THE JOURNAL OF PHILOSOPHY
VOLUME LXXVII, NO. 11, NOVEMBER 1980

BOOTSTRAPS AND PROBABILITIES *

THE mechanism of bootstrapping is roughly this. Some quantities of a theory are measured in an experiment or observation, whereas other quantities are not. Even though it contains unmeasured quantities, a hypothesis, even one of the theory's own, may be tested and confirmed with respect to the theory provided, first, that values for all quantities occurring in the hypothesis can be computed from the measurements by means of logical consequences of the theory; second, that such computed values confirm the hypothesis; and, third, that there are alternative possible values of the measured quantities such that, with these alternative values, the same procedure for determining values for quantities occurring in the hypothesis results in a set of values that *disconfirms* the hypothesis.

So understood, bootstrapping depends upon some account of the confirmation relations among sentences making reference to a common body of quantities. In *Theory and Evidence* [1] I attached bootstraps to Hempel's confirmation theory, slightly modified, and claimed that the connection was inessential, since one could quite as well attach bootstraps to probabilistic theories of confirmation. This essay describes at least one way of carrying out the attachment of bootstraps to probabilities. More generally, I shall discuss some relations between the approach to confirmation developed in *Theory and Evidence* and probabilistic ways of assessing and comparing theories.

Some confirmation theorists think of probability relations as applying locally, within particular restricted contexts determined

* To be presented in an APA symposium on the Philosophy of Science, December 29, 1980. Paul Horwich will comment; see this JOURNAL, this issue, 699–702.
[1] Princeton, N.J.: University Press, 1980.

0022-362X/80/7711/0691$00.90

by an experimental or theoretical setting. Bootstraps can be attached to local probabilities directly enough. I will sketch, informally, an abstract framework for doing so, a framework which can be interpreted in accord with several different conceptions of probability. Let L be a language, Q a set of quantity terms in L, S the set of well-formed claims of L involving Q. We can assume that L contains logical principles and, possibly, mathematical principles as well, which are to hold in every possible situation. Semantically, suppose there is a set Q^* of *quantities* such that a quantity q^* of type n maps members of a subset of any set of ordered n-tuples of individuals into the real numbers R. Let Q^* be closed under the usual algebraic operations. A *possible situation* w^* is a pair consisting of a set of individuals (or domain) and a set of quantities mapping members of the domain, or members of Cartesian powers of the domain, into the reals. The individuals or quantities that are elements of a set that is a member of a situation will be said to *belong* to that situation. W^* denotes the set of all possible situations, and, for an ordered n-tuple I_n of individuals, $q^*(I_n, w^*)$ will denote the value of q^* for I_n in situation w^*. Dom(q^*) will denote the set of all situations to which q^* belongs. I will indicate that sentence s holds in situation w^* by $w^* \vdash s$, and assume that it is clear enough what that will mean.

Now for probabilities. Assume, for appropriate subsets Z of Q^* a probability measure prob$_z$ on the set $\mathrm{Dom}(Z) =_{\mathrm{df}} \bigcap_{q^* \epsilon Z} \mathrm{Dom}(q^*)$. Undoubtedly one wants some restrictions on the probability measures associated with different subsets of Q^*, but I will impose none now. For a sentence S and a set Z of quantities, define $\mathrm{prob}_z(s) =_{\mathrm{df}} \mathrm{prob}_z\{w^* \epsilon \mathrm{Dom}(Z) \mid w^* \vdash s\}$. Conditional probabilities may be defined in the usual way.

Now let Z be a set of quantities, E a collection of sentences, all simultaneously holding in some situation, asserting values for the quantities in Z, T a consistent theory in L, h some hypothesis in L, and $Q(h)$ the set of quantities for which terms occur in h. Then say E confirms h with respect to T if there is a subtheory T_0 of T such that:

i. for each quantity $q^* \epsilon Q(h)$ there is an n-tuple of individuals I_n such that $\{r \epsilon R \mid \exists w^*$ for which $q^*(I_n, w^*) = r$ and $w^* \vdash T_0 \cup E\}$ is nonempty and is a proper subset of

$$\{r \epsilon R \mid \exists w^* \text{ for which } q^*(I_n, w^*) = r\}$$

ii. $\mathrm{prob}_{Q(T_0 \cup E)}(h, T_0 \cup E) > \max\{\mathrm{prob}_{Q(T_0 \cup E)}(h),$
$\mathrm{prob}_{Q(T_0 \cup E)}(h, T_0)\}$

iii. there exists a set E' of sentences asserting values for quantities in Z, and there exists a situation in which the sentences of E' all hold, and E' and T_0 determine values for quantities $Q(h)$ as in clause i, and $\text{prob}_{Q(T_0 \cup E)}(h, T_0 \cup E') < \min\{\text{prob}_{Q(T_0 \cup E)}(h), \text{prob}_{Q(T_0 \cup E)}(h, T_0)\}$

