

## *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 *confirming* observations with *conforming* 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:

$$(1.1) \quad P(e/h, T) = 1 \text{ and } P(e/T) < 1.$$

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:

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

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 realissimum* 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  $h_1, \dots, h_n$ . ( $H$  is the disjunction of the mutually exclusive  $h_i$ ’s.) Applying Bayes’s formula,

$$\begin{aligned} P(H/e) &= \sum_i P(h_i/e) \\ &= \sum_i P(e/h_i)P(h_i)/P(e) \end{aligned}$$

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

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

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

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  $C \neq S$ . 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,