i \in T$ and all $H^T_{i_j} \in \wp(T \backslash \{H_i\})$ it must hold that $p(H_i \mid H^T_{i_j} \wedge E) > p(H_i \mid H^T_{i_j})$. And thus, by Definition 3.1, E bootstrap-confirms T. (To see that, if for even a single H_i the value of $d^*(H_i; \bigwedge H^T_{i_j}, E)$ is lower than or equal to 0 for some set of auxiliaries, then the value of $B(T, E)$ must be strictly smaller than, and hence cannot be equal to, $(n)(2^{n-1}) - 1$, one only has to note that the range of d^* is the open interval $(-1, 1)$, and that -1, respectively, 1 are not within the range because for $p(H \mid H' \wedge E) - p(H \mid H')$ to obtain those values, it would have to hold that $p(H \mid H' \wedge E) = 0$ and at the same time that $p(H \mid H') = 1$, respectively, that $p(H \mid H' \wedge E) = 1$ and at the same time that $p(H \mid H') = 0$, neither of which combinations is possible.) □

APPENDIX C: PROOF OF THEOREM 5.1

THEOREM 5.1. For all T, T', and E, if $T \subset T'$ and E does not bootstrap-disconfirm T', then $B(T', E) \geqslant B(T, E)$; if in addition E bootstrap-confirms T', then $B(T', E) > B(T, E)$.

Proof. Let $T' = \{H_1, \ldots, H_n\}$. Without loss of generality, we can assume that $T = \{H_1, \ldots, H_m\}$ $(m < n)$. From Definition 4.2 it follows that:

$$
\begin{aligned}
B(T', E) &= \sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'}, E) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'}, E) \\
&\quad + \sum_{k=m+1}^{n} \sum_{l=1}^{2^{n-1}} d^*(H_k; \bigwedge H_{k_l}^{T'}, E).
\end{aligned}
$$

Note now that, if E does not bootstrap-disconfirm T', it must hold for all $H_i \in T$ and all $H_{i_j}^{T'} \in \wp(T' \backslash \{H_i\})$ that $d^*(H_i; \bigwedge H_{i_j}^{T'}, E) \geqslant 0$. Thus,

$$
\begin{aligned}
B(T, E) &= \sum_{i=1}^{m} \sum_{j=1}^{2^{m-1}} d^*(H_i; \bigwedge H_{i_j}^{T}, E) \\
&\leqslant \sum_{i=1}^{m} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'}, E).
\end{aligned}
$$

Hence, if $\sum_{k=m+1}^{n} \sum_{l=1}^{2^{n-1}} d^*(H_k; \bigwedge H_{k_l}^{T'}, E) \geqslant 0$, the following must hold:

$$
\begin{aligned}
B(T', E) &= \sum_{i=1}^{n} \sum_{j=1}^{2^{n-1}} d^*(H_i; \bigwedge H_{i_j}^{T'}, E) \\
&\geqslant \sum_{i=1}^{m} \sum_{j=1}^{2^{m-1}} d^*(H_i; \bigwedge H_{i_j}^{T}, E) = B(T, E).
\end{aligned} \tag{7}
$$

But it is easy to see that the condition is satisfied. For since T' is not bootstrap-disconfirmed by E, it must be the case for all $H_i \in T' - T$ that for all $H_{i_j}^{T'} \in \wp(T' \backslash \{H_i\})$, we have $d^*(H_i; \bigwedge H_{i_j}^{T'}, E) \geqslant 0$ (in virtue of Clause 2 of Definition 3.1). If E bootstrap-confirms T', then it must also be the case that for at least one $H_{i_j}^{T'} \in \wp(T' \backslash \{H_i\})$, we have $d^*(H_i; \bigwedge H_{i_j}^{T'}, E) > 0$ (in virtue of Clause 1 of Definition 3.1) so that $\sum_{k=m+1}^{n} \sum_{l=1}^{2^{n-1}} d^*(H_k; \bigwedge H_{k_l}^{T'}, E) > 0$, whence it follows that in Equation (7), '$\geqslant$' can be replaced by '$>$'. □

ACKNOWLEDGEMENTS

The authors are greatly indebted to Franz Huber, Theo Kuipers, Gert-Jan Lokhorst, Bert Postma, Jan-Willem Romeyn, Jos Uffink, Theo van Willigenburg, Christopher von Bülow, and two anonymous referees for *Synthese* for very valuable comments on an earlier version of this paper.

