DISCUSSION:

WHAT REVISIONS DOES BOOTSTRAP TESTING NEED? A REPLY*

JOHN EARMAN

Department of History and Philosophy of Science
University of Pittsburgh

CLARK GLYMOUR

Department of Philosophy
Carnegie-Mellon University
and
Department of History and Philosophy of Science
University of Pittsburgh

A familiar fact is that we use our background knowledge and some of our hypotheses in arguing for or against other hypotheses. But the structure of such arguments is difficult to capture. Glymour's *Theory and Evidence* contains an attempt to give a formal characterization of the relationships involved for theories with a first-order formalization, and Glymour called the process of testing a hypothesis in the theory by using that same hypothesis, or others in the same theory, "bootstrap testing". However, the original conditions for bootstrap testing were not strong enough, as shown by a series of counterexamples by Christensen (1983). Glymour (1983) proposed to strengthen the account by adding an extra requirement, condition (R). Zytkow (1986) claimed that (R) is too strong. As an alternative to Glymour (1983), Zytkow provided his own version of bootstrapping which, he claimed, avoided both the Christensen counterexamples and the objection to (R). We will show that Zytkow's version of bootstrapping is unacceptable and that his worry about condition (R) is unfounded. However, we will also argue that Zytkow's remarks point to what is, perhaps, a more satisfying version of (R).

1. Zytkow's Revision. On Zytkow's version of bootstrapping, E confirms H with respect to T just in case:

(Z1) E, H and T are jointly consistent
(Z2) There is a finite set of values Q of quantities occurring essen-

*Received December 1986.

Philosophy of Science, 55 (1988) pp. 260–264.

tially in H such that (a) Q is a consequence of $T \cup E$ and (b) H is satisfied by Q according to the satisfaction condition S

and

(Z3) a modification E' of E is logically possible and a set of values Q' for the quantities occurring essentially in H such that (a) Q' is a consequence of $T \cup E'$ (b) it is not the case that H is satisfied by Q' according to the satisfaction condition S, and (c) for every subtheory W in T such that W does not contain any consequence of H, it is not the case that W is contradicted by Q' according to the satisfaction condition.

Condition (Z3c) was designed to eliminate Christensen's counterexamples. It fails to do so. Note that for any H and any subtheory W of T, W contains, that is, entails, a consequence of H, namely $H \vee W$ and all of its consequences. Hence (Z3c) is vacuous since there are no theories of the requisite kind. If we assume that Zytkow meant (as he says at an earlier point in his paper) that W does not contain any nontautological consequence of H, the condition is still nearly vacuous, because $H \vee W$ is a consequence of H and of W, and $H \vee W$ is a tautology only if either H or W is a tautology or H and W are inconsistent. If H and W are inconsistent (Z1) fails. If H is a tautology it cannot be tested. And if W is a tautology (Z3c) has no bite. Thus Zytkow's version of bootstrap testing is subject to all of Christensen's counterexamples.

2. Condition (R) and Macho versus Wimp Bootstrapping. For the sake of present discussion take the Glymour conditions for bootstrap testing of H by E with respect to T to be:

(G1) E, H, and T are jointly consistent

(G2) For each quantity Q_i occurring essentially in H, T entails a sentence T_i such that E and T_i jointly entail a value for Q_i, and these values of the Q_i confirm H according to the satisfaction condition S

and

(G3) There is a sentence E' in the language of E such that E' and the T_i are consistent and jointly entail values for the Q_i that disconfirm H according to the satisfaction condition S.[1]

[1]Van Fraassen's (1983) account of bootstrap testing does not require that values for each of the quantities occurring in H can be computed from E. The resulting divergences from Glymour's account will be discussed elsewhere.

This account is subject to the Christensen counterexamples, as illustrated by the following:

Example 1. T: $(x)(Rx \rightarrow Bx)$ & $(x)Gx$
H: $(x)Gx$
E: Ra & Ba
E': Rb & $\neg Bb$

In such a case we would certainly not regard E as confirming H, but (G1)–(G3) are satisfied—in particular, a value for the theoretical quantity Gx can be computed by using $(x)[(Rx \rightarrow Bx) \leftrightarrow Gx]$, which is a consequence of T.

Christensen (1983) proposed to solve the problem by abandoning the equivalence condition.[2] Unwilling to follow suit, Glymour (1983) proposed to add:

(R) H does not entail that any T_i is equivalent to a sentence whose nonlogical vocabulary is a proper subset of the essential nonlogical vocabulary of T_i.

Zytkow's (1986) complaint was that in cases of what can be termed "macho" bootstrapping—that is, cases where a T_i is H itself—(R) can be too strong. Note that (G3) already rules out macho bootstrapping if the satisfaction condition is given by Hempel's criterion, or anything similar; for if the values of the Q_i obtained from E' and the T_i entail the Hempel development of $\neg H$, E' and the T_i will be inconsistent if H is among the T_i. Similarly, van Fraassen's (1983) analysis of bootstrap testing for systems of equations also cuts off the macho straps.