An analogous condition, with the inequalities between probabilities reversed in clauses ii and iii will characterize disconfirmation.[2]

Locally, everything works like probabilities. Hypothesis h is confirmed with respect to T if its denial is disconfirmed, and conversely; positive instances don't necessarily confirm, and the usual probabilistic escapes from the paradoxes of confirmation are all available. There is no difficulty, for example, in combining bootstrapping with a system isomorphic to Carnapian logical probabilities: just take quantities to be the characteristic functions of the extensions of predicates, let situations be determined by state descriptions, and let prob_z be determined by a logical measure function.[3] Again, the probabilities in question can be personal, and one can imagine a local (one might say "fragmented") personalist Bayesian using bootstraps.

Whatever the differences between bootstrapping and probabilistic confirmation, they are not essentially differences of a local kind. Perhaps real differences will emerge on a larger scale. Faced with a range of evidence more or less agreed upon and a collection of contending theories, all logically consistent with the evidence, which theory is to be given the greater credence; which is more worthy of belief? Global probability theorists give a precise recipe: prefer the theory that has, on the total evidence, the highest probability. The only thing obscure is how to determine the probabilities. The

[2] This account is both incomplete and tentative; it needs to be supplemented, for example, with a specification of the status of logical consequences of the evidence. Again, I am uncertain whether the inequality in iii should be strict. The account could (and probably should) be revised, in a more realistic way, by supposing that one possible (small) world contains many situations to which the same quantities belong, so that a situation, as described above, corresponds to a set of possible worlds. The probability measure associated with a set Z of quantities is then a measure on the space of all possible worlds having situations belonging to Dom(Z), and situations correspond to events of such a space. Letting W range over possible worlds, the probability of a sentence s then becomes the unfortunately complicated $\text{prob}_z(s) =_{\text{df}}$

$$\text{prob}_z\{W \mid \exists w^*(w^* \in W \,\&\, w^* \in \text{Dom}(Z)) \,\&\, \forall u^*((u^* \in W \,\&\, u^* \in \text{Dom}(Z)) \rightarrow u^* \vdash s)\}$$

where w^* and u^* range over situations. Thus some hypotheses of a theory may be tested in one situation, others in another.

[3] Indeed, the Carnapian measure that assigns equal probability to all structure descriptions provides a case illustrating the failure of a natural restriction on the probability measures associated with different subsets of Q^*—namely that the probability of a sentence s be unchanged between measures based on $Q(s)$ and measures based on sets of quantities properly including $Q(s)$.

recommendation of *Theory and Evidence* is less precise: compare the theories by whether or not, for each theory, hypotheses jointly sufficient to entail the theory have been individually confirmed with respect to that very theory, by how strong these confirmations are, by the variety, number, and independence of the tests of such hypotheses. No measure or even firm ordering of these criteria is given generally; only for special cases (e.g., systems of linear equations) is a determinate set of conditions for comparison obtained.[4]

The pertinent question is this: is the bootstrap idea, and the recommendations it yields for comparing theories, merely an incomplete, nuts-and-bolts version of global Bayesian confirmation theory? There are reasons to think so. Where precise comparisons are possible, as with certain systems of equations, bootstrap comparisons agree very closely with the comparisons that one obtains (without regard to prior probabilities) from Roger Rosencrantz's objective Bayesian account of confirmation.[5] But there are differences, and some of them seem to me to be important. For example, on Rosencrantz's account (as on any Bayesian account) if a system H of equations is confirmed by a body of evidence E, so too will any deoccamization of H be confirmed by E, where a *deoccamization* of H is obtained by substituting for every occurrence of some quantity term in H an algebraic combination of two or more distinct quantities not otherwise occurring in H. Further, the support for H—which Rosencrantz identifies with the average likelihood of E on H—will be the same as for the deoccamization of H. So also if we add to the deoccamized version of H additional equations involving the new quantities but insufficient to determine their values from values of measurable quantities. And again if we add to H some irrelevant equation. E confirms all these variants in the Bayesian way if it confirms H, although the confirmation need not be to the same degree, and E provides the same support (in Rosencrantz's sense) for each of these variants as it does for H. Whether, on the basis of the evidence, H is to have greater probability than any of these alternatives is determined entirely by the distribution of prior probabilities. In contrast, bootstrap comparisons yield the result that E provides better evidence for H than for any of the alternatives. But this difference is not decisive for the question at issue, since global probabilists might in principle assimilate these