NOTES

[1] Subjective Bayesians may want to deny this. On their account, scientists are free – within the bounds of probability theory – in the probabilities they assign, and thus may also assign a probability to a hypothesis conditional upon the evidence alone (not conjoined to any auxiliaries, that is) that is greater than the unconditional probability assigned to the hypothesis (in which case the evidence confirms the hypothesis – see Section 3). However, the quantitative theory of bootstrap confirmation to be developed in this paper is neutral on the indispensability of auxiliaries in the sense that the measure of bootstrap support to be proposed also takes into account any support the evidence might give a hypothesis in isolation (in addition to the support the evidence may give the hypothesis relative to various auxiliaries).

[2] A point about notation: we use '$T = \{H_1, \dots, H_n\}$' to mean that T has axioms $H_1, \dots, H_n$, not that it has theorems $H_1, \dots, H_n$. And a point about terminology: by 'E confirms H with respect to T' (or 'E confirms H relative to T') we mean that E confirms H when the conjunction of hypotheses in T is taken as an auxiliary.

[3] Cf. e.g., Christensen (1997). Pondering various possibilities of how relative confirmation can provide 'real' confirmation, he conjectures that Glymour "(takes) certain complicated structures of interlocking relative confirmation to constitute real confirmation of a set of hypotheses" (p. 372). As may be clear, we think this conjecture is correct. Earman and Salmon (1992, 52ff) seem to interpret Glymour in the same way as we do.

[4] Or better, the phrases 'confirms with respect to' and 'disconfirms with respect to'; the exact understanding of 'with respect to' will depend on the interpretation of 'confirms'/'disconfirms'.

[5] Though Glymour does not note this, one must also assume that theories are *naturally* axiomatized (in some sense of "natural") lest the notion of bootstrap confirmation is one that is relative to a given axiomatization. To see this, just consider that since every finitely axiomatizable theory is axiomatizable by just one axiom – given any finite axiomatization, take the conjunction of the axioms – and since, given that we want to exclude macho-bootstrapping, a theory with only one axiom cannot be bootstrap-tested, without some notion of natural axiomatization it be may possible to claim of one and the same theory both that it is and that it is not bootstrap-confirmed by the evidence. *With* that notion, we can stipulate that a theory is bootstrap-confirmed by the evidence if its natural axiomatization is bootstrap-confirmed by the evidence. It seems that the notion of natural axiomatization has been around in the logical literature for some time. However, only recently an attempt has been made to explicate it; see Gemes (1993) (also his (1994, 1997); Schurz's (1991) theory of relevant deduction can also be thought of as such an attempt). We do not want to commit ourselves to Gemes' or any other explication; for present purposes the intuitive notion of natural axiomatization seems clear enough. Every theory to be presented by its axioms in this paper, both in the examples

and in the proofs of the theorems, is assumed to be naturally axiomatized in this intuitive sense.

[6] One may insist that the mere fact that a theory is (in a sense) presupposed in its own test is sufficient to make the procedure circular. Of course one may define circularity in any way one likes, but the crucial issue is whether the fact that a theory supplies auxiliaries for testing its own axioms is *vicious*. And we can only challenge anyone who holds that it is to point out why that is so.

[7] The claim is not that quantitative bootstrap confirmation captures *the* notion of coherence. Surely there is more than one sense in which propositions can be said to hang together (cf. Spohn 1999, 155). The bootstrap confirmation analysis of coherence and other formal analyses of coherence that have recently been proposed (e.g., in Spohn 1991, 1999; Olsson 1999; Shogenji 1999; Bovens and Hartmann 2003; Fitelson 2003; Douven's 2002a notion of a non-probabilistically self-undermining set clearly also is some sort of coherentist notion, as is Douven and Uffink's 2003 notion of a genuine preface case) are therefore best not thought of as being in competition with each other but rather as complementing each other, spelling out different but possibly equally valid concepts of coherence.

[8] Mainly thanks to the work of Garber (1983), Niiniluoto (1983), Eells (1985), and van Fraassen (1988). Whether the problem has been fully solved is still controversial, though; cf. Earman (1992, Ch. 5) and Howson (2000, 193ff) for differing opinions on this issue.

