Bas C. van Fraassen

Glymour on Evidence and Explanation

In Chapter VI of *Theory and Evidence* (specifically pages 199-203) and in a subsequent paper, Clark Glymour develops an account of scientific explanation to go with his theory of relevant evidence.[1] Especially significant for me is his use of these ideas in support of his contention that we can have more reason to believe one theory to be true than another even in cases in which the two theories are empirically equivalent. For that contention poses a challenge to the empiricist account of science that I have proposed.[2] Although I do not conceive of the empiricist-realist debate concerning science as addressed directly to an epistemological issue, I am acutely conscious of the fact that the empiricist view of what science is, will not be ultimately tenable in the absence of a tenable epistemology of a certain sort. And so, in response to Glymour's views on explanation and its relation to confirmation, I shall sketch here a preliminary version of such an epistemological position.

Glymour has put forward an account of three topics relevant to our present concern: hypothesis testing and confirmation, theory comparison, and explanation. The first, his theory of testing and relevant evidence, I admire greatly. In a companion paper to this one, I discussed that first topic in more detail, and below I shall draw a little on results of that discussion.[3] Here I shall sketch an alternative account of theory comparison and acceptance, as a case of decision making in the face of conflicting desiderata, and use that to introduce some objections to Glymour's account of explanation as well.

1. Theoretical Virtures; a Story of Conflict

The virtues that may be attributed to a scientific theory in order to support its acceptance, are diverse. I cannot attempt a complete typology, but I shall point to two important sorts. The first I shall call *confirmational* virtues; they are features that give us more reason to believe this theory (or

Research for this paper was supported by an NSF grant which is hereby gratefully acknowledged. The paper was first presented at a symposium with Clark Glymour and Wesley Salmon at the American Philosophical Association, Western Division, Detroit 1980.

part of it) to be true. This is equivalent, I take it, to the assertion that they are features that make the theory (or part of it) more likely to be true. The second sort I shall call *informational* virtues: one theory may be able to tell us more about what the world is like than another, or it may be able to tell us about parts of the world concerning which the other theory is quiet.

Before putting this division to use, some preliminary remarks. I have described both sorts of virtue in the comparative; for it may be that we can never do better than compare theories with each other. Let us be careful to assume none of the theories philosophers and logicians have proposed concerning these comparisons. There are, for example, theories of confirmation and of information that imply the applicability of quantitative measures, or at least linear orderings of degrees of confirmation and amounts of information. When I speak of more or less confirmation or information, my usage will be at odds with that. For if there are quantitative measures of these concepts, then one theory may have greater confirmation (overall!) or contain more information (in toto) even though some specific evidence confirms it less than another, or even though it is quiet about some phenomena that the other tells us much about. In placing some emphasis on this, I place myself on the side of Glymour on at least one important issue.

Second, I identified the two sorts of virtues by means of very different criteria. For the first is identified by what it does for the theory—making it more likely to be true—and the second by what it consists in—information the theory gives about this subject or that. Hence there is prima facie no reason to deny that the two sorts may overlap.

Without saying very much about what theory acceptance is, at this point, I take it for granted that a virtue of a theory is a feature that, if we are told that the theory has it, gives us more reason to *accept* that theory. Virtues are attributed in order to provide reasons for acceptance. One question Glymour and I differ on is whether acceptance of a theory is belief that the theory is true. If it is then all virtues are automatically confirmational virtues. But there is a strong tradition in philosophy of science, so prevalent that it does not even have a name, which speaks against this. Almost all writers on confirmation or evidential support make explicit some such principle as the following:

(1.1) If theory T provides information that T′ does not provide, and not conversely, then T is no more likely to be true than T′.

There are of course measures of how well the theory accounts for the evidence, or is corroborated by it, or whatever, which may be increased when a theory provides more information. But this makes them suspect as reasons for belief. Belief is always belief to be true; I take that as part of the logic of these words:

(1.2) A feature of T cannot provide more reason to believe that T unless it makes T more likely to be true.

Perhaps we should emphasize at this point that assessment of whether a theory provides more information or whether we have more reason to believe it, is always made in the light of background assumptions. These are meant to be kept fixed in principles 1 and 2. Thus a theory may well provide more information from a logical point of view (i.e., ignoring our present background theories) and yet be more likely to be true (in the light of that background); but such equivocation must be avoided. From principles 1 and 2 we must of course deduce

(1.3) Informational virtues are not confirmational.

