Aron Edidin

Bootstrapping without Bootstraps

Clark Glymour has persuasively outlined the advantages of an account of the confirmation of scientific hypotheses in terms of what he calls a bootstrap strategy. The stragey relates sentences in an observational vocabulary to hypotheses in a theoretical vocabulary as follows: Auxiliary hypotheses are found that, when conjoined with the observation-sentences in question, entail in a certain way theoretical sentences that, in turn, confirm the hypotheses in some standard way. When the strategy succeeds, the hypotheses are confirmed by the observation-sentences relative to the auxiliary hypotheses (or any theory that contains them).[1]

Despite its many virtues, though, the account of the bootstrap strategy presented in *Theory and Evidence* needs to be modified. The two modifications I propose leave intact the basic structure of the strategy, although one would render the name "bootstrap strategy" somewhat inappropriate. On the other hand, the modifications dramatically alter the relation of Glymour's strategy to Bayesian confirmation theories.

The first modification addresses a problem that arises from the fact that Glymour's strategy goes only part way toward connecting observational evidence with hypotheses. The strategy itself gets us from observation-sentences to intermediate theoretical sentences, which are connected to the hypotheses in question in some other way:

$$\text{Observation sentence} \xrightarrow[\substack{\text{Auxiliary}\\ \text{hypotheses}}]{\substack{\text{Bootstrap}\\ \text{strategy}}} \substack{\text{Intermediate theoretical}\\ \text{sentence}} \rightarrow \text{Hypothesis}$$

In what Glymour calls the Hempelian version of the strategy, the intermediate theoretical sentences are instances of the hypotheses.[2] In other versions, other sorts of sentences might play the role of intermediate. But whatever sentences are used, an observation sentence can confirm a hypothesis by way of such an intermediate only if the intermediate itself

Research for this paper was supported in part by the National Science Foundation.

counts in favor of the hypothesis, relative to the auxiliary hypotheses used by the strategy. Moreover, this support must not be undermined by the observation sentence itself.

The problems that can arise in such a context are especially striking for the Hempelian version of the strategy. It is well known that under certain conditions an instance of a hypothesis can actually count against the hypothesis.

For example, let $A = (a_1, a_2, \ldots)$ and $B = (b_1, b_2, \ldots)$ be two sequences of objects. One or both may be infinite, or both may be finite. Define them in such a way that no object is repeated in either sequence, i.e.,

$$a_n = a_m \text{ iff } n = m$$
$$b_n = b_m \text{ iff } n = m$$

Let R be a relation.

Consider the hypothesis that every member of A is related by R to at least one member of B. That is,

(1) $(x)\,(x\varepsilon A \supset (\exists y)\,(y\varepsilon B\ \&\ xRy))$.

Then the sentence

(2) $a_2\varepsilon A\ \&\ b_1\varepsilon B\ \&\ a_2Rb_1$

is an instance of the hypothesis. It is an instance on Hempel's account, and it seems clear that it will be an instance of the hypothesis on any satisfactory account. And it seems that this instance does confirm the hypothesis.

But suppose that we also have good reason to believe the following: First, that no two elements of A are related to the same element of B, and, second, that no element of A is related to an element of B with a higher subscript than its own, i.e.,

(3) $(x)\,(y)\,(z)\,(x\varepsilon A\&y\varepsilon A\&z\varepsilon B\&xRz\&yRz \supset x=y)$;

(4) $(x)\,(y)\,(a_xRb_y \supset x\geq y)$.

If we have good reason to believe (3) and (4), then (2), although it is an instance of the hypothesis (1), actually gives us good reason to *deny* the hypothesis. Indeed, the conjunction of (2), (3), and (4) is *inconsistent* with the hypothesis. (1) and (4) together entail that a_1 is related to b_1, for b_1 is the only element of B whose subscript is less than or equal to that of a_1. But (2) and (3) together entail that a_1 is *not* related to b_1. For by (2), a_2 is related to b_1. By the original description of A and B, $a_1 \neq a_2$. And (3) asserts that no two distinct members of A are related to the same member of B; a_2 has preempted b_1, so b_1 is not available for a_1.