[9] Although it must be noted that Cartwright (1989) still makes extensive use of Glymour's theory in her attempt to show that causes can (sometimes) be obtained from probabilities; somewhat surprisingly, she does not address any of the criticisms to bootstrapping that were then already for some time being vigorously discussed in the journals. In this connection, it should also be noted that as late as 1997 bootstrapping was still referred to as one of "the two leading logical approaches to qualitative confirmation" (the other being hypothetico-deductivism; Christensen 1997, 370).

[10] By choosing a Bayesian approach we are following the mainstream in current analytic philosophy. However, it is noteworthy – as an anonymous referee reminded us of – that there exist other quantitative approaches to confirmation besides Bayesianism, such as Shafer's (1976) Dempster – Shafer belief functions, Zadeh's (1978) possibility measures, and Spohn's (1988) ranking functions (see Halpern 2003, Ch. 2 for an excellent overview of the different approaches to represent uncertainty).

[11] Note that the term '$p(H \mid A_1 \wedge \cdots \wedge A_n)$' should not be '$p(H)$', for then the increase in probability of H might well have nothing to do with the evidence. Glymour (1980a; 376), when briefly considering the prospects for placing the bootstrap idea in probabilistic terms, remarks that analyzing the relation 'E tests H relative to T' in terms of the probability of H conditional on $T \wedge E$ "meets with the difficulty that H can sometimes be used to test itself; that is, H is a consequence of T, so that the conditional probability in question becomes unity". But, clearly, now that macho-bootstrapping is prohibited, this problem can no longer arise.

[12] Here it may be helpful to note that if $p(H \mid H' \wedge E) > p(H \mid H')$, it must hold that both $0 < p(H' \wedge E) < 1$, and $p(H' \wedge E) \neq p(H')$ and hence also that $0 < p(E \mid H') < 1$.

[13] According to Duhem, Quine, and others, confirmation *generally* is three-place. Since we know of no air-tight argument showing that confirmation is *necessarily* three-place, we deem it best to at least formally leave open the possibility that

evidence confirms a hypothesis relative to the empty set, i.e., without the aid of any auxiliaries. Strictly speaking, the clauses of Definition 3.1 make no sense in case the subsets of T they refer to are empty. It should be obvious, however, that in that case '$p(H_i \mid T' \wedge E) > p(H_i \mid T')$' is to be read as $p(H_i \mid E) > p(H_i)$; similarly for '$p(H_i \mid T'' \wedge E) < p(H_i \mid T'')$' in the second clause.

[14] As a referee brought to our attention, it is insufficient to require that $T \cup \{E\} \nvdash \bot$ (as is done in Definition 2.1, and as we did in Definition 3.1 in an earlier version of the present paper) given that T may be consistent with E and yet it may hold that $p(T \wedge E) = 0$ (unless we assume all probability functions to be strict, which we don't); and of course one would not want to say that a theory can be confirmed in any sense by evidence conditional on which it has probability 0.

[15] Some might prefer a stricter definition of probabilistic bootstrap disconfirmation, like for instance one that requires that the evidence *substantially* decrease the probability of at least one axiom of the theory with respect to some other axioms of the theory. (Thanks to an anonymous referee for noting this.) We are here following standard Bayesian usage in identifying disconfirmation with any decrease in probability (however slight), but it would certainly seem to be of interest to investigate quantitative theories of bootstrap confirmation that employ stricter definitions of disconfirmation than Definition 3.2.

[16] The formal argument for this goes as follows: By the law of total probability, it holds that

$$p(E \mid H') = P(H \mid H')p(E \mid H \wedge H') + p(\neg H \mid H')p(E \mid \neg H \wedge H').$$

So, since $p(H \mid H') = 1 - p(\neg H \mid H')$, the higher $p(H \mid H')$ is, the smaller will $|p(E \mid H') - p(E \mid H \wedge H')|$ be, and thus (by some simple algebra) the smaller will $|p(H|H' \wedge E) - p(H|H')|$ be. And from this it follows that, on all plausible ways of measuring evidential impact (see Section, 4), and *ceteris paribus,* that impact will be lower the higher $p(H \mid H')$ is.

[17] Glymour's book in fact seems to suggest another desideratum. Glymour lays great emphasis on the confirmation-theoretic importance of the variety of evidence. However, since so far no one has been able to spell out in an even remotely precise fashion what variety of evidence amounts to, we here leave out sensitivity to variety of evidence as a desideratum for a quantitative account of bootstrapping.

