Minnesota Studies in the PHILOSOPHY OF SCIENCE

HERBERT FEIGL, FOUNDING EDITOR

VOLUME X

Testing Scientific Theories

EDITED BY

JOHN EARMAN

UNIVERSITY OF MINNESOTA PRESS, MINNEAPOLIS

Roger Rosenkrantz

Why Glymour Is a Bayesian

In the third chapter of his book *Theory and Evidence*, Clark Glymour explains why he is not a Bayesian. I shall attempt to show, on the contrary, that he is a Bayesian, more so than many who march under that banner.

1. Bootstrapping and Bayesian Inference

The central problem his book addresses is to explain how findings in one (observational) vocabulary can evidence propositions stated in a different (theoretical) vocabulary. The solution offered is that a hypothesis is confirmed with respect to a theory by deducing instances of that hypothesis from the evidence and assumptions of the theory, where these assumptions may include the very hypothesis under consideration. (It is the latter feature that leads Glymour to label the procedure "bootstrapping.") Confirmation is thus a ternary relation linking a bit of evidence e to a hypothesis h by way of a background theory T. In addition, Glymour requires that the observation or experiment that issues in e be such that it could have issued in a disconfirming, rather than a confirming, instance. In short, the experiment must place the hypothesis in jeopardy.

Both features are nicely illustrated by Glymour's discussion of the hypothesis h, common to Ptolemaic and Copernican astronomy, that a planet's period increases with its distance from the center of motion. His point is that h is testable (hence confirmable) relative to the Copernican theory but is not relative to the Ptolemaic. For in Copernican astronomy, observed planetary positions can be used to determine the relative distances of the planets from the sun. And using the earth's known period of 365.2425 days, the directly observable synodic periods (the times between successive superior conjunctions when earth, sun, and planet all lie on a line) determine the sidereal periods (or time for a complete circuit of the sun), and so the latter may be inferred from the observations as well. We find, for example, that the maximal elongation from the sun is larger for Venus than Mercury, whence Venus's orbit of the sun must contain

Mercury's. Then h predicts that Venus will have the longer (sidereal) period, and hence the longer synodic period. This prediction is borne out by observation. If, however, the observed synodic periods satisfied the reverse inequality, we would have instead a counterinstance of h. Relative to Copernican theory, then, the observed positions and synodic periods do place h in jeopardy. But as there is no like determination of the relative sizes of the planetary orbits in Ptolemaic astronomy, h cannot be tested or confirmed relative to that theory. Instead, Ptolemaic astronomers simply *assumed* h in order to fix the order of the planets.

That the hypothesis h stands in this relation to the two theories is clearly a result of the fact that relative distances from the center of motion are deducible from observations in the Copernican theory but not in the Ptolemaic. That is to say, it results from the greater *simplicity* or *overdetermination* of the Copernican theory. As we will see, greater overdetermination renders a theory more highly confirmable on Bayesian grounds. This already suggests a relation between Glymour's account of evidence and a Bayesian account very different from opposition, but let us look more closely.

For e to confirm h relative to T, Glymour first requires that e be an instance of h in Hempel's sense. Hempel's satisfaction criterion effectively equates con*firm*ing observations with con*form*ing observations, and is of course strongly at odds with a Bayesian account of confirmation based on positive relevance. From a Bayesian point of view, Hempel's "positive instances" are confirmatory only when they happen to be consequences or verified predictions of the hypothesis. This suggests opposition, and yet it is surely very striking that Glymour's examples are all most naturally interpreted as Hempelian instances of this more restricted kind! This is perfectly clear in the example from astronomy, in which we can imagine first ascertaining that the maximal elongation is greater for Venus than Mercury. Then the hypothesis h relating period to orbital radius predicts that Venus will be found to have a longer synodic period than Mercury (to overtake Earth less frequently). Therefore, using Copernican theory, the synodic periods are restricted to satisfy a simple inequality. Similarly, in the examples of theories formulated as equations (pp. 112 ff.), overdetermination of the theory expresses itself in the fact that different subsets of the equations can be solved to yield different determinations of a theoretical quantity, and the predictions are then of the form that these different determinations will agree. In all of these cases, talk of deducing an instance of a hypothesis from theoretical assumptions can be translated without loss into talk of verifying a prediction based on the hypothesis and the subsidiary assumptions.

Consider next the (Popperian) requirement that the observation or experiment place the hypothesis in jeopardy. As Glymour phrases it (p. 127), "the deduction is such that it does not guarantee that we would have gotten an instance of the hypothesis regardless of what the evidence might have been." That is, the relevant observation might have issued in other outcomes from which a counterinstance or disconfirming instance of the considered h would have been deducible, using T, as in the example from astronomy. We may think of the evidence e, therefore, as representing a particular subset of the allowed values of the observable quantities. Glymour's first condition is that the actually observed values do indeed fall in the allowed subset. His second (Popperian) condition is that the complement of the allowed subset be nonempty. If we equate possible outcomes with those of positive probability, his account of "e confirms h relative to T" comes to this:

(1.1) $P(e/h,T) = 1$ and $P(e/T) < 1$.

This looks very much like hypothetico-deduction (see Glymour's own formulation on p. 168).

More to the point, the two conditions of (1.1) are sufficient that e confirm h relative to T on a Bayesian account of confirmation. The second condition of (1.1) is definitely needed. Indeed, if we wish, more generally, to incorporate cases in which e is not a consequence of h and T, it is natural to replace (1.1) by the weaker condition:

(1.2) $P(e/T) < P(e/h,T)$,

which merely expresses the positive relevance of h to e against the theoretical background T. (Notice that (1.1) entails (1.2), but not conversely, although the second part of (1.1) is entailed by (1.2).)

Glymour hankers after conditions that further constrain the confirmation relation. As we shall see, his chief objection to hypothetico-deductive and Bayesian approaches is that they are too liberal, admitting as confirmatory items of evidence that we should not countenance as such. From this viewpoint, it is ironic that the Bayesian reconstruction of the bootstrapping argument just offered is far more restrictive than the one based on Hempelian confirmation, for Hempel's criterion, we have seen, is far less austere than the positive relevance criterion (1.2). And inasmuch as Glymour's own examples seem to depend only on Hempelian instances that happen to be verified predictions or consequences of the hypothesis, one would think that Glymour himself would prefer the Bayesian analysis of bootstrapping to the Hempelian.

In fact, he does express misgivings about Hempel's satisfaction criterion (in his closing chapter), pointing out that it does not permit confirmation of sentences having only infinite models or confirmation by "partial instances" (e.g., of "everything bears the relation R to everything" by "*a* bears R to *b*"). Yet these criticisms suggest that Hempel's criterion is too narrow, whereas one would have thought that it is too broad, as shown, for example, by the paradoxes of confirmation. At any rate, in a paper that has since appeared (Glymour 1981), Glymour expands some of the replies he offered to the version of this paper presented at the conference. He shows how to connect bootstrapping to a Bayesian account (essentially as above) but continues to insist that Bayesian methods are too permissive. The main thrust of the paper is to deny what I had argued in my original presentation: that bootstrapping reduces to a nuts-and-bolts form of Bayesian confirmation theory.

2. Is Bayesian Methodology Too Weak?

Glymour is hardly the first to press this line of criticism or urge that an adequate methodology should impose additional strictures of a non-Bayesian kind. Before I take up his specific objections in detail, it may be well to look briefly at some earlier criticisms of a similar kind. This will not only set Glymour's reservations in better perspective, but it will allow us to highlight additional parallels between his account of evidential support and the present author's.