Actually the problem raised by (1) - (4) is taken care of by a clause of Glymour's Booptstrap Condition. Recall that confirmation by Glymour's strategy is relative to a theory. His formulation requires that the theory in question be consistent with the conjunction of the hypothesis being tested and the evidence statement.[3] Now what we don't want is to be able to confirm (1) by way of (2) relative to any theory that includes (3) and (4). But suppose that evidence statement E is such that (2) can be derived from E and auxiliary hypotheses from the theory. Then the theory and E together entail (2). But if the theory contains (3) and (4), it follows that the theory and E together entail the negation of (1), so Glymour's condition is not satisfied.

It is, however, possible to weaken (3) or (4) in such a way that the original problem is revived. Suppose we replace (3) with the claim that it's highly improbable that two distinct members of A are related to any given member of B.

$$(3') \quad (z)(z\varepsilon B \supset \text{Prob}((\exists x)(\exists y)(x\varepsilon A \& y\varepsilon A \& x \neq y \& xRz \& yRz) < .00001))$$

Now the conjunction of (2), (3′), and (4) is *not* inconsistent with (1), so it is possible to satisfy Glymour's condition even if the theory contains (3′) and (4) and the theory and the evidence jointly entail (2). But (4) entails that (1) and (2) can both be true only if both a_1 and a_2 are related to b_1, and (3′) entails that the probability that that's the case is less than 1/100,000. So if we have good reason to believe (3′) and (4), then (2) counts heavily against (1) even though (2) is an instance of (1).

The core of Glymour's strategy doesn't concern the relation between the intermediate theoretical sentences derived by the strategy and the hypothesis to be confirmed. It seems appropriate, then, simply to require that the intermediate confirm the hypothesis relative to the auxiliary hypotheses used in the derivation, in a way that isn't undermined by the observational evidence.

The second modification is the one that, as it were, cuts the bootstraps off the strategy. A noteworthy feature of Glymour's strategy is that there is no restriction on the set of auxiliary hypotheses used in computations save that they be consistent with the evidence and the hypothesis being tested. In particular, the set of auxiliary hypotheses may entail or even include the hypothesis being tested. This doesn't trivialize the whole procedure because the strategy is arranged to ensure that the computations do not guarantee a positive outcome.[4]

But the use of such sets of auxiliaries causes other problems for the strategy, and these problems are best solved by prohibiting the use of sets of auxiliaries that entail the hypothesis being tested.

To see that this is so, we may begin by considering the following case. The hypothesis to be tested is

(H) A = B.

We have no way of measuring A or B, but we do have means of measuring quantities C and D. We also have these supplementary hypotheses:

(h_1)(=H itself) A = B
(h_2) B = C and
(h_3) B = D.

We can compute a value for B by measuring D and using h_3. We can compute a value for A by measuring C and using h_2 and h_1. There is no guarantee that the resulting values will be equal, since there are possible values for C and D such that C ≠ D. But of course we haven't confirmed that A = B. All we've done has been to compute values for B in two different ways and then obtain a value for A directly from one of the values for B via the hypothesis that A and B are equal.

Glymour suggests that the problem with this computation is that although there are values of the measurable quantities that would disconfirm the hypothesis, there are no such values that are consistent with the supplementary hypotheses other than H itself. He writes, "the possibility of producing a counterexample. . . in this case is spurious, for any such conterexample would have to be obtained from observational data inconsistent with the auxiliaries used to obtain it."[5]

The obvious solution is to require that there be possible values of the measurable quantities that would disconfirm the hypothesis *and* that are consistent with the auxiliary hypotheses other than the one being tested. This is Glymour's approach, incorporated in his Bootstrap Condition.[6]

It is important to understand why, when we require that there be possible disconfirming values that are consistent with the auxiliary hypotheses, we must make an exception of the hypothesis being tested. Suppose that there is a set of possible values for the measurable quantities involved in a set of computations, but that this set of possible values is inconsistent with the hypothesis being tested. Far from yielding only a spurious possibility of disconfirmation, the existence of such a set ensures in the strongest way that a positive outcome is not guaranteed.[7]

We must, then, permit computations for which the only possible set of disconfirming values is inconsistent with the hypothesis being tested. But we can't stop there. Any set of values that is inconsistent with the hypothesis being tested is also inconsistent with any auxiliary hypothesis that entails the hypothesis being tested. We must, then, also permit computations for which the only possible set of disconfirming values is inconsistent with any auxiliary hypothesis which entails the hypothesis being tested. If we do so, however, counterexamples are easy to construct.

