Bas C. van Fraassen

Theory Comparison and Relevant Evidence

What is the main epistemic problem concerning science? I take it that it is the explication of how we compare and evaluate theories, as a basis either for theory acceptance or for practical action. This comparison is clearly a comparison in the light of the available evidence—whatever *that* means.[1] My aim in this paper is very modest: to explore the logic of evidential support in the theory of Clark Glymour.[2] The question of how, and to what extent, overall theory comparison is to be related to the data from relevant tests, I broach in another paper.[3]

First I shall sketch a model of theory structure that covers what I take to be a large and significant class of scientific theories. With reference to that schema, I reconstruct notions of relevant evidence as introduced in Glymour's "bootstrap theory" of testing hypotheses. Then I shall explore the relations of support by the same evidence, between different hypotheses. The form chosen for the reconstruction is so as to make clear the application of Glymour's ideas to the case of a *developing* theory which is becoming more testable. Glymour himself and Michael Gardner have examined the development of atomic theory in the nineteenth and early twentieth centuries from this point of view.[4]

1. A Schema for (Simple, Quantitative) Theories

The semantic approach to theories, initiated by Beth and closely related to the approach introduced by Suppes, gives us an account that I shall here use as a framework for relations between tests, evidence, and hypotheses.[5] In doing this we broaden somewhat the initial basis for Glymour's own exposition, but we ignore his logistical reformulation prompted by compar-

This article is reprinted by courtesy of Adolf Grünbaum and Larry Laudan, general coeditors of the *Pittsburgh Series in the Philosophy and History of Sciences* published by the University of California Press. It will appear in the proceedings volume for the Workshop on the Limitations of Deductivism, held at the University of Pittsburgh Center for the Philosophy of Science in November 1980. The research for this paper was supported by National Science Foundation research grant SES-8005827, which is hereby gratefully acknowledged.

isons with Hempel.[6] Without attempting needless rigor, I shall describe a scientific theory of the sort I have in mind by describing what must be given to specify it exactly.

It has, to begin, a set W of possible *situations*. This is broadly conceived, so that the effect of its laws is to entail that many of these situations are not really possible.

Second, it has a set Q of *quantities*. For ease of present exposition I shall postulate that each quantity has a value (either a real value, or a real-valued function of time) in each possible situation. Quantity q has value q(w) in element w of W.

Although it is crucial to Glymour's theory that distinct quantities may yet have the same value in all *really* possible situations, I shall take the set W to be so broadly conceived that a quantity q can be identified with (or via) the function that assigns to each possible situation in W, the values q has in that situation.

The set Q is closed under the formation of new quantities, by means of composition and transformation of functions taken as broadly as you like, but including at least composition and polynomials of the usual sort. Because of the identification in the previous paragraph, there is no division into "atomic" and "molecular" quantities. Thus $q + q - q$ is identically the same quantity as q itself. A theory's laws may imply that q and q′ always have the same value, and this will be stated as $(q = q')$ as usual; that means only that in all the situations w allowed by the law, $q(w) = q'(w)$. This is standard notation.

Formulas used to phrase hypotheses and laws describe relations between values of quantities. Thus Newton's second law says that at each instant of time, the (value of the quantity) impressed force equals the product of the (values of the quantities) mass and acceleration. Equality is one relationship among values; others may appear. I shall focus here on all the relations of equality and inequality ($=$, $\leq$, $<$, $\neq$, etc.) and use the letter R to stand for them. A *basic proposition* or *basic formula* is one of form (tRt′), where t and t′ denote quantities. It is *true* in situation w exactly if t(w) bears R to t′(w).

The expression "solution of an inequality" is familiar and I shall take it for granted, but restrict its use as follows: a *solution* of basic formula A is a function s that assigns a value at least to each quantity that is denoted by a term, or well-formed part of a term, which appears in A, and there is some possible situation w such that

(a) $s(q) = q(w)$ for each q for which s is defined, and
(b) A holds in w.

For example, the function s that assigns values 3, 4, and 12/k to the pressure P, volume V, and temperature T of a gas, is a solution of the basic formula ($PV = kT$), provided however that space W of possible situations has some situation in which these are the correct values.