This is a general conclusion; it does not imply that if the *overall* amount of information is greater in one theory than in another, then the former must be less likely to be true. For it may be that the theories contradict one another, so that the former does not represent *merely* an increase in *overall* information. If one theory provides all the information that another gives, however, and some more in addition, then the second cannot be less likely to be true. A corollary to this is:

(1.4) A theory cannot be more likely to be true than any of its parts; and the relation between the whole theory and the evidence cannot provide better reason to believe it, than the relation between a part of the theory and the same evidence, provides to believe that part.

This is a controversial conclusion from not very controversial-looking premises. It may be sugared a little by the reflections that for the whole theory we can provide a much larger list of relevant considerations, such as tests that it has passed. The point is only that *whatever procedure we then adopt to reach an overall comparison* of how likely it is to be true in view of the evidence must, to satisfy principle 1, give at least as good a grade to any part of that theory.

Finally, in view of the intimate link between virtues and acceptance, we derive the corollary:

(1.5) Some reasons to accept a theory are not reasons to believe it to be true; hence acceptance is not the same as belief; any reason for belief is a fortiori a reason for acceptance, but not conversely.

The situation is in fact a bit worse than I formulated explicitly in principle 1: there is in general a conflict between the desire for information and the desire for truth. (This was the basic theme of Isaac Levi's *Gambling with Truth* and is explored there at length, though by means of quantitative measures that would, in the present context of theory evaluation, not be used by either Glymour or myself.)[4] Thus the problem of theory acceptance has the structure of a practical decision-making problem: conflicting desiderata must be weighed against one another.

Indeed, this fits well with my own view on theory acceptance, which is that it consists in (a) belief that the theory is empirically adequate, and (b) commitment to use of the theory's conceptual scheme as guide to further research. Since we are never so lucky as to get a theory of truly universal scope and completeness, point (a) does not make point (b) inevitable or even automatically advisable. In addition, it may be more rational to pursue a theory that is capable of testing in the short run, or capable of combination with other theories already being pursued. And finally, the commitment may be made while the belief is still tentative or qualified, in which case the acceptance is also called tentative or qualified.

2. Explanation: What Sort of Virtue Is It?

Wesley Salmon and I have theories of explanation that are in some ways as opposite as can be. In explaining explanation, he looks to objective relations among facts correctly depicted by theory, whereas I look to context-dependent relations among questions, answers, and accepted theories.[5] But on one central point we are in complete agreement: we agree that to give an explanation is to give relevant information. For him, the criteria of relevance and informativeness are combined in relations of objective statistical correlation and spatio-temporal connection. For me, the relevance is initially determined contextually, but then statistical relations (implied by accepted background theories) may serve to evaluate relevant answers, which count as explanations, as good or comparatively better. In my account, these statistical criteria of evaluation are offered

only tentatively, and have in any case not a very central role. But the main agreement stands: for both of us, explanation is an informational virtue.

Although I do not know precisely what role Salmon would assign to explanation in the theory of evidence, my own position will be clear from the foregoing. That a theory provides explanations for phenomena in its intended scope is indeed a virtue, and gives us reason to accept the theory; but being informational, it generally gives no added reason for belief. Glymour's paper is devoted to the contrary thesis that explanation does give reason for belief. But then, his theory of explanation is quite different: for he characterizes explanation not in terms of information, but in terms of the elimination of contingency and chaos.

The first question I want to ask is whether explanation, as characterized by Glymour, is perhaps not after all an informational virtue. Since his theory of explanation is new, it does not yet provide explicit answers to some questions of a formal or procedural sort. For example, if asked what *is* an explanation, Hempel would say, an argument, and I would say, an answer to a why-question. The locution "theory T explains fact E" Hempel would gloss as "there is an explanation with the conclusion that E is the case, whose theoretical premises are furnished by T." I would gloss it as "once you accept theory T, you are in a good position to answer the question *why E?* (understood as asked in the present context)." Thus Newton's theory explains the tides in the sense that, once we accept Newton's theory we can say (correctly by our lights) that the sea is subject to tides because of the gravitational influence of the moon, and (still relative to our thus enlarged background) this is a good answer; holders of other theories are not in a position to say that, or at least are not able to give an equally good answer by their lights. I do not quite know how to regiment such terminology to accord with Glymour's theory of explanation. But I think I do know quite well how to handle examples in accordance with it in the way he does, and this should suffice.