In what follows, it will be convenient to deal with sentences rather than equations. If we replace H and h_1, h_2, and h_3 with analogous first-order formulae, the computations look like this (where the positions above and below the arrows indicate stages in the computation, and the sentences to the right of the arrows are auxiliaries):

$$H':(x)(Ax\equiv Bx)$$

$$\begin{array}{llll} Bx & & Ax & \\ \uparrow & h_2': (x)(Bx\equiv Cx) & \uparrow & h_1'(=H'): (x)(Ax\equiv Bx) \\ Cx & & Bx & \\ & & \uparrow & h_3': (x)(Bx\equiv Dx) \\ & & Dx & \end{array}$$

Glymour suggests that the problem with these computations is that the only values for Dx and Cx that would yield a counterinstance of H′ would also be inconsistent with the conjunction of h_2' and h_3', the two auxiliary hypotheses that do not entail H′.

But a counterexample to the modified Bootstrap Condition may be obtained if we now alter these computations by conjoining H′ to each of the auxiliary hypotheses. In the new set of computations, there is no auxiliary hypothesis that does not entail H′. The modified consistency condition is satisfied vacuously: since every auxiliary hypothesis entails H′, the condition applies to none of them. But the new computation clearly inherits the inadequacy of the old. The original computation doesn't confirm H′, and merely conjoining H′ itself to each auxiliary hypothesis doesn't change matters. The new computation satisfies the Bootstrap condition with the expanded exception to the consistency condition, but it does not confirm H′.[8]

A second sort of counterexample, which will be useful later, may be constructed as follows: find a single hypothesis that entails all the original auxiliary hypotheses, and thus entails H′. Replace each auxiliary hypothesis with this one. Now there is only one auxiliary hypothesis, and it entails H′. Duplicate the computations in the original case in this way: a computation from the original set will be replaced by one that uses only the single auxiliary hypothesis but proceeds by first deriving from the single hypothesis those hypotheses that appeared in the original computation and then proceeding as before. This computation, too, satisfies the modified Bootstrap Condition, and it too fails to confirm H′.

The counterexamples can be avoided by prohibiting the use in computations of hypotheses that entail the one being tested. And this is a natural response to the problems they pose. In the original case, the hypotheses other than H′ provided two different ways to compute values for Bx, but no way to compute a value for Ax. This has the effect of ensuring that values for Cx and Dx that yield counterinstances of Ax will be inconsistent with these hypotheses, for such values will have to yield two distinct, incompatible values for Bx. But given this structure for the computations, H′ or hypotheses that entail it will be required in order to get a value for Ax out of one of the values for Bx. So it is only the presence of H′ or auxiliary hypotheses that entail H′ that enable the objectionable computations to yield an instance of H′. By prohibiting the use as auxiliaries of hypotheses that entail the one being tested, the counterexamples are avoided.

I have said that Glymour's strategy should be modified by prohibiting the use of sets of auxiliary hypotheses that entail the hypothesis being tested. It may seem that the arguments to this point support only a weaker conclusion: that the use of *individual* hypotheses that entail the one being tested should be prohibited. Actually they support the stronger conclusion as well. Note first that the problem in the original example doesn't go away if we change the computation so that no single auxiliary hypothesis entails H′. If we replace h_1' by two separate hypotheses:

$$(x)\ (Bx \equiv A'x) \quad \text{and} \quad (x)\ (A'x \equiv Ax)$$

the computation on the right looks like this:

$$\begin{tikzcd} Ax \\ A'x \arrow[u, "(x)(A'x\equiv Ax)"'] \\ Bx \arrow[u, "(x)(Bx\equiv A'x)"'] \\ Dx \arrow[u, "(x)(Bx\equiv Dx)"'] \end{tikzcd}$$

Clearly nothing essential is changed.