(a) High content versus high probability

Perhaps the most noteworthy previous attempt to show that Bayesian methodology is too liberal comes from Sir Karl Popper. His chief criticism seems to be that Bayesians cannot account for the demand for content. For if high probability is the *ens realisimum* of inquiry, it is best attained by putting forth theories of low content that run a minimal risk of exposure to contrary or falsifying evidence. That confirmation is easily attained if sought, is a recurring theme in Popper. This accounts, he thinks, for the otherwise surprising "success" of Freudians and Marxists. And let us admit

that this criticism has real bite when applied to Hempelian confirmation. For if consistency with our hypotheses is all we demand, then confirmation is indeed easy to come by. The moral Popper draws is that genuine confirmation or support can be obtained only by running risks. And we run risks, first by putting forward "bold conjectures" or theories of high content, and second by subjecting our theoretical conjectures to stringent tests and searching criticism. In fact, Popper carries this line of thought right to its logical conclusion, insisting that confirmation can result only from a *sincere* attempt to overthrow or refute a conjecture. (We have already seen that this "nothing ventured nothing gained" philosophy is incorporated in Glymour's account of confirmation.) In resting content with nothing short of a sincere attempt at refutation, Popper enters the shadowy realm of the psycho-logistic. Although this may seem somewhat out of character, it is important to recognize this strain in his thinking, for we shall encounter it below in the writings of Popper's follower Imre Lakatos. From a Bayesian standpoint, it would be most natural to equate a stringent or sensitive test with one that has a low probability of issuing in a conforming outcome if in fact the conjecture is false. But Popper has been at best ambivalent about attempts to capture what he is saying in probabilistic terms.

Let me now sketch a Bayesian response to Popper's criticism, one that I have developed elsewhere in greater detail (especially in chapters 5-7 of Rosenkrantz 1977), although the present treatment contains important additions and qualifications.

To begin with, Popper's notion of content seems unduly narrow. Roughly, he equates a statement's content with the set of "basic statements" it logically excludes. In practice, though, a theory or model does not *logically* exclude any outcome of a relevant experiment. This is patently true of a probability model. As determined by a suitable statistical criterion of fit, the outcomes will be in only more or less good agreement with such a model. This will also be true of a deterministic model, for empirical study of such a model is always coupled with a probabilistic theory of errors of observation. Moreover, any theory, probabilistic or deterministic, will typically have adjustable parameters that must be estimated from the data used to test the theory. And the number of parameters that must be estimated is surely relevant to any assessment of a theory's content.

A natural way of extending Popper's notion to accommodate degrees of fit and numbers of parameters is to measure a theory's content (or simplicity, or overdetermination) *relative to a contemplated experiment* by the proportion of possible outcomes of the experiment that the theory "fits" by the lights of a chosen probabilistic criterion. I term this proportion the theory's *sample coverage* for the given experiment. And for theories with adjustable parameters, sample coverage is just the union of the sample coverages of the special cases of the theory obtained by assigning definite values to all free parameters. The smaller its sample coverage (i.e., the narrower the range of experimental findings it accommodates in a probabilistic sense), the greater a theory's content. And, I hasten to add, the contemplated experiment relative to which sample coverage is computed may be a composite experiment comprising several applications of the theory, or even its entire intended domain of applications.

The concept of sample coverage captures a good deal of what is packed into our ordinary understanding of content or simplicity. Thus quantitative theories are simpler (have more content) than their qualitative counterparts, and unifications of theory (e.g., of electricity and magnetism, Mendelian genetics and cytology, or quantum theory and relativity) represent (usually major) simplifications of theory, for experiments formerly regarded as independent then appear as highly dependent. Above all, we complicate a theory when we enlarge its stock of adjustable parameters, for each parameter we add extends the range of possible findings that the theory can accommodate. (It doesn't follow, however, that we can compare the content of two theories merely by counting parameters.) The explication of content in terms of sample coverage and the relativization to an experiment help us to avert some familiar difficulties, such as irrelevant conjunction (which I discuss below in connection with Glymour's critique of hypothetico-deductivism). But the really essential point is that by using a Bayesian index of support, we can show that simpler theories are more confirmable by conforming data—they have, so to speak, higher cognitive growth potential. And this already provides a partial answer to Popper's charge that Bayesians cannot explain our preference for content or simplicity.

To illustrate the connection, consider a composite hypothesis H with special cases $h_1, \ldots, h_n$. (H is the disjunction of the mutually exclusive h_i's.) Applying Bayes's formula,

$$P(H/e) = \Sigma_i P(h_i/e)$$
$$= \Sigma_i P(e/h_i)P(h_i)/P(e)$$

$$= [P(H)/P(e)][\Sigma_i P(e/h_i)\frac{P(h_i)}{P(H)}].$$

I call this the *generalized Bayes formula* and the bracketed quantity, which mirrors the evidence, the *average likelihood* of H. Thus

$$(2.1)\quad P(H/e) = \frac{P(H)}{P(e)}[\Sigma_i P(e/h_i)\frac{P(h_i)}{P(H)}]$$

expresses the posterior probability of H as the product of its prior probability by the average likelihood divided by P(e), which I term the *expectedness* of e. (Note: P(e) must always be computed relative to the considered partition of hypotheses.) In practice, of course, one has a continuum of special cases corresponding to different settings of a real-valued parameter (or vector of parameters), and then the summation of (2.1) gives way to an integral. Where the parameter is *freely* adjustable (i.e., where the theory itself gives no clue as to its value), an "uninformative" parameter distribution should be employed. In this way we impose the maximum penalty for lack of content. But in any case it is clear that this penalty will be higher when there are more special cases over which to average the likelihood. A simple example will make this clear.

Ptolemaic astronomy tells us that the center C of Venus's epicycle lies (at all times) on the line ES joining Earth and Sun, but it imposes no further constraint. (Even the constraint that C lies always on ES is rather ad hoc; it does not grow organically out of a geocentric conception but is inferred from observation.) Applied to Venus, the Copernican theory may be considered as the special case of the Ptolemaic that locates C at the point S, the center of the sun. Reflect for a moment on the contrast: one theory confines C to a line, the other to a single point of that line! To see the connection with support, let us look first at the situation in qualitative terms. Qualitatively, there are just three possibilities: (a) C lies close to S with S inside the epicycle, (b) S lies between E and C on line ES with S outside the epicycle, or (c) C lies between E and S with S outside the epicycle. As telescopic observation of the phases of Venus first disclosed, possibility (a) is realized. Hence the Copernican special case has an average likelihood of 1, and the Ptolemaic theory has an average likelihood of 1/3. This gives a "Bayes factor" (or ratio of average likelihoods) of 3:1 in favor of Copernicus. This is not very impressive, but if, in quantitative terms, we could show that C = S (within the limits of observational accuracy), the

Bayes factor in favor of Copernicus would be effectively infinite. For the average likelihood of the Ptolemaic theory would be close to zero (we would be integrating over the entire line ES and only special cases corresponding to settings of C close to S would have appreciable likelihoods). Historically, of course, the phases of Venus did not (and could not) show that C = S. I am drawing this comparison only to illustrate the incomparably greater cognitive growth potential of a simpler theory.

Notice that I have been using the average likelihood to compare a theory with a special case of itself. I see nothing wrong with that. Of course, if we wanted to compare the two in terms of probability, we should have to take logical differences, equating (in our example) the Copernican special case with the hypothesis C = S and the Ptolemaic alternative with C ≠ S. As removal of a single point does not affect an integral, the relevant average likelihoods would be the same. Failure to see this possibility seems to be most of what lies behind Popper's oft-repeated equation of simpler hypotheses with *less* probable hypotheses, and the consequent denial that one can account for the importance of simplicity by connecting it to probability.