[18] For a discussion and comparison of these measures, see Eells and Fitelson (2002). They make a strong case for d and l on the basis of symmetry considerations. Other measures to be found in the literature are Kemeny and Oppenheim's (1952) measure $(p(E|H) - p(E|\neg H))/(p(E|H) + p(E|\neg H))$, Nozick's (1981) measure $p(E|H) - p(E|\neg H)$, Gaifman's (1985, 20n6) measure $(1 - p(H))/(1 - p(H|E))$, Christensen's (1999) measure $d(H, E)/p(\neg E)$, Eells and Fitelson's (2000) measure $l(H, E)/p(\neg E)$, and Kuipers' (2000) measure $p(E|H)/p(E)$. Joyce (2004) makes the intriguing point that *many* of the proposed measures are worth having, given that they capture different, but equally important, notions of confirmation.

[19] As indicated in Note 13, we do not want our theory to formally preclude the possibility of evidence confirming a hypothesis relative to the empty set. It seems natural to take $d^*(H; \emptyset, E) = d(H, E)$ to measure the degree of confirmation bestowed by E on H relative to $\emptyset$; in the following we will write $d^*(H; \emptyset, E)$ simply as $d^*(H; E)$.

[20] In fact, Theorem 4.4 holds for any measure with range $(-1, 1)$, but the (log-)ratio measure is not among those.
[21] Thanks to an anonymous referee for pressing us to be clearer about which theorems hold for which measures.
[22] We are assuming, recall, that macho-bootstrapping is disallowed. Else there would (of course) be 2^n such sets.
[23] To reiterate a point previously made, we are not claiming to provide the one true measure of bootstrap confirmation here. In fact, it is easily seen that already the function given in the text allows of many variations (like, for example, variations with respect to the weights the desiderata are given relative to each other), some of which may well be worth exploring.
[24] That is, it holds for all $T' \subseteq T$ that $p(\bigwedge T') = \prod_{H \in T'} p(H)$.
[25] Here and elsewhere, probability functions will be specified without a proof that they *are* probability functions. It is nowadays easy to check that they are, however, by means of the function **InequalityInstance** of *MATHEMATICA* ®(Versions 4.1 and higher); see Fitelson (2001; 93–100) for an explanation of how to do this.
[26] Here and elsewhere, we write $i \neq j \neq k$ as short for $i \neq j,\ j \neq k, k \neq i$; similarly for similar expressions.
[27] Recall that we are assuming that it makes sense to speak of a theory's natural axiomatization, and also that every theory is given by its natural axiomatization (cf. Note 5). Without this assumption, it may occur that $B(T, E) \neq B(T^*, E)$ even though $T \equiv T^*$. Just consider $T' = \{H_1, H_2, H_3 \wedge H_4\}$, with the H_i's and all the probabilities as in the example. As another straightforward calculation shows, $B(T^*, E) \approx .572 \neq 2.412 \approx B(T, E)$. Intuitively that seems undesirable. However, T' is ruled out as a natural axiomatization by Gemes' (1993, 483) definition of a natural axiomatization (at least if we adopt the fourth clause he briefly discusses in Note 3 of his paper and which he presents as being optional), and we may assume that any other sensible definition of natural axiomatization will do the same.
[28] If we assume that also probabilistic (dis)confirmation always requires auxiliaries, this kind of case cannot even occur, of course. But in that case the threshold mentioned in Theorem 4.4 can be lowered to $(n)(2^{n-1}) - (n + 1)$.
[29] See Jeffrey (1983) for a lucid presentation of the theory's basic machinery; also Resnik (1987).
[30] It would take us too far afield here to discuss van Fraassen's reasons for rejecting this view on justification. For criticisms of these reasons, see, e.g., Kukla (1998); Douven (1999), (2002b), (2003a); Niiniluoto (1999); Psillos (1999).
[31] As, e.g., Gillies (2000) and Howson (2000, 2003) have argued (whether a defense of Bayesianism that does not appeal to utilities is possible, as Howson claims, is doubtful, however; cf. Douven (2003b). For one, it is entirely unclear whether an agent's risk-averseness should be reflected in her utility function; cf. Weirich (1986, 2001), Rabin (2000), and Hacking (2001, 100f) for discussion. As a further indication of the unclarity, see the divergent interpretations of utility proposed in, for instance, Hansson (1988), Hampton (1994), and Dreier (1996). Some believe that utilities are just theoretical posits that do not stand in need of any interpretation (this view seems to underlie Ramsey's (1926); Savage's (1954), work in decision theory and is still not uncommon, as Rabin (2000) reminds us). But aside from the difficulties generally related to instrumentalist interpretations of theoretical terms, on an instrumentalist reading, decision theory, and hence also cognitive decision theory, can only be used