The domain of a solution of formula A includes the set of quantities denoted by well-formed parts of A; I shall call that set Q(A). Since the terms t, t′ in (tRt′) may be complex, that set will usually have more than two members. A formula A is called *logically valid* exactly if for each w in W, any map s that assigns $q(w)$ to q for each q in Q(A), is a solution of A. Similarly we may say that A *implies* B exactly if every solution of A that is also well defined on Q(B), is a solution of B.

Finally the theory has a set of *basic postulates*, each of which is a basic proposition. The *possibility space* Sp(T) of theory T is the set of possible situations in which all basic postulates are true. As a theory develops through the addition of postulates, this possibility space is narrowed; the set W of possible situations stays the same. Theorems are basic formulas implied by the postulates.

It will not cause confusion in this context to use the same letter T to denote the theory and its set of theorems, for I shall keep W and Q fixed. What the theory *says* is that any real situation is one in which all its theorems are true.

2. Characterizing Relevant Evidence

We must first describe the possible evidence that can be had. This is especially important here because for Glymour the relation between evidence E and hypothesis H depends in part on what the evidence might or could have been instead.

When a test is carried out, the values of certain directly measurable quantities are ascertained. What tests are possible, which quantities are directly measurable, we must here take as given. (In my own opinion, and I think also in Glymour's, this is a question of contingent, empirical fact.) The result of a test I shall call a *data set* or *data base*: it is a function E that assigns to those measured quantities the values ascertained. So any solution of a basic formula might, prima facie, be a data base; the two defining constraints on *possible data bases* E are, *first*, that each quantity in the domain of E is directly measurable, and *second* that there is a possible situation in which all the basic formulas $q = E(q)$, for quantities q in the domain of E, are true. It will be convenient much of the time to identify E with that set of basic formulas.

An *alternative* E′ to data base E is a possible data base with the same domain as E. It will be noted that the theory itself, in delimiting the class W of possible situations, has implications concerning what evidence could in principle be had. A *hypothesis* is any basic formula. The first Glymourian mode of evidential support is the following.

(2-1) E provides *weakly relevant evidence* for H *relative to* theory T exactly if E has some alternative E′ and T some subset T_0 such that:
(1) $T \cup E \cup \{H\}$ has a solution.
(2) $T_0 \cup E'$ has a solution.
(3) All solutions of $T_0 \cup E$ are solutions of H.
(4) No solutions of $T_0 \cup E'$ are solutions of H.

As a simple example, let A, B, C be directly measurable quantities, and let T have postulates

$$
\begin{array}{ll}
P_1 & A = x + u \\
P_2 & B = u - z \\
P_3 & C = x + z
\end{array}
$$

Let hypothesis H be postulate P_3, and let the evidence E be $\{A = 3, B = 4, C = -1\}$. Using subtheory $T_0 = \{P_1, P_2\}$ and alternative evidence base $E' = \{A = 3, B = 4, C = 1\}$, we can verify at once that E provides weakly relevant evidence for P_3 relative to theory T. The different values given to C in E and E′ show at once that clause (4) holds. To verify clauses (1) and (2) we note that

$$
\begin{array}{l}
s\colon A = 3,\ B = 4,\ C = -1,\ x = 10,\ u = -7,\ z = -11 \\
s'\colon A = 3,\ B = 4,\ C = 1,\ x = 10,\ u = -7,\ z = -11
\end{array}
$$

are solutions of $T \cup E$ and of $T_0 \cup E'$ respectively. The crucial clause (3) also holds because $T_0 \cup E$ implies at the same time that $A - B = C$ (this comes from E) and $A - B = x + z$ (that comes from T_0), hence the two right-hand sides must have the same values, as P_3 says. I shall return to this example at several points below.

The above relation of weak evidential support already accomplishes one of Glymour's aims: it is hypothesis-specific, allows for selective support, and gets us out of methodological holism. But Glymour takes equally seriously the idea of the *empirical determination* of specific quantities. A test of a hypothesis about temperatures, relative to a theory in which temperature equals mean kinetic energy of the constituent molecules, can yield weakly relevant evidence for a hypothesis about molecular kinetic energies. But since in measuring temperature we deal only with the mean, this is a test in which no specific molecular energy is determined, except perhaps in some absolutely minimal sense. More full-blooded, more and more precise empirical determination of exact values of theoretical quantities plays an extremely important role in the building up of evidential support for the theory, according to Glymour.

As my phrasing already suggests, this aspect of tests admits of degrees, or at least of comparisons of more and less. I shall list three possible amendations for (2-1); the first of these was suggested by Glymour in correspondence.