In the remainder of this chapter I shall therefore examine the account of explanation via identifications, and also the way in which a theory may unify our description of hitherto unconnected phenomena, both of which Glymour gives as instances of theoretical explanation. The first specific idea I mean to explore is this: perhaps Glymour agrees by and large with section 1 above, and disagrees with the thesis that giving an explanation must always (or ever) consist at least in part in giving information (or more specifically with the thesis that if we amend a theory so as to make it more

explanatory, we must be making it more informative). I shall argue that this cannot be so, for explanation, elimination of contingency, and unification, in his sense, are all bought at the expense of making the theory more informative.

3. Explanation from Identity and Necessity

We may imagine the following process, which is not too farfetched as a historical account: first a theory is developed in which temperature, pressure, volume, and all the mechanical parameters appear independently, and which contains the molecular hypothesis identifying bodies with aggregates of molecules. Call this theory T. Then a theoretical improvement is introduced. We add the identity:

(3.1) if X is aggregate A of molecules, then the temperature of X is identical with the mean kinetic energy of the molecules in A.

Since every body is such an aggregate (a single molecule being a small aggregate), this identifies the temperature of every body. Thus (3.1) can be rephrased as

(3.2) Temperature is identical with mean molecular kinetic energy.

I do not use the identity sign, for that has in science the typical meaning of mere equality of values. That the values of the two quantities are always the same is already implied by old theory T. That is exactly the fact which we wanted to have explained. Accepting new theory T', which is T plus (3.2), we can explain this universal equality by pointing to an identity that implies it.

At this point Glymour says that if the identity (3.2) were contingent, then it would need an explanation as much as the equality itself did. But it needs no explanation; the requests for explanation stop here. From these two considerations we can deduce that the identity is not contingent. Although this conclusion would have been controversial some years ago, it is much less now. Of course, the identity is not a priori true or false; we cannot possibly give a proof to demonstrate or to refute it. But this is merely an epistemic fact. Glymour can appeal to the theories of Kripke and Putnam to support his view that a priori truth and necessity do not coincide; and that all such identities are necessary if true. Which explains nicely why requests for explanation should stop there, at least within the confines of physics.

But the very distinction between necessity and the a priori establishes that if it is necessary that A, it may still be informative to assert that A.

Indeed, the only information we have before we are told by theory T′ that (3.2) is the case, is that (3.2) is either necessary or impossible. We do not know which. To be told that it is true, and hence necessary, conveys a great deal of information. Therefore I would say that, explanatory though T′ is, it is less likely to be true than its part T. And the fact that T′ is more explanatory than T, does not constitute a reason to believe it, a fortiori.

I have tried to think of this in terms of Glymour's theory of testing. I understand that theory best if it is applied to a theory like the present one, which is stated initially as a set of functional relationships among physical parameters, some of which are directly measurable and some not. But if we add (3.2), then any Glymourian test I can think of for that hypothesis is no more than a test of the equality

(3.3) For all bodies X, the temperature of X is (has the same value as) the mean kinetic energy of the molecular aggregate which is X.

It may be as unfair, however, to ask for a test of a necessary identity as it is to ask for an explanation.

Perhaps we should look instead at the inevitable simplification that acceptance of such identity brings in its train. One term, the simpler one (in this case, "temperature") is deleted from the primitive vocabulary in the description of the theory and reintroduced by definition. Does the new theory—call it T′′—contain more or less information than the original? Here we face a dilemma, it seems to me, depending on how we construe what the definitional extension says. The definition is used to deduce that procedures heretofore called measurements of temperature are measurements of mean kinetic energy. The original theory said only that they yielded numbers equal to the value of that quantity. Thus viewed, the definitional extension is informative, and indeed, explains this further equality. But alternatively we can say that all information about temperature has been dropped from the theory since a definition is merely a stipulation as to the use of the word hence forth. In that case T′′ has less information than T, but it also does not explain anything that T did not explain. If the concept of temperature is truly dropped and disappears, then fewer why-questions can be asked; and this may be an advantage, but the identification to which explanatory power was attributed has also disappeared.

To sum up, then, it appears that the explanation of a universal equality of values by appeal to a theoretical identification is possible only when that identification is an informative statement.

4. Explanatory Unification

Newton's theory covers the laws of fall, planetary motion, pendulums, tides, comets, and much else. He brought unity to the field of physical science. We also say that his theory *explains* the phenomena, the motions, that I have just listed. Glymour suggests that the two assertions are not unconnected; that the unification constitutes the explanation. Thus one might say: Newton's theory explains the tides, because Newton's theory exhibits the tides as one instance of a general pattern that has as instances a number of phenomena heretofore considered diverse.

I find it very hard to evaluate this claim, since it is not supported by a claim that all explanations are thus (explanations by theoretical identification are not, for example), and I don't know how to search systematically for a random sample of theories that unify diverse phenomena to see whether the connection persists in general. But it does seem to me that unification requires the introduction of additional information.