To resume the main thread of argument, we have given a direct and compelling Bayesian reason for valuing high content and simplicity. Some Popperians will scoff, nevertheless, saying that we are just mimicking Popper's methodology in Bayesian terms, trying, as it were, to recreate the flavor of the gospel in the vulgar tongue. For Bayesians still seek high probability first and foremost, even if, coincidentally, the way to obtain it is to find the simplest theory that can be squared with the "hard" data. But the charge is unfounded. Granted that probability is the *yardstick* by which Bayesians compare rival conjectures, it doesn't follow that high probability is the *goal* of any scientific inquiry. The yardstick is simply the means by which we measure our progress towards the goal, whatever the goal may be. And for my own part, I am quite comfortable with Popper's identification of that goal as the attainment of ever more truthlike theories, i.e., of theories that are closer and closer to the truth. Moreover, highly truthlike theories are just those that combine a high degree of content with a high degree of accuracy—in I.J. Good's happy phrase, they are "improbably accurate"—and a precise explication can be given along Bayesian lines by equating a theory's truthlikeness with its expected support, i.e., its support averaged over the outcomes of a relevant experiment. Then a theory is close to the truth when it is strongly supported by those outcomes of the

experiment that are highly probable, conditional on the truth. Insofar as Bayesian support is a (determinate) blend of accuracy and content, the same will be true of our concept of truthlikeness. Again, the probabilistic explication appears to escape notorious difficulties associated with its more narrowly deductive Popperian counterpart (see Rosenkrantz 1980 for a fuller account), but these matters are somewhat peripheral to our present concerns.

Up to this point in our story, it may well appear that I am just offering a sort of Bayesification of Popper's notion of content. Significant differences emerge, however, when our accounts of the role simplicity plays in theorizing are compared.

Popper connects simplicity with falsifiability and quotes with approval William Kneale's remark that "the policy of assuming always the simplest hypothesis which accords with the known facts is that which will enable us to get rid of false hypotheses most quickly." (Popper 1959, p. 140) There is, to be sure, a pervasive equivocation in Popper on "falsifiability," which is used in both a semantical sense (namely, the number of basic statements a theory excludes) and a pragmatic sense (namely, the ease with which a false conjecture can be exposed as such). And it is not generally true that conjectures that are more falsifiable in the semantic sense are more readily disposed of. But perhaps this is a quibble. The more serious criticism levelled at Popper is that mere elimination of false pretenders does not necessarily leave one closer to the truth. For in theorizing, one seldom has an exhaustive list of theoretical possibilities at hand. Indeed, there is a certain temptation to stand on its head Popper's taunt that confirmation is easily obtained if sought, and maintain that it is rather falsification that is easily obtained if sought. One can easily imagine all sorts of Goodmanesque (gruelike) alternatives to well-established hypotheses that would be easy to falsify. At the very least, Popper unduly neglects considerations of plausibility in theory construction; and more than that, there is something seriously askew in his view that interesting truth is most efficiently attained via elimination of false conjectures. Perhaps we can best appreciate my misgivings by turning forthwith to a Bayesian account of these matters.

We must recognize, to begin with, that *Bayesian* confirmation or support is *not* easily obtained. For it requires both accuracy and simplicity. In fact, the ideal case is that in which the theory fits all and *only* those experimental outcomes that actually occur (e.g., just the actually observed frequencies with which the planets retrogress). From this perspective, it is not at all surprising to find that "particle physicists are in the habit of thinking that anything not expressly forbidden by nature is compulsory." (Calder 1979, p. 186) And in the same vein, C. Lanczos writes:

> In 1929 he [Einstein] talked of the "Promethean age of physics," in which one is no longer satisfied with the discovery of the laws of nature, but one wants to know why nature is the way it is *and cannot be anything else*... The impressive feature of Einstein's gravitational theory was that if one wanted to characterize a Riemannian geometry by the simplest set of field equations, one automatically arrived at Einstein's gravitational equations, which gave a complete explanation of Newtonian gravity, without the necessity of a special force of gravitation. (1967, pp. 185-186)

There is much more to efficient theorizing, however, than fitting all and only what occurs. For one thing, the "hard facts" vary in hardness, and it will often be impossible to accommodate all the mass of partially conflicting data. And, in any case, it seems advisable to begin with special cases of the complex system or process of study and "put in our ingredients one at a time." (Bartlett 1975)

What are some of the things to be said for starting with a deliberately oversimplified model? First, there is mathematical *tractability*. We can construct and explore the properties of simple models rather easily; highly complicated models may require techniques that lie beyond the present reach of mathematics. Second, there is *economy of effort*. About his search for an adequate model of DNA, James Watson writes:

> We could thus see no reason why we should not solve DNA in the same way. All we had to do was to construct a set of molecular models and begin to play—with luck, the structure would be a helix. Any other type of configuration would be much more complicated. Worrying about complications before ruling out the possibility that the answer was simple would have been damned foolishness. Pauling never got anywhere by seeking out messes... (1968, pp. 47-48)

And later he adds:

> Finally over coffee I admitted that my reluctance to place the bases inside partially arose from the suspicion that it would be possible to build an almost infinite number of models of this type. (p. 139)

A third advantage that springs to mind is *feedback*. A workable model of even a highly schematic version of the system studied provides information about how the full system works in special circumstances or when certain variables are controlled or confined to subsets of their allowable ranges, and this allows the model builder to see precisely how his simplified model breaks down when complicating factors are introduced. This provides insight into what sorts of complications will most dramatically improve goodness-of-fit.

In sharp contrast to Popper's account, then, far from aiming at rejection of false theoretical alternatives, theoreticians seek a model that works tolerably well in circumscribed contexts (a sort of "first approximation") and then ("putting in their ingredients one at a time") seek ways of complicating or refining the picture to capture "second-order effects" or finer details. In short, the development of a theory occurs less by eliminative induction than by successive approximation or "structured focusing." And Popper's account is weakest in describing what might be called the "developmental phase." Popper and Lakatos demand that a "progressive" modification of theory increase testability and content and that some of the excess content be corroborated. But, almost by definition, a complication of theory will increase sample coverage and thereby reduce content, so that, in effect, Popperian methodology condemns any complication of theory out of hand, no matter how much it improves accuracy! The more liberal Bayesian approach, on the other hand, qualifies a complication as "progressive" just in case the accuracy gained is enough to offset the loss of content, as determined by the precise yardstick of average likelihood. Bayesians may speak, accordingly, of *support-increasing* or *support-reducing* complications of theory. Persistent failure to find a support-increasing complication to account for discrepant data certainly looms as a difficulty for any theory (and its proponents), but no automatic rejection is implied.

To illustrate these rather abstract remarks, consider again the problem of the planets. The heliocentric scheme has all the planets orbiting the sun in simple closed curves. To capture the salient features of planetary motion, we might begin with an oversimplified heliocentric model based on uniform coplanar sun-centered circles. This model is not very accurate, but its major simplifications already stand out quite clearly. For the relative sizes of the orbits can be determined from observation of just a few positions per planet, and all the major irregularities of planetary motion—the number and frequency of retrogressions, variations in apparent brightness and diameter, and so forth—are accounted for at one stroke as effects of the earth's motion (as Copernicus emphasized in Book I of *De Revolutionibus*). Moreover, the theory fits *only* the behaviors actually observed. The contention that the complexity of the full system Copernicus proposed obscured these simplifications strikes me as highly questionable. Astronomers like Brahe and Kepler distinguished quite clearly between the simplifications inherent in the heliocentric picture and the complexities of Copernicus's own filling out of the details. (Kepler even accuses Copernicus of not being Copernican enough in needlessly complicating his system by placing the center of planetary motion at a point near, but distinct from, the center of the sun.) And, in point of fact, a rather minor complication of the oversimplified model based on eccentric circles with motion uniform about an equant point and orbital planes slightly inclined to the ecliptic but all passing through the sun, would have produced unprecedented accuracy.