[4] This vagueness is certainly in part the result of analytic incompleteness, but it may also in part be due to an intrinsic vagueness in the subject; for "what confirms what" is vague just as "what is not but would be the case" is vague.

[5] *Inference, Method and Decision* (Boston: Reidel, 1977).

discriminations in either of two ways. On the one hand, it can be supposed that these bootstrap discriminations are merely clumsy, mechanical ways of obtaining discriminations that are really founded on differences in prior probabilities. On the other hand, if one recognizes, for example, the widespread preference for occamized over deoccamized theories, but doubts that it can plausibly be accounted for by prior probability distributions, one can suppose that the preference is founded not on probability but on something that combines with probability to determine the relative merits of theories. A favorite suggestion is to introduce epistemic utilities, and to suppose that a rational preference ordering of theories is the ordering given by expected epistemic utility. I am not quite sure how to specify prior distributions that would do the job, and even less clear how to construct reasonable epistemic utility scales that reflect the structural virtues and liabilities of theories, but I have no reason to believe that either could not be done.

There are further reasons to suppose that bootstrapping should be understood as simply part of the nuts and bolts of global Bayesian confirmation. Suppose for example, that the probability measures prob_Z can all be obtained by conditionalizing a common probability prob_ϕ on the space of all situations. Then we can think of prob_ϕ as giving the Bayesian priors for any sentences in the language, and with evidence E we conditionalize to the subset of $\text{Dom}(Q(E \cup T_0))$ in which E holds. Then E bootstrap-confirms h with respect to T only if there is a subtheory T_0 of T such that $T_0 \cup E$ is positively relevant to h. What this shows is that, if local probabilities are not too fragmented, bootstrap confirmation can be seen as a species of global Bayesian confirmation.[6]

That is short of showing that bootstrap comparisons of theories result from Bayesian comparisons of theories. Bootstrapping suggests that evidence E may fail to bear on hypothesis K all by itself, and likewise on hypothesis L, but that E may confirm K with respect to L, and also E may confirm L with respect to K. It is then claimed that, in these circumstances, E provides grounds for believing K & L with respect to K & L, or, put more naturally, that E confirms K & L. Bayesians simply compare the prior probability of K & L with its posterior probability. It turns out, however, that, in the circumstances given, if E bootstrap-confirms K with respect to K & L and if E bootstrap-confirms L with respect to K & L, then (provided only that $\text{prob}_{Q(E \cup K)}$ and $\text{prob}_{Q(E \cup L)}$ are conditional

[6] Essentially this point is made about nonprobabilistic versions of bootstrapping by A. Edidin, "Glymour on Confirmation," *Philosophy of Science*, forthcoming.

probabilities from a common probability measure), E must provide a Bayesian confirmation of K & L.[7] Thus the chief bootstrap principle for putting confirmations together to assess theories turns out to be a Bayesian principle. It begins to look very much as though bootstrapping succeeds only at getting at global Bayesian principles in a different way.

Even so, it is neither my expectation nor my hope that bootstrap confirmation will reduce to Bayesian confirmation. It is not my expectation because I am inclined to doubt that, in many situations, we have either objective probabilities or subjective degrees of belief of a sufficiently global kind upon which we can rely to relate evidence to theory. When theories are proposed for novel subject matters (as in some contemporary social science) or when new theories are seriously considered which deny previously accepted fundamental relationships (as, say, with the introduction of general relativity) we may be at a loss for probabilities connecting evidence to theory. It is in such cases that we find bootstrap arguments used explicitly to establish relevance. What we know and argue with are fragmented probabilities and structural relations—bootstrap relations among others—and more global probabilities must follow along. So my expectation is that the mechanism of bootstrapping can function where global probabilities do not. That is perhaps only a puerile independence, which matures to probability.