Let us again use a fictional history. Suppose that in one country, people develop theories of fall, planetary motion, tides, comets, and pendulums, and that the laws of motion they state for these are *exactly* those implied by Newton's theory. Then a quite different person, call him Newton*, is born into this country, sees all those diverse theories, and proposes Newton's theory. All the claims about unification and explanation presumably still hold. Yet Newton's theory entails a great deal that the sum of the diverse theories does not. It entails, for example, laws of planetary motion for solar systems other than this one, and tides for planets with moons different from ours. Indeed, just because he does not merely unify the phenomena we know already, but because he unifies them together with a large range of unexplored and unobserved phenomena, do we say he truly unifies them. Thus the addition of information appears to be crucial to the process of unification.

5. Testing and Comparision of Theories

I have so far examined the one alternative: that Glymour will accept the basic epistemological principles I stated at the outset, while denying that explanation is an informational virtue. It appears that this position would not be tenable. Hence it is high time to examine the second alternative, that Glymour will deny those principles. The crucial ones are (1.1) and (1.2). It is possible to deny both, or else to deny the second while affirming the first, provided one *denies*:

(5-1) It is inconsistent to say that you have more (respectively, less) reason to believe T while saying that T is less (respectively, more) likely to be true.

(with each part relativized, as always, to the light of accepted background evidence and theories). In order to explore this second alternative, we must of course look to Glymour's writings, and especially his book, to see what he says about this. The matter is, however, not clear there.

To begin, Glymour does not hold with rules for acceptance, except for the trivial one that we ought not to accept a theory if we know of a better one (Chapter V, pages 152-155, "Comparing Theories"). This reduces the problem of acceptance, which I take him to equate with that of belief, to the problem of comparing theories. Glymour's account of testing and relevant evidence focuses on the notion:

(5-2) E confirms hypothesis A relative to Theory T

where E is a body of evidence. (In another paper I examine, or rather reconstruct, that account, and give warrant there for the claims I shall make about it here.[3])

In order to say anything at all about having better reasons to believe one theory rather than another on a given occasion, Glymour *must* go beyond the purely relative notion (5.2). He must "derelativize" if we are to have a comparison of theories with each other. A naive suggestion would be to compare the hypothesis of two given theories with respect to confirmation received relative to previously accepted theory. This would make nonsense of Glymour's general account. For his central argument is that when a theory is advocated, it is on the basis of successful tests of hypotheses of that theory relative to that theory. The wealth of examples, and the sophisticated analysis of the logical complexities of this procedure, that Glymour gives us, convince me that he is right. The bootstrap account of testing is a major achievement.

Glymour makes various suggestions for the correct "derelativization," both explicit and implicit, and I shall examine what I take to be the main one. In all cases, the burning question for us here is how the suggested comparison relates to increase or decrease in information. I shall quote, to begin, his fifth explicit suggestion:

(5-3) It is better that a body of evidence test a set of hypotheses sufficient to entail the whole of the theory than that it test a

> logically weaker set, for if the logically sufficient set can be established the rest will follow. (*Theory and Evidence*, pp. 153-154)

The content and general theme of the book make it clear that the testing in question is to be confirmation of those hypotheses relative to the theory itself. And it appears, in this passage, that confirmation is inherited by the theorems from the axioms. But this is not so in the strict sense in which Glymour explicates confirmation. The following two facts are crucial features of the account, which are used to defuse philosophical problems elsewhere in the book:

(5-4) If E confirms A relative to T, and A logically implies B, it does *not* follow that E confirms B relative to T.

(5-5) If E confirms A relative to T, and also confirms B relative to T, it does *not* follow that E confirms (A & B) relative to T.

(See my other paper for examples; it is shown there also that various obvious weakenings of the denied principles fare no better.)

It looked for a moment as if Glymour were prepared to accept some such principles as I stated at the outset; now it seems that this appearance was deceptive. But then, what warrant can he claim for the quoted passage (5.3)? It is clear that "the rest will follow" in the sense that the theorems are as likely to be true, at least, as the axioms. But the crucial question is whether the extent to which they are confirmed by the evidence is as great.

At this point I am tempted to suggest the introduction of an ancestral to the confirmation relation. We could say that although the theorems need not be confirmed, we nevertheless have as much reason to believe them, because of the support they *inherit* from the axioms. I find it difficult to conceive of any other sort of warrant for the quoted passage; but if Glymour says this, he will accept the epistemological principles I set out to begin, it seems, and our second alternative has run out.