as an explanatory device, and not as a guide to decision making (see, e.g., Satz and Ferejohn 1994). So, in particular, cognitive decision theory could on that reading not inform us about when it is rational for us to believe a particular hypothesis or theory; at most it could be used post factum to explain why someone preferred to accept one rather than another hypothesis or theory.

[32] See Goosens (1976) for a more systematic critique of the concept of cognitive utility.

[33] A way in which bootstrap confirmation could play a role in determining the justificational status of a theory that respects principle (P) is to assign a justificatory role to bootstrap support only after probabilistic considerations have been taken account of. This is, for instance, what the following principle does:

(P*) If (i) $p(T \mid E) > p(T' \mid E)$ *or* (ii) $p(T \mid E) = p(T'|E)$ and E bootstrap-confirms T but not T' *or* (iii) $p(T|E) = p(T'E)$ and E bootstrap-confirms both T and T' but $B(T, E) > B(T', E)$, then we are more justified in believing T than we are in believing T'.

Again another response to the puzzle would be to claim that justification is to be evaluated at the level of single hypotheses, and not at the level of theories. It is perfectly compatible with one theory as a whole being more probable than another theory as a whole that the probability of any of the axioms of the latter exceeds the probability of each of the axioms of the former. (This would be much along the lines of Merricks' 1995 response to Klein and Warfield's 1994 claim that coherence is not generally truth-conducive. The discussion concerning the truth-conduciveness of coherence has in fact many parallels with the discussion in the present section. Given the close conceptual ties between coherence and bootstrap support pointed to in Section 2, this should come as no surprise.)

REFERENCES

BonJour, L.: 1985, *The Structure of Empirical Knowledge*, Harvard University Press, Cambridge MA.

Bovens, L. and S. Hartmann: 2003, *Bayesian Epistemology*, Oxford University Press, Oxford.

Cartwright, N.: 1989, *Nature's Capacities and their Measurement*, Oxford University Press, Oxford.

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.

Christensen, D.: 1997, 'What is Relative Confirmation?', *Noûs* **31**, 370–384.

Christensen, D.: 1999, 'Measuring Confirmation', *Journal of Philosophy* **96**, 437–461.

Douven, I.: 1999, 'Inference to the Best Explanation Made Coherent', *Philosophy of Science* (Proceedings) **66**, S424–S435.

Douven, I.: 2002a, 'A New Solution to the Paradoxes of Rational Acceptability', *British Journal for the Philosophy of Science* **53**, 391–410.

Douven, I.: 2002b, 'Testing Inference to the Best Explanation', *Synthese* **130**, 355–377.

Douven, I.: 2003a, 'Empirical Equivalence, Explanatory Force, and the Inference to the Best Theory', in R. Festa, A. Aliseda, and J. Peijnenburg (eds.), *Logics of Scientific Cognition: Essays in Debate with Theo Kuipers*, Rodopi, Atlanta GA, in press.

Douven, I.: 2003b, 'On Bayesian Logic', in M.C. Galavotti (ed.) *Observation and Experiment in the Natural and Social Sciences*, Kluwer, Dordrecht, pp. 321–326.

Douven, I.: 2004, 'Basic Beliefs, Coherence, and Bootstrap Support', in S. Roeser, R. Rood, and R. van Woudenberg (eds.), *The Epistemology of Basic Belief* (publisher to be determined), in press.

Douven, I. and J. Uffink: 2003, 'The Preface Paradox Revisited', *Erkenntnis* **59**, 389–420.

Dreier, J.: 1996, 'Rational Preference: Decision Theory as a Theory of Practical Rationality', *Theory and Decision* **40**, 249–276.

Duhem, P.: 1906/1954, *The Aim and Structure of Physical Theory* (translated by P. Wiener), Princeton University Press, Princeton NJ.