(2-2) If q belongs to Q(A) then the following are possible:
- (1) The set of values assigned to q by solutions of $T_0 \cup E$ is properly contained in the set of values assigned to q by solutions of T_0.
- (2) All solutions of $T_0 \cup E$ assign a value to q in the same interval I.
- (3) All solutions of $T_0 \cup E$ assign the same value to q.

One of these conditions should be added as fifth clause (2-1), to arrive at a stronger criterion of relevant evidence. Next we can strengthen the criterion by adding as sixth clause a corresponding condition for the alternative data base E′. To mark the extreme, in which every quantity in the formula is calculable from the data base via the theory, I shall give it a name too.

(2-3) E provides *strongly relevant evidence* for H relative to theory T exactly if E has some alternative E′ and T has a subtheory T_0 such that
- (a) clauses (1)-(4) of (2-1) hold
- (b) clause (3) of (2-2) holds, and holds also when E is replaced by E′ for all q in Q(H).

In section 4 I shall give examples of strongly relevant evidence.

Thus the evidence is relevant to A if it allows selective testing of A (relative to the theory), but the degree of relevance is determined by how closely the evidence allows us to calculate (relative to the theory again) the values of the quantities that appear in this formula. The reader is asked to think up names for intermediate degrees of relevance.

3. Significance of Glymour's account

How good is my reconstruction of Glymour's account? I take it to be good; good but not perfect. So I shall explain some points of divergence and also say why I believe the present account to share the considerable virtues that can be claimed for Glymour's own.

Most obviously omitted from my version is the notion of computation. I have been unable to find a precise reconstruction to which the use of that notion makes a difference; and I believe that Glymour is now also of the opinion that this had a didactic use only.[7]

Second, Glymour gives his account twice over, once for theories stated in equational form, and once for those formulated in a first-order language. I do not find the latter idealization very interesting, and believe that it actually has definite harmful effects on philosophical discussions. But it must be presumed that, although Glymour does not say so explicitly, his account for equational theories needs some emendation corresponding to clause (2) in the statement of the *Bootstrap Condition* for the first-order case.[8] I shall not discuss such minor variations in further detail; as Glymour himself says, we should just be flexible in their admission as problems arise.

My reasons for stating the account in several definitional stages are two. First, I want to allow for approximations. Even if the evidence (data base) is totally precise (and I take allowances for imprecise data to introduce no difficulties of principle), it may not determine the *exact* value of other quantities in the hypothesis. Second, as a theory develops, as in the example of the atomic theory in the nineteenth century, more and more quantities may become empirically determinable; conversely, a good deal of testing is already being done, while some of the crucial quantities still remain indeterminate. For example, Gardner states that Herapath derived the equation $PV = Nm\,v\,2/3$, and used this to make "the first calculation ever published of molecular speeds." I have no reason to doubt that his contemporaries saw it that way, and considered him to have answered the objection that molecular speed is an empirically indeterminable quantity. Yet it is equally clear that the most the formula will help us deduce is a mean value, or some characteristics of the distribution, of that parameter in the total body of gas characterized by P and V. Between 1803, when Dalton proposed the theory, and a century later, when Einstein worked on Brownian motion, the developing theory became ever *more testable*, and this must be reflected in our account.

As long as some hypothesis of a theory is not supported by strongly relevant evidence, relative to that theory, that can be raised as an objection to the theory. Glymour raises this sort of point to explain the felt deficiency of "de-occamized" theories; and theories that admit of distinctions said not to have any physical meaning (e.g., phase of the wave function). As long as the theory is still developing, however, we are better advised not to turn our backs on it if we cannot immediately get strongly relevant evidence, but to look to more tenuous sources of evidential support as well. And finally, in the theory's days of triumph, when the development is thought to be complete, mathematicians may be turned loose to remedy such apparent deficiency (e.g., replacement of wave functions by statistical operators to represent states) because it is no longer felt that any and all potential resources should be husbanded for the rainy days ahead.

As I said, the virtues that may be claimed for Glymour's theory are considerable. By explicating selective testing of hypotheses in bootstrap fashion, he has illuminated an important and pervasive aspect of scientific methodology. By focusing our attention on the empirical determination of values, for most directly measurable quantities, in such tests, he has shown how a developing theory may become more and more testable, and how more and more strongly relevant evidence may become available.