A later passage appears to rule out this maneuver anyway. One recurring theme for Glymour has been the assertion that the evidence may provide better support for a theory than for the observational consequences of that theory taken by themselves (Chapter V, pages 161-167, "The Theoretician's Dilemma"). I am champing at the bit here to argue that nothing like what Glymour and others call the observational consequences constitutes anything like the empirical import of the theory. But that is irrelevant for the present dispute; it suffices to read Glymour as saying here that a certain

subtheory of a theory comes out the loser in the comparision between the two. It is clear that the theory entails its subtheory; hence we have here, it appears, a clear denial of the principle that we have at least as much reason to believe the theorems as we have to believe the axioms. The property that the whole theory has and the subtheory lacks is this:

(5-6) There is a body of evidence E, and a complete set A of axioms for T such that E confirms each member of A relative to T itself.

The principle we can apparently glean from what Glymour has said so far, then, is this:

(5-7) If we have no evidence that disconfirms any theorem of T relative to T, while T but not T′ has property (5.6), then T is better confirmed than T′.

Now I would deny this, because in the case in which T′ is a subtheory of T, hence consists of theorems of T, I would say that as far as any nonrelative evidential support we have for theories on that occasion, is concerned, logical consequences inherit support from confirmed hypotheses. We can read (5.7) as saying that there is no inherited support. If a theory does very well in testing, we cannot cite that as evidence for the proposition that any of its consequences are true, at least not until we take the great leap and decide to believe that the theory itself is true. (For belief, presumably, is inherited!)

I find this very surprising. I should think that if a theory is doing well, that is good evidence that certain relevant parts are true, the evidence being the better when the part is more circumscribed. But a clash of intuitions won't get us anywhere. Instead I must ask Glymour to confront the problem I have with (5.3) and its apparent consequence (5.7). The only sort of reason I can think of to warrant the former is at odds with the latter. In the nonrelative or derelativized sense of evidential support (which will be needed if we are to compare theories with each other), it seems to me sensible to say that evidential support for the axioms is especially important, *only if* we also say that this (nonrelative) support is inherited by their consequences. But in *that* case, I can see no way in which we shall ever have less reason to believe a subtheory than we have to believe the whole. All I can see—and I urge this on Glymour as the correct diagnosis—is that we might have less reason to *accept* the subtheory (questions of belief left aside), because, taken as a theory, it has considerably fewer virtues than its parent does.

6. The Real Significance of a Test

If theory comparison and acceptance are a matter of striking a balance between competing desiderata, we should expect a procedure designed to elicit simultaneously the strengths of the theory on several different counts. That is exactly how I see the testing procedure, and Glymour's account I take as support for my view. To pass a well-designed test provides confirmation and evidential support for a theory, but to admit such a test in the first place the theory must be sufficiently informative. Hence to say of a theory that it has passed well-designed tests speaks simultaneously to the two competing criteria of confirmation and information, and the support testing provides is not to be equated with simple confirmation.[6]

Notes

1. Clark Glymour, *Theory and Evidence* (Princeton University Press, 1980) and Explanations, Tests, Unity and Necessity. *Nous, 14* (1980) 31-50.

2. *The Scientific Image* (Oxford: Oxford University Press, 1980).

3. Theory Comparison and Relevant Evidence, this volume.

4. Isaac Levi, *Gambling with Truth* (New York: Knopf, 1967; Cambridge, Mass.: MIT Press, 1973); see also his *The Enterprise of Knowledge* (Cambridge, Mass.: MIT Press, 1980).

5. See Chapter V of *The Scientific Image* where Salmon's theories are discussed in sections 2.2 and 2.6.

6. *Note on inheritance of support*. Should evidential support for a theory be inherited by its consequences? A few logical distinctions help. Let us call a judgment *diachronic* if it takes the form "T is now more X than formerly." There "X" may stand for "believed" or "supported by the available evidence" etc. Clearly there is no inheritance in the diachronic case, since the change in attitude or evidence may have to do directly with one part of T and not another. Call the judgment *synchronic* if it compares several theories from a single historical point of reference. Then there is obvious inheritance in the case of belief (if (A&B) is more strongly believed than C, then so is A), and obvious non-inheritance for the effect of *new* evidence (same reason as in the diachronic case). The final disputed case concerns therefore a synchronic judgment concerning support by the total evidence: could the total evidence on a given occasion support a theory more than one of its parts? Note that when Bayesians define "E confirms H" as "$P(H|E) > P(H)$" their concern is either with the diachronic or synchronic/new evidence case.