But there's a more general argument. Recall the second sort of counterexample, in which the auxiliary hypotheses were replaced by a single, stronger hypothesis that entails each of them. Now consider any set of computations that uses a set of auxiliary hypotheses that jointly entail the hypothesis being tested. To any such set of computations, there corresponds (in the way indicated in the counterexample) an equivalent set of computations using a single auxiliary hypothesis that is itself equivalent to the conjunction of the first set of hypotheses. If the latter set of computations is ruled out, the former should be as well. But if individual hypotheses that entail the hypothesis being tested may not be used, the latter set of computations *is* ruled out, because its single auxiliary hypothesis does entail the one being tested. It follows that sets of computations using auxiliary hypotheses that jointly entail the one being tested should be excluded.

It may be that there is a way of altering the Bootstrap condition that avoids the counterexamples but falls short of prohibiting the use of sets of auxiliaries that jointly entail the hypothesis being tested. But I think that the cases I've decribed create a strong prima facie case in favor of the restriction I propose.

Philosophers who believe that testing a hypothesis essentially involves putting the hypothesis in jeopardy have a further reason to prohibit the use of sets of auxiliaries that entail the hypothesis being tested. For unless this restriction is adopted, it will be possible to confirm hypotheses via sets of computations such that no values of the evidential quantities would *disconfirm* the hypothesis relative to any theory. Indeed, this will be the case whenever the set of auxiliaries used in the computation entails the hypothesis being tested. The Bootstrap condition requires that there be values of the evidential quantities that yield a counterinstance of the hypothesis by way of the computations. But those values only disconfirm the hypothesis if they confirm its negation. As was noted in the discussion of disconfirmatory instances above, they can confirm the negation of the hypothesis relative to a theory only if the theory is consistent with the negation of the hypothesis and also includes all of the auxiliary hypotheses used in the computations. But if those auxiliaries entail the hypothesis, no theory that includes them is consistent with the negation of the hypothesis. It follows that if the hypotheses used in a set of computations entail the hypothesis being tested, those computations cannot disconfirm the hypothesis relative to any theory at all. If the computations satisfy the Bootstrap Condition they won't guarantee a positive outcome—they might fail to confirm the hypothesis—but they do guarantee that there won't be a negative outcome—they cannot disconfirm the hypothesis. And it might well be thought that computations that shield a hypothesis from disconfirmation in this way cannot confirm it.

The case against the use of sets of auxiliaries that entail the hypothesis being tested is strengthened by the apparent absence of any good reason to include them. Since Glymour on several occasions notes that his strategy permits the use of such computations, and indeed uses such a computation in the very example with which he introduces his strategy, it might be thought that the restriction cuts to the very heart of the strategy. That is not so.

Although Glymour claims that cases violating the restriction abound in the scientific literature, it is striking that no such case appears in any of the chapters concerning applications and historical examples of the use of his strategy.[9] The cases he has in mind seem to be cases in which the hypothesis being tested is used to determine the value of a *constant* occurring within it. In this they resemble the example Glymour uses to introduce his strategy. Let us then consider this example.

To be tested is the hypothesis that for any sample of gas, so long as no gas is added to or removed from the sample, the product of the pressure and volume of the gas is proportional to its temperature. This amounts to the claim that, under the given conditions,

$$PV = kT,$$

where k is an undetermined constant. If we can measure P, V, and T, but cannot measure k, "the hypothesis may be tested by obtaining two sets of values for P, V, and T, using the first set of values together with the hypothesis [itself]. . . to determine a value for k, i.e.,

$$k = \frac{PV}{T},$$

and using the value k thus obtained together with the second set of values for P, V, and T either to instantiate or to contradict the hypothesis."[10]

It is clear, though, that in this case the hypothesis can be tested in Glymour's manner without itself being used in the computation. The hypothesis that Glymour formulates as the claim that under certain conditions PV = kT may equivalently be formulated as the claim that under those conditions,

$$\frac{PV}{T} = \frac{P'V'}{T'}$$

for any two sets of values for pressure, volume, and temperature. From the same two sets of values that were used in Glymour's computation, we may now instantiate or contradict the hypothesis directly, without the need to compute values for any additional quantities.

