Testing Hypotheses: Prediction and Prejudice

Peter Lipton

Observations that fit a hypothesis may be made before or after the hypothesis is formulated. Can that difference be relevant to the amount of support that the observations provide for the hypothesis? Philosophers of science and statisticians are both divided on this question, but there is an argument that predictions ought to count more than accommodations, because of the risk of "fudging" that accommodations run and predictions avoid.

In the case of "accommodation," a hypothesis is constructed to fit an observation that has already been made. In the case of "prediction," the hypothesis, though it may already be partially based on an existing data set, is formulated before the empirical claim in question is deduced and verified by observation. Well-supported hypotheses often have both accommodations and successful predictions to their credit. Most people, however, appear to be more impressed by predictions than by accommodations. Edmond Halley was able to account for the observed comets of 1531, 1607, and 1682 as a single object with a perturbed elliptical orbit. Natural philosophers took some notice when he published his views in the *Philosophical Transactions* in 1705, but it was only when his prediction of the return of the comet in 1758 was confirmed that the entire European intellectual world embraced Halley's Comet. The single prediction appears to have been far more impressive than the three accommodations (*1*).

Was this reaction rational? It is surprisingly difficult to establish an advantage thesis: to show that predictions tend to provide stronger support than accommodations. The content of the hypothesis, of the statements needed to link the hypothesis to the observation, of background beliefs, and of the observation itself may all be unaffected by the question of whether the observation was accommodated or predicted, and these seem to be the only factors that can affect the degree to which a hypothesis is supported by evidence. The difference between accommodation and prediction seems merely one of timing. Some observations are more reliable and more telling than others, but the date when they were made seems to be irrelevant (*2*–*6*).

To make the case against the advantage thesis more vivid, consider a fictitious case of twin scientists. These twins independently and coincidentally generate the same hypothesis. The only difference between them is that one twin accommodates an observation that the other predicts. If there really were an advantage to prediction, we ought to say that the predictor has more reason to believe the hypothesis than the accommodator, though they share hypothesis, data, and background beliefs. This is counterintuitive, but things get worse. Suppose that the twins meet and discover their situation. It seems clear that they should leave the meeting with a common degree of confidence in the hypothesis they share. If they came to the meeting with different degrees of rational confidence in their hypothesis, at least one of them ought to leave with a revised view. But what level should they settle on: the higher one of the predictor, the lower one of the accommodator, or somewhere in between?

There seems no way to answer the question. Moreover, if there is a relevant difference between prediction and accommodation, then the twin who should revise her view after she meets her sibling must not revise simply because she knew all along that someone like her twin might have existed. If revision were in order merely because of this possibility, then the difference between accommodation and prediction would vanish. Whenever data are accommodated, we know that there might have been someone who produced the hypothesis earlier and predicted the data instead. But how can the question of whether there actually is such a person make any difference to our justified confidence in the hypothesis? Any adequate defense of the putative difference between prediction and accommodation will have to explain how an actual meeting could be different from a hypothetical meeting. Those who reject the distinction seem to be on firm ground when they maintain that no such explanation is possible.

Ad Hoc Hypotheses, Real Tests, and Best Explanations

Here are three popular defenses of the advantage thesis. One is that accommodation allows a hypothesis to be built around the data, but such a hypothesis would be ad hoc and hence only poorly supported. As it stands, this is not a good argument. The expression "ad hoc" literally means "purpose-built." To say that an accommodation is ad hoc in this sense is just to repeat that it is accommodation; it is not to say or to show that the hypothesis is poorly supported or otherwise deficient. On the other hand, the expression "ad hoc hypothesis" is often used in a derogatory sense that implies that the hypothesis is poorly supported. But on that reading, the argument becomes circular: Accommodating hypotheses are poorly supported, therefore they are poorly supported. On either reading, the ad hoc argument fails.

A second argument for the advantage thesis is the argument from testing. According to this argument, predictions are worth more than accommodations because it is only through its predictions that a hypothesis gets properly tested, and it is only by passing a test that a hypothesis gains genuine support. The idea is that a test is something that could be failed, and it is only a prediction that a hypothesis can fail. Here the hypothesis is made to commit in advance and say how things will be, so that we may go on to discover that things actually are not that way. So if the hypothesis passes the test of prediction, the hypothesis has earned some credit. This thought is closely related to Karl Popper's emphasis on the importance of getting hypotheses to stick their necks out, though Popper himself took the radical view that there is no such thing as positive evidence (*7*). In accommodation, by contrast, the hypothesis does not stick its neck out: It cannot be shown to be wrong, because the hypothesis is constructed after the data are known and compatibility is thus guaranteed.

An analogy helps to bring out the intuitive strength of the argument from testing (*8*). Suppose that Jacob, my elder son, takes his trusty bow and arrow and shoots at a target on the side of a barn, hitting the bull's-eye. We are duly impressed. Now Jonah, my younger son, steps up to a different barn, pulls back his bow, and shoots his arrow at the barn. Then he walks up to the side of his barn and paints a bull's-eye around his arrow. We would give him rather less credit, for archery anyway. Accommodation is like drawing the bull's-eye afterwards, whereas in prediction the target is there in advance. The argument from testing seems clearly to show why successful prediction should count more than accommodation.