But while lauding these achievements, I want to add two criticisms. These concern virtues that cannot be claimed, and should instead be marked down as open problems.

The theory of evidence has a number of standard problems. One is the apparent additional value evidence takes on if it is collected after the theory has predicted it. I say "apparent" because perhaps there is only the value any evidence has when it is clear that the theory has not been "cooked up" to account for it. Glymour provides a formidable problem to Bayesian approaches that is more or less a variant of that familiar one: how can a Bayesian see *any* confirmation for a theory in evidence that was assimilated before the theory was proposed? Attempts to meet this objection by reference to probabilities formed prior to the gathering of that evidence fail for a number of reasons. However this may be, we must not ignore the fact that Glymour's theory has itself nothing to say on this point.

A second such standard problem is that a *variety* of evidence appears to provide better support for a theory than a narrow spectrum of evidence. Again I say "appears," for it is not clear just how objective the classification of phenomena needed to make this distinction is. (Do the sudden shocks felt by lovers kissing on a rug, Benjamin Franklin's kite, balloons sticking to the wall, St. Elmo's fire, and the Leyden jar form a narrow or a broad spectrum of phenomena?) Glymour takes it seriously, objects to various accounts that have been given of it, and offers his own.[9] But his proffered justification of the methodological requirement of variety of evidence, consists in the observation that without that requirement, his account would lead us to say that a hypothesis is well confirmed when in fact it is not. Specifically, he uses Kepler's ubiquitous laws once more to show a shortcoming of the bootstrap conditions that for its remedy require a demand for variety of evidence. I don't quite know how to distinguish this sort of merit from an objection.

In this connection a point may be made concerning tactics. Just by explicitly *building in* the requirement that all relevant quantities be determinable by the computation used in confirmation (i.e., opting for what I call strongly relevant evidence in his definition of confirmation), Glymour automatically precludes his acount from providing an explanation of why scientists should consider it a shortcoming if not all relevant quantities are empirically determinable.

We all have convictions concerning the importance of these notable features of evidence, and use them to gauge evidential support. But why? What are the basic aims of the activity of testing, experimenting, evidence gathering that must explain, if anything does, our predilection to gauge evidential support in these ways? Bayesians pride themselves on being able to offer some rationale, and I have sketched an approach to this problem in my other paper on Glymour's theories.[10] Glymour's own account is much closer to the mire and blood of research practice; but it does not, it seems to me, justify these general features of scientific methodology.

It is time to return to the exploration of relevant testing and evidence. I shall do so by means of a series of logical thought experiments.[11]

4. Gauging Evidential Support

In the debate over scientific realism, one epistemological question is always lurking in the background: can the evidence give greater support to one theory than another even if the two are empirically equivalent? Glymour has argued the affirmative by attempting to give us reason to think that an extension of a theory, which makes no new empirical predictions, can be better supported than the original.[12] This is very audacious; but it accords with the quite familiar phenomenon that a lengthy, circumstantial story is often more plausible than a bare statement of some of its implications. On the other hand, it is in apparent conflict with the opinion, certainly widespread among philosophers, that a longer story has more ways of being false, and is therefore less likely to be true. This is a puzzling situation and needs to be clarified. I shall return to it in the next section, after a look at the logic of Glymourian testing.

While recognizing that his theory of evidence does not immediately lead to an account of theory comparison, Glymour offers some criteria for comparing entire theories that, he says, emerge from the bootstrap theory.[13] The notions of being *testable* and *tested* that he uses in this discussion can be defined as follows.

(4-1) Hypothesis A is [*strongly*] *testable* [respectively, *tested*] *with respect to* theory T exactly if there exists some data-set [respectively, we have in our possession some data-set] that provides [strongly] relevant evidence for A relative to T.

(4-2) Theory T is *optimally* [*strongly*] *testable* [respectively, *tested*] exactly if it is axiomatized by a set of hypotheses each of which is [strongly] testable [respectively, tested] with respect to T itself.

Glymour shows (and I shall give further examples below) that a subtheory may be less well tested than the whole theory of which it is part. Indeed, in some cases the subtheory is not testable at all. Conversely, any subtheory or hypothesis will be at least as well tested with respect to a larger, logically stronger theory than with respect to a smaller one.