Kepler's refinement of the picture clearly embodies the methodological principles stated here. Thus, in complicating a model to improve fit, one should complicate it minimally. Kepler's ellipses are minimal complications of circles, and, in addition, his second law retains the feature of uniform circular motion that the radius vector sweeps out equal areas in equal intervals of time. Finally, his third law represents a quantitative sharpening (or simplification) of the empirical rule-of-thumb (discussed earlier) that a planet's period increases with its distance from the center of motion. Because this law relates the motions of different planets, Kepler's laws as a whole provide a model of the planets that is, I would surmise, comparable in simplicity to the model based on uniform circles. (Newton's gravitation law represents an additional simplification, imposing dynamical constraints that exclude various kinematically possible systems of Keplerian orbits as unstable.) In any case, the vastly improved accuracy of Kepler's model renders it support-increasing. And, in addition, Kepler's model lends itself to a natural causal or physical interpretation in a way that Ptolemaic and Tychonic models did not. Planets speed up as they approach the sun and planets closer to the sun go round faster, pointing clearly to the sun as a causal agent.

Let us look now at Glymour's position on these matters, for again, we find much substantive agreement in the face of proclaimed disagreement. First, Glymour is one of the very few codefenders of the view, espoused in chapter 7 of my 1977 book, that the Copernican theory really is simpler than the Ptolemaic and that its greater simplicity has evidential value. "On several grounds," he writes, (1980, p. 198), "Copernican theory is superior to Ptolemaic astronomy: there are properties of the bodies of the solar

system that are presupposed by both theories, but that are indeterminable in Ptolemaic theory whereas they can be determined within Copernican theory." (1980, p. 198) And he goes on to urge that this virtue rendered the Copernican alternative the better confirmed of the two.

Similar agreement with my view that simplicity has evidential force is found in his discussion of the classical tests of general relativity, a discussion ostensibly designed to show that Bayesians cannot account for the judged relative importance of the different tests. (pp. 277 ff.) After pointing out that the anomalous advance of the perihelion of Mercury could be accommodated by a number of theories, he writes:

> Perhaps the most common and influential objection to these contenders against general relativity was that, unlike Einstein's theory, they saved the phenomena only by employing an array of arbitrary parameters that had to be fitted to the observations. Eddington, barely concealing his contempt, objected against ether theories of the perihelion advance and gravitational deflection that they were not "on the same footing" with a theory that generated these phenomena without any arbitrary parameters. It was pointed out that Poincaré's extension of Lorentz's theory could, by proper adjustments of a parameter, be made consistent with an infinity of perihelion advances other than the actual one. Conversely, Einstein's derivations of the phenomena were praised exactly because they involved no arbitrary parameters—and, the exception that proves the rule, also criticized because they did. (p. 284)

No one who has digested even the very sketchy discussion of average likelihood in this paper will have the slightest difficulty accounting for such judgments in Bayesian terms. Glymour's own essential agreement with the deliverances of Bayesian analysis comes out most clearly in his chapter VIII on curve-fitting. The application of the average likelihood index of support to polynomial regression is taken up in chapter 11 of Rosenkrantz (1977), and its performance is compared with that of various non-Bayesian (or "orthodox") tests. Glymour does not discuss either Bayesian or orthodox approaches to curve-fitting, but he does offer a way of assessing the severity of a test that uses a notion quite reminiscent of sample coverage:

> How is severity to be compared? Suppose that we have the hypotheses H and G and that, given prior data, the range of outcomes on a new experiment that would be in contradiction with H is properly contained in the range of possible outcomes that would be in contradiction with G. . . In this sort of case, I think it is natural and proper to regard G as more severely tested than H. (pp. 333-334)

And later he adds, "one can try to develop, not just the crude comparison of severity of tests I have used, but a *measure* of severity of tests. . . ." (p. 339) If "consistency with the data" is understood in a probabilistic sense admitting of degrees (really the common usage in the sciences), sample coverage provides just such a measure. Then Glymour's suggestion that polynomials of lower degree "are preferred to more complex families that also fit the data because . . . the data provide more and severer tests of the simpler hypothesis than of the more complex one" (p. 335) will follow readily from a Bayesian analysis in terms of average likelihoods, if by "fit the data" we understand "fit the data equally well."

It is unfortunate that a misreading of my (possibly obscure) 1976 paper on simplicity prevented Glymour from appreciating that Bayesian analysis delivers precisely what his own intuitions demand. He says that I fail to show "that in curve-fitting the average likelihood of a linear hypothesis is greater than the average likelihood of a quadratic or higher degree hypothesis." But of course I don't want to show *that*, for it isn't true! What can be shown is that the average likelihood of the quadratic family will be higher than that of the linear family when the data fit the quadratic hypothesis *sufficiently better* than the linear one, whereas the latter will enjoy higher average likelihood when the two families fit equally well.

Obviously it has not been my intention to attack Glymour's intuitions about simplicity. By and large, I see in him a kindred spirit, one who recognizes both the central role simplicity plays in the deliberations of theoreticians of all stripes and its objective evidential force. His tendency to think that Bayesians cannot account for its role and force is perhaps understandable in light of the extent to which the subjectivist form of the Bayesian approach has dominated the scene, until quite recently. (Indeed, many writers still use "Bayesian" and "subjectivist" interchangeably.) Unlike objectivists, such as Sir Harold Jeffreys, subjectivists have laid very little stress on average likelihood or on models with adjustable parameters, quite possibly because the need to adopt a parameter distribution or weighting function when employing the average likelihood index somewhat vitiates the subjectivists' claim to be able to separate cleanly the "subjective element" (the prior) from the "public element" (the import of the data). At any rate, no theory of evidence that fails to handle models with free parameters or account for the felt diminution of support that results from adding parameters or accommodating more outcomes that might have been but were not observed, can be taken very seriously. Simplicity looms as the central problem in the whole field of theory and evidence (and a glance at Glymour's index would tend to vindicate this judgment). To Popper must go much of the credit for keeping the issue of simplicity alive at a time when methodologists of a positivist persuasion were more inclined to dismiss it as a will-of-the-wisp or consign it to the limbo of the "purely pragmatic" or the "merely aesthetic."

(b) Novel predictions and old evidence

Another old chestnut, closely related to the Popperian demand for placing our conjectures in jeopardy, is the maxim that hypotheses are more strongly confirmed by their ability to predict facts not already known. Some would go even further and say that theories are not confirmed at all by already known facts or previously available data. Yet, to all appearances, Bayesian methodology is at odds with this principle. For if we think of support or likelihood as a timeless relation between propositions (akin to logical implication in this respect), then P(E/H) does not depend on whether or not E was known prior to proposing H.