Earman, J. (ed.): 1983, *Testing Scientific Theories*, University of Minnesota Press, Minneapolis.

Earman, J.: 1992, *Bayes or Bust?*, MIT Press, Cambridge MA.

Earman, J. and C. Glymour: 1988, 'What Revisions Does Bootstrap Testing Need?', *Philosophy of Science* **55**, 260–264.

Earman, J. and W. Salmon: 1992, 'The Confirmation of Scientific Hypotheses', in M. Salmon (ed.), *Introduction to the Philosophy of Science*, Prentice Hall, Upper Saddle River NJ, pp. 42–103.

Edidin, A.: 1983, 'Bootstrapping Without Bootstraps', in Earman (ed.), pp. 43–54.

Eells, E.: 1985, 'Problems of Old Evidence', *Pacific Philosophical Quarterly* **66**, 283–302.

Eells, E. and B. Fitelson: 2000, 'Measuring Confirmation and Evidence', *Journal of Philosophy* **97**, 663–672.

Eells, E. and B. Fitelson: 2002, 'Symmetries and Asymmetries in Evidential Support', *Philosophical Studies* **107**, 129–142.

Fitelson, B.: 2001, *Studies in Bayesian Confirmation Theory*, PhD. dissertation, University of Wisconsin–Madison.

Fitelson, B.: 2003, 'A Probabilistic Theory of Coherence', *Analysis* **63**, 194–199.

Gaifman, H.: 1985, 'On Inductive Support and Some Recent Tricks', *Erkenntnis* **22**, 5–21.

Garber, D.: 1983, 'Old Evidence and Logical Omniscience in Bayesian Confirmation Theory', in Earman (ed.), pp. 99–131.

Gemes, K.: 1993, 'Hypothetico-Deductivism, Content, and the Natural Axiomatization of Theories', *Philosophy of Science* **60**, 477–487.

Gemes, K.: 1994, 'A New Theory of Content I: Basic Content', *Journal of Philosophical Logic* **23**, 596–620.

Gemes, K.: 1997, 'A New Theory of Content II: Model Theory and Some Alternatives', *Journal of Philosophical Logic* **26**, 449–476.

Gillies, D.: 2000, *Philosophical Theories of Probability*, Routledge, London.

Glymour, C.: 1980a, *Theory and Evidence*, Princeton University Press, Princeton NJ.

Glymour, C.: 1980b, 'Bootstraps and Probabilities', *Journal of Philosophy* **67**, 691–699.