Moreover, the same procedure can be used whenever it is necessary to use the hypothesis being tested to compute the value of a constant occurring within it. In such a case, the hypothesis must be equivalent to a statement of the form

$$k = f(Q_1, \ldots, Q_n),$$

where values for $Q_1, \ldots, Q_n$ are either measurable or computable from observations by way of other auxiliaries, and k is an undetermined constant. But then, the hypothesis is also equivalent to a statement of the form

$$f(Q_1, \ldots, Q_n) = f(Q'_1, \ldots, Q'_n)$$

for any two sets of values of the quantities, so the hypothesis can be instantiated or contradicted directly.

It seems that we have yet to see an example of an ineliminable use of the hypothesis being tested in a legitimate computation. Moreover, one of Glymour's own arguments concerning variety of evidence suggests that computations using the hypothesis being tested don't yield very good tests of the hypothesis anyway. One reason variety is desirable is that you don't have much support for hypotheses A and B if you've only tested A by using hypotheses that entail B and you've only tested B by using hypotheses that entail A. Glymour develops this point in some detail in his discussion of Kepler's laws[11] and reiterates it in his discussion of criteria of theory choice,[12] where he suggests that a theory containing such a pair of hypotheses is ceteris paribus less well tested than one that does not. But if the only available test of A uses hypotheses that entail A, the result is a degenerate case of the problem (for A = B). (Note that this does *not* seem to be a problem for Glymour's example concerning the hypothesis PV = kT. I take this to confirm my claim that the use of the hypothesis in that computation is inessential.)

It would seem, then, that the use of sets of auxiliary hypotheses that entail the hypothesis being tested is dispensible. Since dispensing with the use of such sets also enables us to avoid various counterexample to the unaltered strategy, I conclude that the use of such sets should be prohibited.

Before leaving the ideal gas law too far behind, it will be worthwhile to discuss one feature of Glymour's strategy that should *not* be changed. Glymour notes that "the consequence condition, which requires that every logical consequence of any collection of confirmed hypotheses be confirmed as well, is not met." But he claims that "there is no difficulty. . . in simply imposing the consequence condition independently, so that all logical consequences of whatever consistent class of sentences is confirmed by the bootstrap condition are taken to be confirmed as well."[13]

But the example we've just been considering, concerning the hypothesis that PV = kT, suggests that imposing the consequence condition would itself have unfortunate consequences. It seems clear that at least two sets of values of pressure, volume, and temperature are needed to test this hypothesis. Both Glymour's computation and my modified version use two sets of values. No single set of values can test the hypothesis by way of the Bootstrap Condition alone, because no single set of values can yield a counterinstance. But if the Consequence Condition is added to the Bootstrap Condition, then *any* single set of values will not only test but confirm the hypothesis. Let (a, b, c) be such a set of values. Then the set (a, b, c) confirms the hypothesis that

$$PV = \frac{ab}{c}T.$$

But that hypothesis entails that for some k, $PV = kT$, which is just to say that it entails $PV = kT$ where k is an undetermined constant. If every logical consequence of a confirmed hypothesis is itself confirmed, then, the single set of values a, b, and c confirm that $PV = kT$.

I have argued that Glymour's strategy of confirmation should be modified in two ways. The satisfaction condition should be altered to require that the conjunction of the instance obtained and the evidence used count in favor of the hypothesis relative to the auxiliary hypotheses used in the computation, and sets of auxiliary hypotheses that entail the hypothesis being tested should be excluded. I have argued further that neither of these modifications seriously violates the intent of the strategy. But if these two modifications are adopted, the resulting strategy is easily derivable from a Bayesian confirmation theory.