My hope that bootstrapping is not just global probability in disguise derives from a deeper point. Probabilist accounts of the grounds for preferring theories tread a delicate path between the demand for informative, explanatory, simple theories, on the one hand, and the demand for credibility, on the other; devices that tend to show that the two demands can jointly be met are especially welcome. That is, for example, part of what makes Rosencrantz's account of support attractive (whether or not one agrees with it); for its intent is to show that intuitively simpler theories are better supported than are more complex theories, by evidence in agree-

[7] The circumstances amount to the following Bayesian assumptions:

i. $\text{prob}(L, K \,\&\, E) > \text{prob}(L, K)$
ii. $\text{prob}(K, E) = \text{prob}(K)$
iii. $\text{prob}(K, L \,\&\, E) > \text{prob}(K, L)$
iv. $\text{prob}(L, E) = \text{prob}(L)$

It is immediate that $\text{prob}(K \,\&\, L, E) > \text{prob}(K \,\&\, L)$, as follows:

$$\begin{aligned}(\text{prob}(L \,\&\, K, E))^2 &= (\text{prob}(L, K \,\&\, E)\ \text{prob}(K, E))(\text{prob}(K, L \,\&\, E)\ \text{prob}(L, E))\\ &> (\text{prob}(L, K)\ \text{prob}(K, E))(\text{prob}(K, L)\ \text{prob}(L, E))\\ &= (\text{prob}(L, K)\ \text{prob}(K))(\text{prob}(K, L)\ \text{prob}(L))\\ &= (\text{prob}(L \,\&\, K))^2\end{aligned}$$

ment with both. Global probabilists are pulled in either of two ways: toward making explanatory power, informativeness, and simplicity confirmatory virtues, and thus reasons for belief, or else toward making them virtues of another kind, dubbed "epistemic," and reasons not for belief but only for preference. But whichever path is taken, the probabilist cannot justify a full-blooded realism. Consider two belief states. In the first, one believes the atomic theory, including its claims about the various kinds of atoms, their relative and absolute weights, and so on. In the second one believes only those consequences of the theory which are about observable features of macroscopic systems, and one is agnostic about the rest of the theory. The distinction is vague, but that scarcely matters. Whichever way he is pulled regarding explanation, simplicity, informativeness, and the like, the global probabilist cannot judge the first state of mind to be better founded on the evidence than is the second, and preferable to it for that reason. For the probability of the consequences must always be at least as high as that from which they are consequent. Appeal to epistemic utilities can of course make the theory preferable to its collection of consequences about middle-sized dry goods, but it strikes me as a mistake to regard such preferences as preferences for some states of *belief* over others: they are, at best, preferences for use, or for acceptance.

This will not be much bother to those who are not realists, or who are not tempted ever to compare states of belief. They will properly say that the business of confirmation theory and of an account of rational belief is not to compare such bizarre doxastic states, but only to give the conditions for the rational distribution and change of degrees of belief. I hope for something more, for I believe we believe, and I think it would not be rational, knowing of the atomic theory and of the evidence for it, to believe in the theory's empirical adequacy but not in the theory. And likewise with many other cases. My hope is that bootstrapping and other devices will provide a canon by which to justify and regiment my prejudice. The reasons for the hope are straightforward to describe.

Bootstrap confirmation relations turn out to be Bayesian confirmation relations, but not conversely. Bayesian confirmation is open-handed and generous; bootstrap confirmation is meaner. This meanness gives to theories, in some circumstances, a special advantage over anything that might be called their observational consequences, or their consequences for measurable quantities. The structural relations determined by bootstrapping go some distance, however short, toward providing grounds for preferring belief in theories to belief in their observational consequences *alone*. For example, it has been shown, using a Hempelian version of bootstrapping, that each of the axioms of a theory can be confirmed with respect to a theory by a body of evidence, even though the evidence does not confirm each of the "observational" consequences of the theory, or any set of axioms therefore, with respect to this same body of "observational" consequences. I expect that similar examples can be constructed for the probabilistic version of bootstrapping described in this essay.[8] What is not clear is how common or generalizable such examples are.[9]