Goosens, W.: 1976, ‘A Critique of Epistemic Utilities’, in R. Bogdan (ed.), *Local Induction,* Reidel, Dordrecht, pp. 93–113.
Hacking, I.: 2001, *An Introduction to Probability and Inductive Logic*, Cambridge University Press, Cambridge.
Halpern, J.: 2003, *Reasoning about Uncertainty*, MIT Press, Cambridge MA.
Hampton, J.: 1994, ‘The Failure of Expected-Utility Theory as a Theory of Reason’, *Economics and Philosophy* **10**, 195–242.
Hansson, B.: 1988, ‘Risk Aversion as a Problem of Conjoint Measurement’ in P. Gärdenfors and N.-E. Sahlin (eds.), *Decision, Probability, and Utility,* Cambridge University Press, Cambridge, pp. 136–158.
Howson, C.: 2000, *Hume’s Problem: Induction and the Justification of Belief*, Clarendon Press, Oxford.
Howson, C.: 2003, ‘Bayesian Evidence’, in M.C. Galavotti (ed.), *Observation and Experiment in the Natural and Social Sciences*, Kluwer, Dordrecht, pp. 301–320.
Jeffrey, R.: 1983, *The Logic of Decision* (2nd ed.), University of Chicago Press, Chicago IL.
Joyce, J.: 2004, ‘On the Plurality of Probabilist Measures of Evidential Relevance’, manuscript.
Kaplan, M.: 1981a, ‘Rational Acceptance’, *Philosophical Studies* **40**, 129–145.
Kaplan, M.: 1981b, ‘A Bayesian Theory of Rational Acceptance’, *Journal of Philosophy* **78**, 305–330.
Kemeny, J. and P. Oppenheim: 1952, ‘Degrees of Factual Support’, *Philosophy of Science* **19**, 307–324.
Klein, P. and T. Warfield: 1994, ‘What Price Coherence?’, *Analysis* **54**, 129–132.
Kuipers, T.: 2000, *From Instrumentalism to Constructive Realism*, Kluwer, Dordrecht.
Kukla, A.: 1998, *Studies in Scientific Realism*, Oxford University Press, Oxford.
Lehrer, K.: 1990, *Theory of Knowledge* (2nd ed.), Routledge, London.
Levi, I.: 1967, *Gambling with Truth*, MIT Press, Cambridge MA.
Maher, P.: 1993, *Betting on Theories*, Cambridge University Press, Cambridge.
Merricks, T.: 1995, ‘On Behalf of the Coherentist’, *Analysis* **55**, 306–309.
Niiniluoto, I.: 1983, ‘Novel Facts and Bayesianism’, *British Journal for the Philosophy of Science* **34**, 375–379.
Niiniluoto, I.: 1998, ‘Verisimilitude: The Third Period’, *British Journal for the Philosophy of Science* **49**, 1–29.
Niiniluoto, I.: 1999, *Critical Scientific Realism*, Clarendon Press, Oxford.
Nozick, R.: 1981, *Philosophical Explanations*, Harvard University Press, Cambridge MA.
Olsson, E. J.: 1999, ‘ “Cohering With” ’, *Erkenntnis* **50**, 273–291.
Psillos, S.: 1999, *Scientific Realism: How Science Tracks Truth,* Routledge, London.
Quine, W. V.: 1953, ‘Two Dogmas of Empiricism’, in his *From a Logical Point of View,* Harvard University Press, Cambridge MA, pp. 20–46.
Rabin, M.: 2000, ‘Risk Aversion and Expected-Utility Theory: A Calibration Theorem’, *Econometrica* **68**, 1281–1292.
Ramsey, F. P.: 1926, ‘Truth and Probability’, in his *The Foundations of Mathematics,* Routledge and Kegan Paul, 1931, London, pp. 156–198.
Resnik, M.: 1987, *Choices,* University of Minnesota Press, Minneapolis.
Satz, D. and J. Ferejohn: 1994, ‘Rational Choice and Social Theory’, *Journal of Philosophy* **91**, 71–87.

Savage, L.: 1954, *The Foundations of Statistics,* Wiley, New York.
Schurz, G.: 1991, 'Relevant Deduction: From Solving Paradoxes Towards a General Theory', *Erkenntnis* **35**, 391–437.
Shafer, G.: 1976, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton NJ.
Shogenji, T.: 1999, 'Is Coherence Truth Conducive?', *Analysis* **59**, 338–345.
Spohn, W.: 1988, 'Ordinal Conditional Functions: A Dynamic Theory of Epistemic States', in W. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics* (Vol. 2), Reidel, Dordrecht, pp. 105–134.
Spohn, W.: 1991, 'A Reason for Explanation: Explanations Provide Stable Reasons', in W. Spohn, B.C. van Fraassen, and B. Skyrms (eds.), *Existence and Explanation*, Kluwer, Dordrecht, pp. 165–196.
Spohn, W.: 1999, 'Two Principles of Coherence', *Erkenntnis* **50**, 155–175.
van Fraassen, B. C.: 1980, *The Scientific Image*, Clarendon Press, Oxford.
van Fraassen, B. C.: 1983a, 'Theory Comparison and Relevant Evidence', in Earman (ed.), pp. 27–42.
van Fraassen, B. C.: 1983b, 'Glymour on Evidence and Explanation', in Earman (ed.), pp. 165–176.
van Fraassen, B. C.: 1988, 'The Problem of Old Evidence', in D. Austin (ed.), *Philosophical Analysis,* Kluwer, Dordrecht, pp. 153–165.
Weirich, P.: 1986, 'Expected Utility and Risk', *British Journal for the Philosophy of Science* **37**, 419–442.
Weirich, P.: 2001, 'Risk's Place in Decision Rules', *Synthese* **126**, 427–441.
Zadeh, L. A.: 1978, 'Fuzzy Sets as a Basis for a Theory of Possibility', *Fuzzy Sets and Systems* **1**, 3–28.

Department of Philosophy
Erasmus University
Rotterdam
E-mails: douven@fwb.eur.nl, meijs@fwb.eur.nl