Glymour's initial allowance of macho bootstrapping was motivated by reflecting on tests of the perfect gas law, $PV = kT$, where for purposes of discussion we are to think of P, V, and T as experimentally measurable quantities and k as a theoretical quantity. Now suppose that two sets of values of P, V, and T are obtained. The gas law is then used to compute

[2] The equivalence condition demands that if E confirms H with respect to T, then likewise E' confirms H' with respect to T' if $\vdash E \leftrightarrow E'$, $\vdash H \leftrightarrow H'$, and $\vdash T \leftrightarrow T'$. This condition seems essential both to avoid equivocation and to make the account of bootstrapping conform with our understanding of deductive argument and entailment. It may well be that first-order representations of theories are inadequate to capture entailment relations, truth conditions and other logical and semantical features; but whatever the logical theory chosen, the equivalence condition should be met. Alternatively, testing could be relativized to an axiomatization of T. But then how is this extra argument place to be filled so as to produce a test simpliciter? There are two types of possibilities. First, attention could be restricted to the simplest axiomatization. But even assuming that we have an acceptable explication of simplicity, why should confirmation and disconfirmation turn on simplicity of presentation? Second, the axiomatization could be restricted to ones where the evidence gives warrant for each axiom. But this presupposes that the problem at issue has been solved.

two values for k which are checked against one another. But as Edidin (1983) and van Fraassen (1983) pointed out, it can be held that what one is actually doing in such a case is either directly testing the empirical relation G_1: $(t_1)(t_2)[((P(t_1)V(t_1))/T(t_1)) = ((P(t_2)V(t_2))/T(t_2))]$ or else bootstrap testing G_2: $(t_1)(t_2)[k(t_1) = k(t_2)]$ with respect to the theory consisting of the conjunction of G_2 and G_3: $(t)[k(t) = ((P(t)V(t))/T(t))]$. Condition (R) is not violated in the bootstrap test of G_2 since G_2 does not imply that G_3 is equivalent to a sentence whose nonlogical vocabulary is a proper subset of that of G_3. Zytkow's examples of testing the law of conservation of momentum and Ohm's law are more complicated, but essentially the same points emerge.

3. Further Remarks on Condition (R). In the following case there is no confirmation of H with respect to T, at least not if (R) is imposed:

Example 2. T: $(x)[Gx \,\&\, (Gx \leftrightarrow Ox)]$
H: $(x)Gx$
E: Oa
E': $\neg Oa$

Since theoretical hypotheses in this form rarely if ever occur, Glymour proposed that the revision with (R) was an acceptable approximation.

Setting aside worries about whether or not condition (R) is too strong, there remains the problem that it was postulated *ad hoc* as a response to Christensen's counterexamples and that as it stands it lacks an independent motivation. Zytkow reasoned that the counterevidence E' in example 1 would falsify $(x)(Rx \rightarrow Bx)$ and thus would "remove the inferential support" for $(x)[(Rx \rightarrow Bx) \leftrightarrow Gx]$; this latter statement cannot be confirmed or disconfirmed by data limited to values of Rx and Bx, and its only validation is that it can be inferred from $(x)(Rx \rightarrow Bx)$ and $(x)Gx$. Therefore, $(x)Gx$ can be disconfirmed only if $(x)[(Rx \rightarrow Bx) \leftrightarrow Gx]$ is already discredited as the basis of a computation of a value for Gx (Zytkow 1986, pp. 104–105). The trouble with this reasoning is that it assumes we already have in hand a theory of evidential relevance. But perhaps Zytkow's point can be put thus: E' is unconvincing as a counterexample to $(x)Gx$ since, if $(x)Gx$ is true, the auxiliary hypothesis used to compute a value for Gx is equivalent to another auxiliary (namely, $(x)(Rx \rightarrow Bx)$) that cannot serve as a basis for a computation of a value for Gx but is disconfirmed by E'. This suggests that (R) be replaced by:

(R′) H does not entail that for some i, $\vdash T_i \leftrightarrow T_i'$ where E and T_i' do not jointly entail a value for Q_i and where E' and T_i' are inconsistent.

From the point of view of (R′), condition (R) goes in the right direction since it is the reduction of vocabulary under the assumption of *H* that leads to a violation of (R′) in the Christensen example. We do not know whether adopting (R′) in place of (R) leads to interesting differences for bootstrap confirmation.

In sum, our answer to "What revisions does bootstrap testing need?" is that it needs a condition that is similar to (R) or (R′) but that is better motivated than either of these.

REFERENCES

Christensen, D. (1983), "Glymour on Evidence and Relevance", *Philosophy of Science 50*: 471–481.

Earman, J. (ed.) (1983), *Testing Scientific Theories: Minnesota Studies in the Philosophy of Science, Vol. X.* Minneapolis: University of Minnesota Press.

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

Glymour, C. (1980), *Theory and Evidence*. Princeton: Princeton University Press.

———. (1983), "Revisions of Bootstrap Testing", *Philosophy of Science 50*: 626–629.

van Fraassen, B. C. (1983), "Theory Comparison and Relevant Evidence", in Earman (1983), pp. 27–42.

Zytkow, J. M. (1986), "What Revisions Does Bootstrap Testing Need?", *Philosophy of Science 53*: 101–109.