(4-3) If T implies T′ and E is consistent with T, and E provides [weakly, strongly] relevant evidence for consequence A of T′ relative to T′, then E also provides this for A relative to T.[14]

Thus for any given hypothesis, we may increase the relevant evidence (*tout court*, relative to the theory we accept) without doing new experiments, by introducing stronger postulates. Of course, it may be objected that in such a case our *overall* theory will then have less relevant evidence, or at least be less tested, than before we accepted these stronger postulates. That this objection is mistaken follows exactly from the fact that, on the contrary, a whole theory may be better tested than a given subtheory.

I do not mean this as a reductio. It is a credo of Wilfrid Sellars, and also of Gilbert Harman, that all inference consists in placing (or attempting to place) ourselves in a position where we have a better and more explanatory account of the world, and that when we have *that* our epistemic position is more secure.

Glymour provides us with an instructive example of the situation I have just described in the abstract, pointing to celestial mechanics: "Seventeenth century astronomers were able to confirm Kepler's first law only by using his second, and they were able to confirm his second only by using his first."[15] An artificial example using three postulates is easily constructed; in fact the example of section 2 above will do. Recall that A, B, C are the directly measurable quantities:

$$
\begin{array}{llll}
P_1 & A & = & x + u \\
P_2 & B & = & u - z \\
P_3 & C & = & x + z
\end{array}
$$

For each of the three postulates we can find a data base that provides weakly relevant evidence for it relative to the whole theory. In section 2 we looked at relevant evidence for P_3 relative to this theory. Yet neither P_3 nor P_2 is testable relative to the subtheory formed by dropping the first postulate.

So if someone originally wished to advocate the theory whose postulates are just P_2 and P_3, there is no doubt that his advocacy would be substantially improved if he decided to accept P_1 as well. Before he does so, his theory simply cannot pass, or fail, any test at all. Nor can it be of more interest than the hypothesis that gravity is a form of love, not even if he tells us that his terms B, C, x, u, z denote such eminently interesting and quantifiable magnitudes as performance on an I.Q. test, number of cars owned, aptitude for sailing, femininity quotient, and so forth. Depending on what these terms are, or how they are interpreted, the theory may be meaningful, possibly true or false, audacious, or even outrageous; but it is not testable. The evidence *cannot* give us reason to accept it, in this sense.

In commentary on this paper, Glymour has suggested a plausible way to state this fact in general: the evidence can give us less reason to accept a theory T taken as a whole, than to accept a larger theory T′, of which T is a part, taken as a whole. And in one straightforward sense, that is just what the preceding examples show, and cannot be denied by anyone. But it seems to me that the situation is more puzzling yet than these examples bring out. For even relative to the large theory, one of its larger parts may receive more suport than one of the smaller parts thereof. Relevant evidential suport is not inherited by consequences.[16]

Recall that Q(B) is the set of quantities denoted by expressions found in B. Let us examine the following question:

(4-4) If A implies B, and E provides [strongly] relevant evidence for A relative to T, must E provide [strongly] relevant evidence for B?

The answer is no, simply because B may be irrefutable by any data base alternative to E; for example if B is logically valid. The form of this counter example shows at once that adding the proviso-clause $Q(B) \subseteq Q(A)$ to (4-4) will not get us a positive answer either.

Would it help to add "provided B is not logically valid"? No, for it may be that any subtheory T_0 needed in the calculation leading from an instance of E (or its alternative) to values for quantities in A, will imply B and hence make refutation impossible. Here is an example. Let f_X be the a characteristic function of set X. That means that for each quantity q we have another quantity $f_X q$ that takes value 1 exactly when q takes a value in X, and that takes value 0 otherwise. We can furthermore use multiplication; thus $q(f_X q)$ will take value q or *zero* depending on whether q takes a value inside X or outside it. Let us now attempt to test two hypotheses

$$H_1 \quad q = 1$$
$$H_2 \quad 0 \leq q \leq 1$$

and suppose that there is only one usable subtheory, consisting of a single postulate of the theory (plus the theorems implied by that postulate alone), namely

$$P_0 \qquad A(f_{[0,\ 1]}A) \leq q \leq A$$

where A is directly measurable and q is not.

If evidence E gives A a value x; then with the help of P_0 we calculate $x \leq q \leq x$ if x lies in [0, 1], but we can conclude only that $0 \leq q \leq x$ if x lies outside that interval (thus P_0 rules out negative values for A, as well as for q, and has therefore some immediate empirical import). In the first case the alternative data base $E' = \{A = 1/2\}$ can be cited to show that E provides strongly relevant evidence for hypothesis H_1. But there is no data base at all that can provide even weakly relevant evidence for H_2. The reason is that we can find no E′ which, together with T_0 implies that $q \nleq 1$; hence no alternative to E of the sort needed for relevant evidence.