Scientists have, though, a curious ambivalence about this time-honored precept. Specifically, they never fail to pay it lip-service and never fail to disregard it in practice whenever it tends to weaken the evidence for their own theories. Almost all the empirical support for Dalton's atomic theory, including the laws of constant and multiple proportions, was already known, yet it was cited as evidence for the theory. And in his popular account of relativity (Einstein 1916), Einstein quite expressly states that all the facts of experience that support the Maxwell-Lorentz theory of electromagnetic phenomena also support the special theory of relativity, since the latter "has crystallized out of" the former. Einstein cites in particular the experiment of Fizeau as having "most elegantly confirmed by experiment" the relativistic version of the law of addition for velocities, even though that experiment had been performed more than fifty years earlier (before relativity was even a twinkle in Einstein's eye) and had moreover been explained by the Maxwell-Lorentz theory. In point of fact, there was no evidence of a "novel" sort to which Einstein could point, since it was not then technically feasible to accelerate small particles to speeds approaching that of light or to perform mass-energy transformations in the laboratory. What Einstein did point out instead was the ability of relativity theory to account for "two classes of experimental facts hitherto obtained which can be represented in the Maxwell-Lorentz theory only by the introduction of an auxiliary hypothesis"—in other words, he pointed to the greater simplicity of relativity. About the negative result of the Michaelson-Morely experiment he writes:

> Lorentz and Fitzgerald rescued the theory from this difficulty by assuming that the motion of the body relative to the aether produces a contraction of the body in the direction of motion, the amount of contraction being just sufficient to compensate for the difference in time mentioned above. Comparison with the discussion of Section XII shows that also from the standpoint of relativity this solution of the difficulty was the right one. But on the basis of the theory of relativity the method of interpretation is incomparably more satisfactory. According to this theory there is no such thing as a "specially favored" (unique) coordinate system to occasion the introduction of the aether-idea, and hence there can be no aether-drift, nor any experiment with which to demonstrate it. Here the contraction of moving bodies follows from the two fundamental principles of the theory, without the introduction of particular hypotheses; and as the prime factor involved in this contraction we find, not the motion in itself, to which we cannot attach any meaning, but the motion with respect to the body of reference chosen in the particular case in point. (1916, p. 53)

The situation was not so very different in the case of general relativity. Einstein laid great stress on the equality of inertial and gravitational mass, a brute fact in the old physics, but a necessary consequence of the general principle of relativity in the new physics. Here too the difference is one of overdetermination and has nothing to do with novelty per se. And of course the advance of the perihelion of Mercury, predicted by general relativity, had long been established and measured with precision (the predicted advances in the perihelia of other planets were too small to be detectable).

Faced with these and other obvious exceptions to the precept, Lakatos and Zahar (1975) fall back on a modified form of it that accounts a prediction "novel" when it was not consciously used to arrive at the theory (so that the theory explains it in passing). If taken quite literally, this proposal would require us to read a theorist's mind before being able to assess the evidence for his theory (see Michael Gardner's contribution to this volume). In any event, they are able to argue on this basis that the stations and retrogressions of the planets, the brightness of a planet at perigee, and the bounded elongation of an inner planet from the sun, etc., all count as "novel" predictions of the Copernican theory, though not of the Ptolemaic. They observe that "although these facts were previously known, they lend much more support to Copernicus than to Ptolemy, within whose system they

were dealt with only in an ad hoc manner, by parameter adjustment" (Lakatos and Zahar, 1975, p. 376). That is, the Ptolemaic theory could account for these effects of the earth's motion only by fitting additional parameters or making additional assumptions. Could it be more clear that the real appeal here is, not to "novelty," but to overdetermination?

There is, to be sure, another sense of "novelty" that plays a more important role: namely, a prediction is novel when it is unexpected on rival theories (or on rival theories of comparable simplicity). And, of course, Bayesians have no difficulty accounting for the force of predictions that are "novel" in this sense.

The solution of the problem Glymour poses about old evidence (1980, pp. 86-92) should also be clear. The puzzle is this: if an item e of evidence is already known, then it must have probability one, and consequently, even if a hypothesis h entails it, $P(h/e) = P(e/h)P(h)/P(e) = P(h)$, using Bayes's formula, and no confirmation is registered. (This is a sort of obverse of the charge that Bayesians are unable to account for the peculiar force of novel predictions.) On *objectivist* Bayesian grounds, however, the likelihoods $P(e/h_i)$ of the alternative hypotheses are timeless relations, and of course $P(e)$ must be computed relative to a considered partition of hypotheses, $h_1, \ldots, h_n$ by the partitioning formula, $P(e) = P(e/h_1)P(h_1) + \ldots + P(e/h_n)P(h_n)$. And this quantity will be less than one, unless e is a necessary truth. For purposes of comparing hypotheses, then, the probability of old evidence is *not* one, and may even be quite small. This only shows, of course, that old evidence poses no difficulty for an objectivist Bayesian position—a point that Glymour readily conceded at the conference. (For a subjectivist's way of handling the problem, see Daniel Garber's contribution to this volume.)

What does cry out for explanation is our conviction that the ability of general relativity to fit the already measured advance of the perihelion of Mercury can afford just as striking a confirmation (and seem quite as "miraculous") as the ability of that theory to predict the precise magnitude of the deflection of starlight passing close to the sun. While I was listening to Glymour describe Einstein's vicissitudes in finding a covariant theory that would account for the advance of Mercury's perihelion, the solution of this puzzle suddenly became quite clear. The point is that Einstein's success was not assured. What is generally overlooked is that one is not interested in finding any old theory to explain an anomaly; one seeks, in practice, a (reasonably simple) theory *of specified form*. Thus Einstein sought a theory that satisfies the general principle of relativity. And we can think of such a quest in the following way. One is interested, at bottom, in the hypothesis that there *exists* a (not unduly complicated) theory of such-and-such form capable of accommodating the data from a certain class of experiments, only some of which have already been performed. That there does exist a theory of the required form that fits the output of an already performed experiment of the class in question affords, on straightforward Bayesian grounds, a more or less striking confirmation of the existential hypothesis in question. And the longer or more tortuous the derivation, and the more different (and tenuous) the theoretical assumptions involved, the more striking the confirmation (of all the implicated principles) will be (as in Bohr's derivation of the Balmer series for hydrogen).

(c) Projectibility

Hypothetico-deductive accounts face the difficulty that an observation may be a consequence of more than one hypothesis, and, in particular, of a "counterinductive" or "unprojectible" hypothesis. Examining an emerald before time t and finding it green is a consequence of "All emeralds are grue," as well as of "All emeralds are green." And this seems disturbing, inasmuch as the grue hypothesis licenses the prediction that emeralds not examined before time t are blue, hence emeralds of a different color. Since consequences of a hypothesis are confirmatory on a Bayesian account, some restriction of the Bayesian confirmation relation seems called for. And, quite apart from this concern, we have been witnessing, since the early 1950s, a search for a basis for excluding such "counterinductive inferences."

To be sure, Bayesian inference blocks this alleged paradox at many points. For one thing, there is no Bayesian consequence condition that would allow one to confirm the prediction of blue emeralds after time t. And, more generally, there are ways of handling irrelevant conjunction. Yet these considerations do not seem to go to the heart of the matter. For the more serious issue here, in my view, is *whether* (not how) we should drive a wedge between "projectible" (or "lawlike") and "unprojectible" hypotheses.

The grue hypothesis belongs to a class we might label *bent* or *crooked*. Such hypotheses posit a breakdown of a *straight* counterpart in some nonlocal region of space or time. The grue hypothesis is, admittedly, a rather extreme case in that it posits a sharp discontinuity, but presumably those who view such hypotheses as absolutely unconfirmable would regard their continuous or gradualistic modifications as equally unsavory.

And yet science is riddled with bent or crooked hypotheses, and this should certainly make us wary of any proposal to banish them wholesale. Nelson Goodman's own attempt to do so, the entrenchment theory, would list, among others, the hypotheses of relativity theory among the unprojectible! For example, the Einsteinian hypothesis "All particles subject to constant force have linearly increasing relativistic momentum" is "overridden," in Goodman's sense, by its Newtonian counterpart, "All particles . . . have linearly increasing momentum." For the latter had, circa 1905, much the better entrenched consequent predicate and was, up to that time, unviolated, supported, and unexhausted. In effect, the hypotheses of special relativity posit departures from their Newtonian counterparts that become experimentally detectable only at speeds close to the speed of light. They are, in this respect, perfectly representative bent hypotheses. It is no defect of Bayesian methodology that it gives such hypotheses a hearing.