Nevertheless, as it stands, this too is a poor argument. It confuses the scientific hypothesis with the scientist who formulates it. What is true is that only in the case of pre-

Department of History and Philosophy of Science, Free School Lane, Cambridge CB2 3RH, UK. E-mail: Peter.Lipton@Kings.cam.ac.uk

diction does the scientist run the risk of having been mistaken, because it is only in the case of prediction that the scientist may have to admit to having made a false prediction. But we ought to concern ourselves here with the science, not with the scientist. When it comes to the hypothesis, there is no contrast between prediction and accommodation. If the data been different, the prediction would have been false and the hypothesis refuted or disconfirmed. But just the same goes for accommodated data: If they had been different, the hypothesis that was built around the actual data would also have been refuted. It is also true that, had the accommodated data been different, the scientist would have built a different hypothesis, but that is not to the point. As far as the specific hypothesis itself is concerned, the situation is symmetrical: If a prediction is a test, then so is an accommodation. So the argument from testing fails.

The third popular argument for the advantage thesis is the argument from the best explanation (*9*). In the case of accommodation, there are two explanations for the fit between hypothesis and data. One is that the hypothesis is true; another is that the hypothesis was designed to fit the data. We know that the second, the accommodation explanation, applies to the case; and this seems to preempt an inference to the truth explanation. In the case of prediction, by contrast, we know that the accommodation explanation does not apply, which leaves the truth explanation in the running. In one case, the truth explanation cannot be the best explanation, whereas in the other it might be. This is why prediction is better than accommodation.

This argument from the best explanation has considerable intuitive force but also a number of weaknesses. The most important of these is that it is unclear whether the accommodation explanation actually does preempt the truth explanation. They certainly could both be correct, and to assume that accepting the accommodation explanation makes it less likely that the hypothesis is true is once again to beg the question against accommodation. The issue is precisely whether the fact that a hypothesis was designed to fit the data in any way weakens the inference from the fit to the correctness of the hypothesis, and the argument from the best explanation does not help to settle it.

Choice and Fudging

Are there any better arguments for the advantage thesis? I will suggest two. To make sense of them, we need first of all to note that there are many relatively uncontroversial factors that affect the degree of support that an observation provides for a hypothesis (*10*, *11*). These can be roughly divided into features of the evidence and features of the hypothesis. On the list of evidential virtues, we may put down that more supporting evidence is better than less, but this is not the only entry. Variety in the data is also an evidential virtue. Someone who just repeats the same experiment over and over eventually reaches a point of diminishing returns, whereas a hypothesis supported by a variety of experiments rightly inspires greater confidence. Having accurate and precise data is another evidential virtue, as is having the results of controlled experiments, in which the scientist can be confident of the absence of disturbing influences. The same applies to so-called ''crucial'' experiments, in which the evidence simultaneously supports one hypothesis while undermining some of its rivals, and to evidence that would be very improbable unless the hypothesis were true.

One can also construct a list of theoretical virtues. One is the prior plausibility of the hypothesis: how natural it is and how well it fits with other claims that are already accepted. Simplicity is another theoretical virtue: With an appeal to Ockham's razor, the simpler hypothesis is often given a better chance of being correct. Other theoretical virtues include the plausibility of the auxiliary statements that have to be used to wring testable consequences out of the hypothesis, and the absence of good competing hypotheses.

The two arguments I want to make for the advantage thesis make connections between the contrast between prediction and accommodation and these relatively uncontroversial evidential and theoretical virtues. The first and simpler of the two arguments is the argument from choice. Scientists can often choose their predictions in a way in which they cannot choose which data to accommodate. When it comes to prediction, they can pick their shots, deciding which predictions of the hypothesis to check. Accommodated data, by contrast, are already there, and scientists have to make out of them what they can. But how can this be used to show that predictions tend to provide stronger support than accommodations? Whatever Karl Popper may have advised, scientists often try to make the strongest possible case for their own hypotheses. So they have a motive for choosing predictions which, if correct, will give maximum support to their hypotheses, not because they are predictions but because they will exhibit the sort of evidential virtues just mentioned. Thus, the scientist will choose predictions that allow for very precise observations that would substantially increase the variety of data supporting the hypothesis, and so on. Successful predictions tend to provide stronger support than accommodations, not directly because they are predictions, but indirectly because scientists have control over which predictions to check; control that is not available in the case of accommodation. The earlier availability of the hypothesis may lead to the gathering of different data and to data that yield stronger support.