A related phenomenon concerns inductive systematization: can a theory in combination with evidence provide better confirmation of some relation about measurables than is provided by the observational consequences of the theory in combination with that same evidence? Can it do so when the probability of the theory itself is high? The answer to the first question is controversial for probabilistic, Bayesian confirmation theories; the answer to the second appears to be definitely negative for such accounts of confirmation. But, exactly because it is more stingy with confirmation, bootstrap relations can in certain contexts automatically provide inductive systematization of these kinds. I give an elementary example:[10]

Let the sentence S of L assert that two quantities are equal in value, or that the value of one quantity is not greater than that of another; let each quantity have at most one value in each situation, so the possible situations are in effect given by any set of real values for some or all of the quantities. Let A, B, C be measurable quantities, q an unmeasurable quantity, and let $f_{[,]}(X)$ be the quantity that is the characteristic function of the quantity X on the closed interval $[,]$. Let $0 < \text{prob}_{\{A,B,C,q\}}(Af_{[0,1]}(A) \leq C \leq A) \ll 1$. Consider the theory

i. $Bf_{[0,1]}(B) \leq q \leq B$

ii. $C = f_{[\frac{1}{4},1]}(q)$

Let the evidence be $\{A = 1, B = 1\}$. Then the hypothesis $Af_{[0,1]}(A) \leq C \leq A$ is confirmed (and, indeed, has probability 1) with respect

[8] See my *Theory and Evidence*, *op. cit*, pp. 166/7.

[9] For certain kinds of linear theories it can be shown that the axioms of a theory are never better tested, in the bootstrap way, than are the axioms of the set of measurable consequences of the theory. Compare my "The Good Theories Do," *Proceedings of the* 1979 *ETS Colloquium on Construct Validity*, to appear.

[10] This example was suggested to me by a similar example given by Bas van Fraassen.

to the theory, but not with respect to those consequences of the theory which state only relations among measurable quantities, since those consequences do not, in combination with the evidence, constrain the value of C.[11]

I scarcely think that all good confirmation is bootstrap-like, or that more generous confirmation relations never apply. I think, for example, that the provision of some kinds of explanations provide grounds for belief in theories, and I think that it is quite all right to confirm deoccamized theories when we have independent evidence that the "redundant" quantities are real enough. And so on. The demand for bootstrap confirmation is, I am sure, at best prima facie and defeasible. My hope remains that prima facie requirement will, in combination with other considerations—such as the structure of explanations—help untangle the web of reasoning by which we warrant our beliefs.[12]

CLARK GLYMOUR

University of Pittsburgh

THE DISPENSABILITY OF BOOTSTRAP CONDITIONS *

My comments on Clark Glymour's paper will be divided into three parts: (I) discussion of his general strategy (Is his proposal a substantial revision of probabilism, or does it affect only certain pe-

[11] It could be objected that the theory used in this example is not confirmed by the evidence, let alone highly probable. That defect is easily remedied. Consider the version of probabilistic bootstrapping described in fn 2. Let the data from one situation be $\{C = 1, B = 1\}$ with no value stated for A. With appropriate probabilities, each axiom of the theory (that is, i and ii) is confirmed with respect to the other by the evidence, but the hypothesis is not tested. The probability of the theory can be quite high on such evidence. Let the data from a second situation be as in the text. Then, on the total evidence from the two situations, the theory is confirmed and its probability is high, the probability of the hypothesis with respect to the theory is high and has been increased by the evidence, but the hypothesis is not confirmed with respect to the measurable consequences of the theory.

[12] Having devoted this essay to attempting to correct an error of tactics in *Theory and Evidence*—the development of the bootstrap idea without connection to probability—I want also to correct an error of fact of some consequence. On pages 14–21 of that book I incorrectly attribute to Wesley Salmon, on the basis of his "Verifiability and Logic" in P. Feyerabend, ed., *Mind, Matter and Method* (Minneapolis: University of Minnesota Press, 1966), the proposal that a sentence be taken as true if and only if all its first-order observational consequences are true. Professor Salmon does not hold such a view, and did not propose it in the essay referred to.

* Abstract of a paper to be presented in an APA symposium, commenting on a paper by Clark on Glymour; see this JOURNAL, this issue, 691–699.

0022-362X/80/7711/0699$00.50