From a Bayesian point of view, lawlikeness admits of degrees and is chiefly a function of simplicity and theoretical assimilability (as reflected in a prior distribution). I am quite content to let it go at that, for I am convinced that there is no fundamental distinction to be drawn between hypotheses that are projectible or confirmable and those that are *absolutely* unconfirmable. (I argue this point at greater length in Rosenkrantz 1982, pp. 86-91)

3. Informal Assessments and Epistemic Utilities

A good theory can explain the salient facts without recourse to special assumptions of an arbitrary kind. I have been urging that the Bayesian theory of evidence is a theory of precisely this sort. It dispenses with ad hoc prescriptions and so-called epistemic utilities. *Genuine* epistemic utilities, like content, are automatically reflected in support (and, in effect, this provides a criterion for distinguishing the genuine from the spurious among them). From a strict Bayesian point of view, support is all in all. It is not surprising to find, therefore, at least one sympathetic reviewer of my 1977 book (Jaynes 1979) wondering why I even bother with the adhockeries that disfigure so much of the literature of scientific method. Why, indeed, do I attempt a precise explication of simplicity when, however defined, simplicity matters only insofar as it is reflected in support? At the other extreme, some critics of the Bayesian approach question its applicability to actual scientific evidence. Glymour raises such doubts in his recent paper when he writes:

> I am inclined to doubt that, in many situations, we have either objective probabilities or subjective degrees of belief of a sufficiently global kind upon which we can rely to relate evidence to theory. When theories are proposed for novel subject matters (as in some contemporary social science) or when new theories are seriously considered which deny previously accepted fundamental relationships. . . , we may be at a loss for probabilities connecting evidence to theory. (1981, p. 696)

These two questions may seem unrelated, not to say oppositely directed, but, in essence, they elicit the same reply.

Although most criticism of the second kind focuses on the alleged arbitrariness of prior probabilities of theoretical hypotheses (the passage from Glymour tends that way), the real difficulty, more frequently, is to compute the relevant likelihoods—a point that Patrick Suppes has emphasized on numerous occasions. It often happens that we can calculate conditional outcome probabilities for a "null hypothesis" of chance or randomness, but we cannot calculate them for the hypotheses (of association, or causal connection) of real interest to us. For a very simple example, consider R.A. Fisher's celebrated case of the tea-tasting lady, who claims an ability to discriminate whether the tea or milk was infused first in a mixture of milk and tea. Fisher's design calls for the lady to classify eight cups, of which four are milk-first and four are tea-first (and the lady knows this). It is then easy to find the probability that she classified r of the eight cups correctly, given that she is merely guessing; but there is no way to calculate these probabilities on the supposition that she has some skill. The prevalence of cases like this one explains the widespread use of tests of statistical significance. Such tests are used to make rather informal assessments of evidence even in cases in which no well-defined alternative hypotheses are in view.

Now my answer to both points can be given at once. First, epistemic utilities are important in precisely those contexts in which Bayesian methods cannot be applied for inability to compute the relevant likelihoods. (My earlier, often outspoken, criticism of epistemic utilities is here softened to this extent.) At the same time, however, our informal assessments in these cases are (and ought to be) guided by the methodological insights that formal Bayesian analysis affords in the contexts in which it does apply.

To begin with, we might seek a qualitative analogue of the average likelihood. The latter, you recall, is a determinate blend of accuracy and content; it measures, roughly speaking, the improbability of a theory's accuracy. The ideal case is that in which the theory fits all and *only* those possible outcomes that actually occur. Demands of accuracy and simplicity alike narrow the range of outcomes that a theory can accommodate. Now in cases in which likelihoods cannot be computed, we may still have an intuitive rank ordering of experimental outcomes as agreeing more or less well with the theoretical conjecture of interest. Then we can mimic average likelihood in a qualitative way by the proportion of possible outcomes that (by the intuitive yardstick) fit the hypothesis *at least as well* as the outcome observed. The size of this proportion will again reflect accuracy and simplicity in a determinate way, and, moreover, in a way that tends to yield assessments qualitatively similar to those yielded by average likelihood where both methods apply (see the last section of Rosenkrantz 1976 on this). I call this proportion the *observed sample coverage*. In principle, any two hypotheses, whether mutually exclusive or not, can be compared by this informal measure. More generally, using a suitable null hypothesis, we can compute the *chance probability* of agreement with the hypothesis of real interest as good as (or better than) that observed.

To illustrate, if someone claims an ability to detect water with a hazel prong and boasts of a ninety percent rate of success, we should not be impressed unless his success rate is materially higher than that achieved by digging at random in the same area (i.e., the chance rate). If that cannot be shown, his accuracy is not improbable and his claim is unsubstantiated.

Informal assessments of evidence are often aimed at establishing improbable accuracy. I recently came across a beautiful example in Thor Heyerdahl's interesting book, *Early Man and the Ocean* (1979, chapter 3). The hypothesis of interest is that the cultural flowering that occurred in ancient Meso-America had sources (Sumerian, Egyptian, Hittite, or Phoenician) in the Near East. Heyerdahl protests the tendency of "isolationists" to dismiss the parallels between these cultures singly, rather than confronting them collectively, for there is a compounding of improbabilities. That one or two such parallels should arise by mere coincidence does not strain credulity, but the probability of finding well over a hundred by chance seems infinitesimal.

Heyerdahl's point is well taken, and even the partial list of over fifty parallels he compiles is nothing if not impressive (1979, pp. 84-92). Yet, the evidence from cultural parallels could be marshalled more convincingly by introducing missing ingredients of the informal Bayesian paradigm I have sketched. What we lack is a sense of how much similarity typifies cultures between which there has been no contact. We also need some assurance that dissimilarities are being systematically taken into account.

To this end, we need a well-defined sample space, in effect, an ethnographic survey of many cultures based on a single workable typology, and then we need a measure of similarity between cultures based on this typology. A computer could then be programmed to calculate the proportion of pairs of surveyed cultures manifesting a degree of similarity at least as great as that of the pair for which contact is hypothesized. That proportion (the observed sample coverage) estimates the probability that two cultures *chosen at random* would manifest at least as much similarity (i.e., the chance probability).

Such comparisons are, of course, no better than the typology and similarity measure on which they are based. Imagine that given items of the typology are treated as branching classification trees. As a first step toward measuring similarity with respect to that item, proceed down the tree to the last branch point at which the two cultures A and B of a comparison agree, then compute the proportion of surveyed cultures (*including* the pair A,B) which proceed at least that far down the same branch of the tree. Then the square of this proportion (necessarily positive) estimates the probability that two cultures chosen at random would agree on the given item to at least that level of specificity. In this way, our measure of similarity reflects both the specificity and statistical rarity of a shared custom or artifact, and dissimilarities are systematically taken into account. This desideratum stands out very clearly in Heyerdahl's discussion, which will suggest other desiderata and ways of refining our measure. My purpose here is no more than to indicate the general lines along which one might proceed.