The astonishing consequence is that in H_2 we have a hypothesis that is certainly not logically valid, that can be established on the basis of theory plus evidence (namely as a consequence of H_1), and that cannot even possibly receive the support of even weakly relevant evidence relative to that theory.

It may be thought that in any such case, either the subtheory used in the calculations will itself be untestable, or else the hypothesis will be capable of support from data bases used to test that subtheory. Combining the strategies of the two sorts of examples we have had, I shall now show that this surmise too is incorrect. Let the directly measurable quantities be A and B and the "theoretical" quantity q, and let the two hypotheses to be considered again be H_1 and H_2 above. But now let the usable subtheory consist of two postulates, namely P_o above plus:

$$P_{oo} \quad Bf_{[0,\ 1]}B \leq q \leq B.$$

As data bases, consider:

$$E = \{A = 1,\ B = 1\}$$
$$E' = \{A = 1,\ B = 1/2\}$$
$$E'' = \{A = 1/2,\ B = 1\}$$

Relative to any theory including P_o and P_{oo}, we find that E provides strongly relevant evidence for each; hence the little theory whose sole postulates are these is optimally testable. (For P_o plus $(A = 1)$ implies that $q = 1$, which together with $B = 1$ implies P_{oo}; while the formulas $q = 1$, $B = 1/2$ have no solution that is also a solution of P_{oo}. This establishes the claim of strong evidence for P_{oo}; *mutatis mutandis* for P_o using E and E''.)

Exactly similarly, E provides strongly relevant evidence (relative to the theory whose postulates are these two) for hypothesis H_1. But the logical consequence H_2 thereof again fails to be even weakly testable relative to this theory. No possible evidence can provide relevant evidence for it.[17]

It would be, for me, a welcome surmise that we are dealing here only with a minor technical point, and that the defined notions of relevant evidence are merely meant to "generate" the correct relation of derelativized evidential support. The suggestion that what is needed in addition is some sort of "ancestral," or "inherited support," relationship, is surely one that must ocur at once to any reader at this point in the story. But I do not believe that this would be welcome to Glymour, given the uses he makes of these features of his account.

5. Comparison of Theories

As Glymour, and a number of others, have pointed out in discussion, a longer story is often more plausible or credible than a short one.[18] At the same time, even the most rudimentary (comparative, nonquantitative) discussions of probability entail that no matter how much A helps to support B, the conjunction (A & B) can be no more likely to be true than B alone (in the light of whatever evidence we have). We seem to have a difficult dilemma: either the way we reckon up the weight of evidence may favor hypotheses or theories less likely to be true, or else even the most rudimentary notions of probability are radically mistaken.

The dilemma is real, and we should embrace the first horn, not feel impaled by it. As I have argued elsewhere,[19] theory comparison and acceptance are a matter of decision making by, and balance striking between, conflicting criteria. The conflict between the desire for more informative theories (predictive power, explanation, empirical strength) with that for likelihood of truth is the main example. A well-designed test does not speak merely to the second desideratum: only powerful theories are in a position to pass powerful tests. Tallying up the support of relevant tests is a way of gathering reasons for the acceptance of theory on several counts at once. (It follows as a corollary that acceptance is not simply belief, for nothing counts as a reason for belief, as such, that detracts from the likelihood of truth.) To reconcile ourselves further to this choice of one horn of a dilemma, here is a more circumstantial account of just how longer, more circumstantial, stories may be more credible than short ones.

If tests are a way of gathering support of several sorts, then it is not surprising that a theory with features that detract from likelihood of truth may do better. In a certain laboratory, small water samples may be more likely to be hot than large ones at a certain time. Yet more of the large ones may yield a reading above 90°C than the small ones on massive, cold thermometers. In this case the thermometer reading reflects several features of the sample, and although the thermometer is the official method of registering hotness, a certain feature that in this case makes hotness less likely, also makes an accurate temperature test more feasible.

This takes care of one way in which theories that are more credible may yet be less likely to be true (for "more credible," in this case, read "better tested.") The second way is brought out by an example due to Nancy Cartwright.