Consider a set of computations that satisfies the modified Bootstrap Condition. If E is the evidence statement and $h_1, \ldots, h_2$ the auxiliary hypotheses, then the conjunction of E and h_1 through h_2 entails an instance of the hypothesis being tested. Moreover, the instance in question counts in favor of the hypothesis relative to the evidence and auxiliary hypotheses. So we have

(1) $E \mathbin{\&} h_1 \mathbin{\&} \ldots \mathbin{\&} h_n \rightarrow S$,

and, if the prior probability of E is less than 1,

(2) $p(H/S \& E \& h_1 \& \ldots \& h_n) > p(H/h_1, \ldots, h_n)$,

where H is the hypothesis being tested, S is an instance of H, and E is the evidence-statement used in the computation. Since the auxiliaries can't entail H, this is no empty identity.

Now (1) entails that (if coherence is preserved)

(3) $p(H/S \& E \& h_1 \& \ldots \& h_n) = p(H/E \& h_1 \& \ldots \& h_n)$

(2) and (3) together entail that

(4) $p(H/E \& h_1 \& \ldots \& h_n) > p(H/h_1 \& \ldots \& h_n)$.

But (4) is just the formal way of saying that E is positively relevant to H, relative to the auxiliaries $h_1, \ldots, h_n$. Thus if E confirms H relative to theories that entail $h_1, \ldots, h_n$ by the modified version of Glymour's strategy, E is positively relevant to H relative to $h_1, \ldots, h_n$ on a Bayesian theory. What was advanced as an alternative to the Bayesian theory turns out to be derivable from the theory.

The principal advantage Glymour claims for his strategy is that it explains how evidence selectively confirms some hypotheses but not others. He argues that Bayesian theorists cannot do the same without recourse to ad hoc restrictions on prior probabilities. But if I'm right about the need to modify the strategy as I suggest, and about its subsequent derivability from the Bayesian theory, then it may be that the principal benefit of the strategy is that it shows how Bayesians can explain the selective relevance of evidence without ad hoc restrictions.

Notes

1. The strategy is described in greater detail in Glymour's contribution to this volume, and in his *Theory and Evidence* (Princeton; Princeton University Press, 1980), Chapter 5, esp. pp. 130-131. I discuss the various virtues of Glymour's account, as well as certain problems which are examined in greater detail below, in "Glymour on confirmation," *Philosophy of Science* 48 (1981).

2. This is the version presented in Chapter V of *Theory and Evidence*. Carl Hempel's account of confirmation by instances appears in Studies in the Logic of Confirmation. In *Aspects of Scientific Explanation* (New York; Free Press, 1965).

3. *Theory and Evidence*, p. 130 (clause i. of the Bootstrap Condition).

4. Thus, in the Hempelian version of the strategy, there must be possible values of the evidential quantities that yield a counterinstance of the hypothesis via the computations actually used to derive an instance. See *Theory and Evidence*, pp. 130-131 (clause iv. of the Bootstrap Condition).

5. *Theory and Evidence*, p. 117n.

6. Ibid., clause iv., pp. 130-131.

7. This is in fact the case in the example which Glymour uses to illustrate his strategy. See *Theory and Evidence*, p. 111.

8. Paul Horwich has noted that this counterexample can be avoided by allowing the hypothesis being tested to be used as an auxiliary, but prohibiting the use of any other auxiliaries that entail it. But this added condition remains vulnerable to other counterexamples. The following computation satisfies the proposed restriction but inherits the inadequacy of the original example:

$$\begin{tikzcd} Ax & \\ (x)((Ax \equiv Bx) \supset (Bx \equiv Cx)) \& Cx \arrow[u, "(x)(Ax \equiv Bx)"'] & Bx \\ Cx \arrow[u, "(x)((Ax \equiv Bx) \supset (Bx \equiv Cx))"'] & Dx \arrow[u, "(x)(Bx \equiv Dx)"'] \end{tikzcd}$$

9. *Theory and Evidence*, Chapters VI-IX.

10. Ibid., p. 111.

11. Ibid., pp. 140-141.

12. Ibid., p. 153.

13. Ibid., p. 133.