As for prior probabilities, admittedly they are of little importance in preliminary investigations where we lack a sharply delimited set of theoretical alternatives. Observed sample coverage can still be applied to assess support in such contexts, without regard to alternative hypotheses. But where the theoretical possibilities have been effectively narrowed, we can expect the informal, qualitative counterparts of prior probabilities, which I will call "initial plausibilities," to play a major role. Indeed, Heyerdahl's famous voyages were mounted to explode the supposed implausibility of certain migration routes. His point is that routes that seem implausibly long in miles may actually be short when powerful ocean currents and trade winds are taken into account. His voyages demonstrated the feasibility of a journey across the Atlantic or across the Pacific from Peru to Polynesia in the highly seaworthy wash-through reed vessels or balsa rafts of the Egyptians and Incas (highly specialized constructions whose occurrence in all three places is itself one of the important bits of evidence pointing to contact between these cultures). Finally, by using the procedure of the last paragraph, one could hope to rule out alternative migration routes.

My suspicion is that informal counterparts of the three main elements of a formal Bayesian analysis—prior probabilities, likelihoods, and alternative hypotheses—figure importantly in nearly all informal assessments of evidence, and that more explicit use of the informal Bayesian index of support (the observed sample coverage) would often render assessments of this sort more systematic and more objective.

4. Glymour's Misgivings

With this much background, we can turn at last to Clark Glymour's reservations about Bayesian methods (some of which have already been touched on in passing) and the additional constraints he wishes to impose.

I think of Bayes' theorem as a refinement of the hypothetico-deductive approach. We seek hypotheses conditional on which actually occurring outcomes have high probability while nonoccurring outcomes have low probability. More precisely, Bayes's formula implies that a hypothesis h_i of a partition $h_1, \ldots, h_n$ is confirmed by an outcome e just in case e has a higher probability on h_i than it has on the average, relative to the members of the partition (i.e., iff $P(e/h_i) > P(e/h_1)P(h_1) + \ldots + P(e/h_n)P(h_n)$). And, by the same token, a member h_i of a partition of hypotheses is not disconfirmed by outcomes that are highly improbable on h_i unless those outcomes are substantially more probable on alternative hypotheses of the partition. It is widely conceded that this scheme characterizes, in a general way, both the precepts and practice of working scientists and model-builders. Glymour too concedes it, yet he denies that hypothetico-deduction is a sound scheme in all respects (1980, pp. 29 ff.).

His main fear is that it cannot handle irrelevant conjunction. If *e* is held to confirm *h* just by virtue of *h*'s entailing it, then, equally, *e* must confirm h&k as well, where *k* is any hypothesis you like. Again, if degree of confirmation is measured by the ratio P(h/e):P(h) = P(e/h):P(e) of posterior to prior probability, then if *e* is a consequence of *h*, P(e/h) = P(e/h&k) = 1, and *e* will accord h&k precisely the same degree of confirmation it accords *h* alone. That seems objectionable when *k* is extraneous, and even more objectionable when *k* is probabilistically incompatible with *h* in the sense that P(k/h) is low. Personally, I have always considered this reason enough to reject the ratio measure in favor of the difference measure:

$$(4.1)\quad dc(e,\ h) = P(h/e) - P(h)$$

writing dc(e, h) for the *degree of confirmation e* accords *h*. This measure is easily seen to satisfy the following condition:

$$(4.2)\quad dc(e,\ h\&k) = P(k/h)dc(e,\ h) \text{ when } e \text{ is a consequence of } h.$$

And this little theorem seems to deliver precisely what intuition demands. For, on the one hand, we certainly don't want to say that a consequence of *h* should *dis*confirm h&k. But neither should it confirm h&k as strongly as *h*. Indeed, the degree of compatibility of *k* with *h* should control the rate of depreciation, and this is what (4.2) says.

The difference measure can be applied to conclude, for example, that examining a sample of emeralds for color before time t and finding them green (*e*) accords "All emeralds are green" a higher degree of confirmation than "All emeralds are grue." For the former hypothesis is the conjunction of *h*: "All emeralds examined before time t are green," with *k*: "All emeralds not examined before time t are green," whereas the latter is the conjunction of *h* with *k*′: "All emeralds not examined before time t are blue." Given our background knowledge that emeralds do not change color all at once, either individually or as a class, P(k/h) >> P(k′/h). And the asymmetry in question is language independent. By contrast, the Hempelian account of confirmation registers confirmation for both h&k and h&k′, and leaves us unable to discriminate between them. Worse still, because that account satisfies the consequence condition, *e* will also confirm *k*′—the dreaded counterinductive inference. Irrelevant conjunction is, therefore, very much a two-edged sword.

Notice too how our explication of content handles irrelevant conjunction. On Popper's account, conjoining an extraneous hypothesis represents a simplification, since more states of the world are then logically excluded. But, in our probabilistic version, this is not so, for content is relativized to a contemplated experiment. Thus conjoining, say, a hypothesis about the velocity of neon light in beer to a Mendelian model of a given mating experiment will have no effect on the latter's sample coverage for that experiment. No simplification results, but prior probability is necessarily reduced.

I come next to "deoccamization" (see Glymour 1980, pp. 30-31). At first blush, one is tempted to say that a deoccamized theory (one in which a parameter is replaced throughout by a function of several other parameters) differs only notationally from the theory it deoccamizes. To the extent that two theories fit the same outcomes of the same experiments to the same degree, I regard them as equivalent. And so it troubles me not at all that a theory and a deoccamization of it may have the same sample coverage or the same support. The only considerations that would lead anyone to prefer one such notational variant to another one are, I should think, considerations of elegance or of a heuristic nature. And I see no reason to issue prescriptions on this matter.

There is nevertheless something about Glymour's position that troubles me. He leaves it as an exercise for the reader to show that deoccamization will reduce a theory's testability. (pp. 143-144) But let the theoretical term t of theory T be replaced throughout by the sum t′+t″ of two new parameters, t′ and t″, yielding the deoccamization T′ of T. (To borrow one of his examples, "force" in classical mechanics might be uniformly replaced by the sum of "gorce" and "morce.") Now it seems to me that any instance of a hypothesis *h* of T deducible from observations and T is ipso facto an instance of the corresponding hypothesis *h*′ of T′. For any determination of t is likewise a determination of t′ + t″. So T and T′ have, on Glymour's own showing, the very same evidence. I think he escapes this conclusion only by imposing a further requirement, namely, that for a hypothesis to be tested by given data, *every* theoretical parameter of that hypothesis must be determined. It is not enough that t′+t″ be determined from the observations; each of t′ and t″ must be determined (a reading suggested by Glymour 1980, p. 357).

It will come as no surprise that I consider this requirement overly stringent. In fact, I think it goes against the grain of Glymour's own approach. For if observations determine a sum of two theoretical quantities, why shouldn't we be willing to count that as a test of any hypothesis in which they occur—albeit a weaker test? After all, two different determinations of such a sum must yield the same value, and this constrains the data. That all quantities be individually determinable is the ideal case, but our admiration for the ideal should not lead us to disparage the good.

Glymour himself concedes that "pure deoccamization perhaps never occurs in science, but what does sometimes occur is deoccamization together with additional, untested claims about the new quantities." (p. 364) The clear implication is that such quasi-deoccamization is as much to be shunned as the real thing. I wonder about that too. Where there are additional claims, there is additional content, even if it lies beyond the reach of present experimental techniques. A theory like Einstein's, which says that mass (or energy) is really a sum of two terms, rest energy and energy of motion, one of which becomes appreciable only at speeds close to that of light, seems to be a theory of exactly this sort. When it was proposed, there was no way to test it. And similarly, particle physics is riddled with hypotheses stating that an elementary particle is really made up of a pair of such particles, but where the new predictions that follow are presently inaccessible to experimentation. Consider the following illustrative passage about "charm" from Nigel Calder's popular book *The Key to the Universe:* "Nor could the gipsy itself help in settling the issue in favor of charm. Supposing that the new particle did indeed consist of the charm/anti-charm combination, the charm was thoroughly hidden because it was self-cancelling. With zero net charm the gipsy could not be expected to show direct signs of charmed behavior." (p. 111) This looks very much like another case of quasi deoccamization, one that should give us pause.