A rumor that undergraduate women will be permitted to live off-campus as of age 21 (circulated at Pittsburgh in 1963) may not find much credence. But if this rumor is enlarged to include a supposed reason for this change in relations, such as that the university finds itself in the difficult situation of having overadmitted undergraduates, it will be much more believable.

The intuitive force of the example is clear. I am not sure that it supports Glymour's contention, because although the conclusion is the same, the apparent cause of the increased plausibility is quite different. In Glymour's case, the enlarged theory is more testable. In Cartwright's example, the enlarged theory is more plausible because a reason or cause for the main prediction is included. Although that reason is itself independently confirmable (and presumably still only an unconfirmed rumor), it shows how the asserted change could have come about.

We can imagine therefore that the standard reaction to a hypothesis has several stages. The first stage is the question, "How could that be?" Until we have satisfied ourselves that there is indeed some plausible way the event in question could come about, we refuse to go to the testing and evaluation stage at all. This transition from one stage to the next, however, does not require enlarging the theory: it requires only that some way of enlarging the account so as to embed the theory in a plausible story is possible. We can imagine that the undergraduates' first reaction is incredulity, because they can see no way the university, not known for its liberal innovations, could have come to such a decision. This lack on their part is merely one of imagination, however (in more technical cases it may be our well-known lack of logical omniscience), and the theory can be taken seriously when that obstacle is removed.

Theory comparison, theory choice, theory acceptance are not merely a matter of seeing which is most likely to be true. Glymour's penetrating analysis of testing has shown, in my view at any rate, that neither is evidential support merely a matter of gauging likelihood of truth in the light of the evidence. In this way, Glymour's analysis destroys the whole basis for what used to be known as confirmation theory. The correct reading is not that testing can provide us with more reason to believe an audacious theory than it provides for belief in the empirical adequacy thereof, but rather that it can give us other sorts of reasons for acceptance.

Notes

1. Evidence itself is an abstraction that should be more controversial than it is; the confrontation between orthodox and Bayesian views may yet make it so. From several points of view, all we have to play the role of evidence is some distribution of comparative certainties

and uncertainties over a set of propositions. Use of an idealization that takes such propositions as representing the evidence needs justification, especially if that distribution is a function of feedback from theories considered in its light.

2. *Theory and Evidence* (Princeton: Princeton University Press, 1980). See also his Explanations, Tests, Unity, and Necessity. *Nous* 14 (1980): 31-50, and Bootstraps and Probabilities. *Journal of Philosophy*, 77 (1980): 691-649 (with *erratum* 78 (1981), page 58).

3. Glymour on Evidence and Explanation, this volume.

4. *Theory and Evidence*, pp. 226-263, and M. Gardner, Realism and Instrumentalism in the Nineteenth Century. *Philosophy of Science* 46 (1979): 1-34. As far as the main thrust of Gardner's paper is concerned, I regard it of course as based on naive identification of debates about whether to accept the atomic theory, with ones about whether or not to believe it to be true. I call this naive not because of identification itself is, but because Gardner is ostensibly addressing the issue of realism versus (what he calls) instrumentalism, which is trivialized by that identification.

5. Semantic analysis of physical theory harbors several approaches, some more extensionalist and some less (Suppes's work being an example of the former). In *The Scientific Image* (Oxford: Oxford University Press, 1980) I have given a number of references in notes 22 and 23 to Chapter 3 (page 221), and note 29 to Chapter 6 (page 228).

6. See *Theory and Evidence*, pp. 111-123.

7. Specifically, the following definition is equivalent to (2-3):

D*. E provides *strongly relevant evidence* for A relative to T exactly if there is a set of quantities Q*, an alternative E′ to E, and a subset T_q of T for each q in Q* (whose union is a subset T_0 of T) such that:

(1) $T \cup E$ has a solution
(2) All solutions of $T_q \cup E$ assign the same value, r_q, to q
(3) All solutions of $\{q = r_q : q \in Q^*\}$ are solutions of A
(4) $T_0 \cup E'$ has a solution
(5) All solutions of $T_q \cup E'$ assign the same value r'_q, to q
(6) No solution of $\{q = r'_q : q \in Q^*\}$ is a solution of A
(7) $Q(A) \subseteq Q^*$

The idea of this definition is that for each q, we compute a precise value from E via a set T_q of theoretical hypotheses; and this same computation leads from E′ to an alternative value for q; the former imply an instance of A and the latter an instance of some contrary of A. But different subsets of T_0 cannot lead from E (or from E′) to different precise values of q, on pain of inconsistency. Second, placing a constraint on a set of solutions is really equivalent to saying that they are all extendable to some quantity so as to yield the same value for *that*. (For example, the constraint $q \geq 0$ is equivalent to implication that $f_{[0,\infty]}q$ has a value 1 (where f_X is the characteristic function for X).)