To be sure, Glymour's comments on my presentation (repeated in Glymour, 1981) make it plain that he does not object to deoccamization when there are positive reasons for thinking that the new quantities have distinguishable denotata. He avers that "the demand for bootstrap confirmation [wherein every quantity is individually determined] is, I am sure, at best *prima facie* and indefeasible." But then it is left for the rest of us to wonder what all the hoopla is about if, as he admits, *pure* deoccamization never occurs. The only substantive issue that divides us is whether or not to insist that every theoretical quantity be individually determined in any test of a hypothesis. And this requirement strikes me as highly representative of those that cry out for justification, either in terms of a more comprehensive methodology or theory of rationality or as facilitating the achievement of cognitive objectives.

Many actual cases of bootstrapping seem to violate this additional stricture. Newton was able to test his gravitation law by comparing "the force requisite to keep the moon in her orb with the force of gravity at the surface of the earth" and finding them "to answer pretty nearly." By equating the centripetal force acting on the moon with gravitational force (and neglecting the sun), one obtains:

$$m_M v^2/R = Gm_E m_M/R^2;$$

and equating the moon's velocity v with the circumference of its orbit, $2\pi R$, divided by its period T, one has the following expression for T:

$$T^2 = 4\pi^2 R^3/Gm_E$$

where m_E is the mass of the earth and G is the gravitational constant. In this test of the law, Newton was not able to determine G and m_E separately, but he could determine their product as $Gm_E = gr^2$, where r is the earth's radius and g the acceleration of free fall, using the obvious relation $mg = Gm_e m/r^2$. This gives a theoretical determination of the moon's period which could be checked against observation. Would Glymour deny the force of the very test that apparently clinched the matter for Newton?

Here is another example. When Venus is at maximal elongation from the sun, earth, Venus, and sun lie on a right triangle and measurement of the angle SEV at E yields the ratio VS:ES of the orbital radii. On the other hand, at inferior conjunction, when E, V, S lie on a line in that order, we have ES = EV + VS. Assuming that ES is known, we have a determination of EV, the distance from the earth to Venus at inferior conjunction. Now Venus is fairly close at inferior conjunction, and we might hope to determine this distance directly by triangulation. One slight hitch is that we can't really observe Venus at inferior conjunction, since its orbit is nearly coplanar with the earth's orbit. But we can measure its apparent diameter at points very close to inferior conjunction, so let us ignore this difficulty for the sake of argument. The more serious problem is that we lack an independent determination of the actual diameter of Venus. Still undaunted, we make the natural but wholly untested assumption that the diameter of Venus does not differ appreciably from that of the earth. Now, for the punch line, imagine that our two independent determinations of EV agree within experimental error. Would this confirm the (heliocentric) hypothesis that the center of Venus's epicycle is at S? Here the apparent diameter determines only the product of EV by the actual diameter. Still, I am inclined to think that some confirmation is registered, if only because the apparent diameter is *determined* by an assumption about the actual diameter within the heliocentric theory but not within the geocentric theory. In fact, any other epicycle of Venus (compatible with its observed angular variations) containing the sun must intersect the sun-centered epicycle, and at the points of intersection we would have conflicting predictions of apparent diameter. Still, all I want to claim is that some slight confirmation of all the implicated hypotheses would be registered by agreement of our two determinations of EV. Does Glymour disagree? In defense of bootstrapping he writes:

> One claims that if certain principles of the theory are true, then certain empirical data in fact determine an instance of some theoretical relation, . . . This is some reason to believe the hypothesis, but a reason with assumptions. Of course it is possible that the assumptions—the hypotheses used to determine values of theoretical quantities—are false and a positive instance of the hypothesis tested is therefore spurious, or a negative instance equally spurious. But this does not mean that the test is circular or of no account. (1980, p. 352)

And that is why I said earlier that the requirement that all quantities be separately or independently determinable goes against the grain of Glymour's own conception of bootstrapping. In the case before us, we achieve this only by making a wholly untested assumption. But that does not make our test "of no account."

Glymour's remaining objection to the Bayesian account of confirmation is that it does not satisfy the *consequence condition:* that whatever confirms a hypothesis H confirms any consequence K of H. His intuitions tell him that this holds in at least some cases. But presumably his intuitions also allow that hypotheses are confirmed by their consequences or verified predictions in at least some cases. And he knows that this principle cannot be combined with the consequence condition to yield a non-trivial confirmation theory, unless one of the conditions is suitably restricted. The Bayesian theory restricts the consequence condition, satisfying it only for those consequences K of H such that P(K/H) >> P(K/notH). (Such K might be called "explained consequences," inasmuch as alternative explanatory hypotheses are effectively excluded.) True, this opens the door to irrelevant conjunction, but the alternative to admitting that a consequence E of H confirms the conjunction of H with any H′ is, we saw, far less palatable.

And failure of the consequence condition removes much of the sting anyway, for even though E confirms H&H′, it may disconfirm H′. Moreover, on the difference measure, dc(E, H) = P(H/E) – P(H), the degree to which E confirms H&H′ drops to zero when H is inconsistent with H′. No other confirmation theory, I submit, can steer a safer passage between the implausibilities of the various corner positions.

Wherever one looks for substantive disagreement between the deliverances of the bootstrapping and the Bayesian accounts of confirmation, one fails to turn them up, *unless* additional strictures that fly in the face of much scientific practice and cry out for justification are introduced.

References

Bartlett, M.S. 1975. Epidemics, In *Probability, Statistics and Time,* New York: Wiley, pp. 111-123.

Calder, N. 1979. *The Key to the Universe*. New York: Viking.

Calder, N. 1980. *Einstein's Universe*. Harmondsworth: Penguin.

Einstein, A. 1916. *Relativity*. New York: Crown. (Reprinted 1961)

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

Glymour, C. 1981. Bootstraps and Probabilities. *Journal of Philosophy* LXXVII: 691-699.

Heyerdahl, T. 1979. *Early Man and the Ocean*. New York: Doubleday.

Jaynes, E.T. 1979. Review of Rosenkrantz (1977). *Journal of the American Statistical Association 74:* 740-741.

Lanczos, C. 1967. Rationalism and the Physical World. *Boston Studies in the Philosophy of Science,* V. III, Dordrecht: Reidel.

Lakatos, I., and Zahar, E. 1975. Why Did Copernicus's Research Programme Supercede Ptolemy's? In *The Copernican Achievement,* Berkeley and Los Angeles: University of California Press, pp. 354-383.

Morrison, D.F., and Henkel, R.E., eds. 1970. *The Significance Test Controversy,* Chicago: Aldine.

Popper, K. 1959. *Logic of Scientific Discovery*. London: Hutchinson.

Rosenkrantz, R.D. 1976. Simplicity as Strength. In *Foundations of Probability and Statistics and Statistical Theories of Science,* v. 1, ed. W.L. Harper and C.A. Hooker, Dordrecht: Reidel, pp. 167-196.

Rosenkrantz, R.D. 1977. *Inference, Method and Decision,* Dordrecht: Reidel.

Rosenkrantz, R.D. 1980. Measuring Truthlikeness. *Synthèse* 45: 463-487.

Rosenkrantz, R.D. 1982. Does the Philosophy of Induction Rest on a Mistake? *Journal of Philosophy* LXXIX, 78-97.

Watson, J. 1968. *The Double Helix*. New York: Athenaeum.