8. See *Theory and Evidence*, p. 131; this clause serves to evade the difficulty discussed on p. 132. Suppose for example that T has postulates (where A, B, C are measurable):

T1. $E - 1 = 0$
T2. $E = A \cdot B$
T3. $C = 1$

The data base $I = \{A = 1, B = 1\}$ provides strongly relevant evidence for T1 (using T2). But also, using only T2 and T3, I can provide relevant evidence for the hypothesis H: $E - C = 0$. This is analogous to the problem example discussed on page 132; a problem since we don't suppose that introducing theoretical parameter E in this fashion allows us to confirm a hypothesis about the color of swans by looking at black shoes. If we restrict ourselves to equational theories, all basic propositions can be put in the form $f(\text{—}) = 1$ using characteristic functions, and conjunctions can be captured: $X = 1$ and $Y = 1$ exactly if $X \cdot Y = 1$. Thus we can provide a parallel to clause (v). If we allow inequalities, we can use that manoeuvre, via characteristic functions applied to quantities as in the preceding note.

9. *Theory and Evidence*, pp. 139-142.

10. Glymour on Evidence and Explanation, this volume.

11. It is clearly important to see how Glymour's account accords with the history of science, and more specifically whether it throws light on puzzling developments. Glymour supports his account of testing this way through a large and varied class of historical examples. To investigate logical interconnections, thought experiments and fictional histories seem to me to be more appropriate. Especially in the case of the relations between theory comparison and evidence, the historical example method leaves much to be desired. To be shown several examples in which a later theory (a) is in retrospect superior, and (b) represents a gain in evidential support by a certain criterion, established little or nothing about that criterion. Should anyone argue for the curative powers of Vitamin C in that way, Glymour could be the first to show his error.

I do not mean to suggest that Glymour's own support for his theory takes that naive form. On the contrary, he gives us logical analyses of the structure of supporting arguments that cite evidence, in the case of figures as diverse as Copernicus and Freud, so perceptive and captivating that we can immediately perceive the evident value of his methodological insights. But even the best historical support must remain an illustration: we are no more able to deduce the principles of methodology from observed phenomena of scientific activity than Newton could deduce the laws of motion from observed physical phenomena.

12. *Theory and Evidence*, p. 161-167; see also his paper in *Nous* cited in note 2 above, and my response in the paper cited in note 3.

13. *Theory and Evidence*, pp. 152-155.

14. This follows at once from the definition. Note that consistency of T and E means simply that $T \cup E$ has a solution.

15. *Theory and Evidence*, p. 141.

16. This appears to contradict the opinion stated in *Theory and Evidence*, sentence straddling pp. 153-154, whence I derive the notion of optimally tested theory.

17. I must add here that I have not exploited a possibility that Glymour mentions: that a hypothesis A itself be used in the calculation of quantities, from a data base, to provide relevant evidence for A. Any defects or surprising features of the account following therefrom, I would tend to discount, because Glymour would lose nothing if he decided after all to forbid it. The example he gives to support it seems to me to involve an equivocation. Suppose we test $PV = rT$ by:

(a) first measuring P, V, T at time t, and calculating a value r_0 for r by means of that hypothesis

(b) secondly measuring P, V, T at a later time, and confirming that $PV = r_0T$.

We certainly have a test here of something, but I submit that it is a test of the entailed hypothesis that the value of PV/T is constant in time. If it be insisted that r itself is a physical quantity, then we are testing a little theory consisting of two postulates:

1. $(t)\,(t')\,(r(t) = r(t'))$
2. $(t)\,(P(t)V(t) = r(t)T(t))$

and we can use evidence plus either hypothesis to confirm the other. In neither case do we confirm one hypothesis via a calculation which uses that hypothesis. The same point is made by Aron Edidin, this volume.

18. Wesley Salmon, Susan Hollander, Nancy Cartwright, and Ian Hacking (in the order in which they made these comments).

19. Glymour on Evidence and Explanation, this volume.