That is the argument from choice. It is relatively straightforward, but it does not show quite as much as one might hope. For one thing, scientists do not always have that much choice about what predictions to check: Halley is a case in point. Nevertheless, the argument from choice does show why successful predictions tend sometimes to be more powerful than accommodations, and more powerful because they are predictions. This is what we might call the weak advantage thesis. The point is that the scientist can sometimes use a hypothesis as a guide to data that would provide particularly strong support in the case of prediction, but not in the case of accommodation, because the hypothesis is only antecedently available in the case of prediction. Unfortunately, the argument from choice does not give a reason for the more ambitious claim—the strong advantage thesis—that a single, particular observation that was accommodated would have provided more support for the hypothesis in question if it had been predicted instead. The following analogy may help to clarify this distinction between the weak and strong advantage theses. The fact that I can choose what I eat in a restaurant but not when I am invited to someone's home explains why I tend to prefer the dishes I eat in restaurants over those I eat in other people's homes, but this obviously gives no reason to suppose that lasagna, say, tastes better in restaurants than in homes. Similarly, the argument from choice may show that predictions tend to provide stronger support than accommodations, but it does not show that the fact that a particular datum was predicted gives any more reason to believe the hypothesis than if that same datum had been accommodated. To defend this strong advantage thesis, we need another argument: the ''fudging'' argument.

The fudging argument depends on an interesting feature of the two lists of virtues given above, namely that some of the evidential virtues are in tension with some of the theoretical virtues. Here is an example. On the evidence side, scientists want the supporting evidence to be extensive and varied. On the theoretical side, they want the simplest hypothesis. It is easy to have either one of these virtues on its own. If one just wants lots of varied evidence, one can assemble an encyclopedia of facts; but the hypothesis that is their conjunction will be incredibly complex because the facts are so heterogeneous. On the other hand, if all that matters is simplicity, that too is easy, so long as one does not mind about covering much evidence. What is hard and what scientists want is to satisfy both constraints simultaneously. They want simple hypotheses that nevertheless handle a great diversity of evidence.

Now for the fudging argument. When scientists have data to accommodate, they do the best they can. If the data are diverse, however, this can lead to a sacrifice in simplicity and other theoretical virtues. The epicycles that Ptolemaic astronomers were forced to insert into their planetary model in order to account for the available astronomical data are often taken to be a blatant case of fudging (*12*).

Practicing scientists ought to be able to generate more subtle examples from the recent histories of their own disciplines. The point is that the investigator may, sometimes without fully realizing it, fudge the hypothesis, putting in a few extra epicycles or kinks to ensure that more of the data gets captured. In a case of prediction, by contrast, there is no motive to introduce anything unnatural into the hypothesis, because the investigator does not know the right answer in advance and so would not know what kink to introduce even if one were required. So in that case, the scientist will use the best hypothesis and, if the prediction is successful, will have manifested both empirical and theoretical virtues.

The advantage that the fudging argument attributes to prediction is thus in some respects similar to the advantage of a double-blind medical experiment, in which neither the doctor nor the patients know which patients are getting the placebo and which are getting the drug being tested. The doctor's ignorance makes her judgment more reliable, because she does not know what the "right" answer is supposed to be. The fudging argument makes an analogous suggestion about scientists generally. Not knowing the right answer in advance—the situation in prediction but not in accommodation—makes it less likely that the scientist will fudge the hypothesis in a way that makes for poor empirical support (*13*, *14*).

The claim that the need to accommodate data may lead to fudging ought to be uncontroversial, but does it really show that scientists ought to give any weight to whether the data were accommodated or predicted? A counterargument is that whatever fudging may occur, this is something the scientists can determine directly, by inspection of the hypothesis and the data, once these are given. But this may be to exaggerate scientists' abilities or equivalently to underestimate the complexity of the factors that determine the degree to which the hypothesis is supported by data. Fudging may be neither fully conscious nor readily visible, so the indirect evidence that information about whether the evidence was accommodated or predicted provides may be relevant.

The Twins Revisited

From the perspective of the fudging argument, we are now in a position to see the germs of truth in the three popular arguments for the advantage thesis that we considered above. Recall first the argument that accommodations are worse than predictions because accommodations are ad hoc. My objection was that this leaves the question unanswered or begged, because calling a hypothesis ad hoc either just means it is designed to accommodate, or it simply asserts that an accommodating hypothesis cannot enjoy strong empirical support. The fudging argument provides an independent reason why accommodating systems tend to be ad hoc in the second sense (only weakly supported by the evidence they accommodate); namely, that they tend to suffer from theoretical vices such as excessive complexity and poorly motivated assumptions.

The second argument that I initially rejected was the testing argument: that predictions are better because only they test the hypothesis, because a test is something that can be failed. My objection was that although it may be that only predictions test the scientist, an accommodating hypothesis can be as falsifiable as a predicting one. In such a case, if the accommodated evidence had been different, the hypothesis would have been disconfirmed. The fudging argument is related to the notion that we should after all test the scientist and that she should test herself. It is not simply that she ought to run the risk of being wrong, though that is a consequence of the real point: She should place herself in a situation where she does not know the right answer in advance, because she is then less likely to fudge.

Lastly, the fudging argument suggests an improved version of the argument from the best explanation. In its original form, that argument has it that prediction is better because the best explanation for predictive success is truth, whereas the best explanation for accommodation is instead that the hypothesis was designed for that purpose. My objection was that it was not clear that the accommodating explanation preempts the truth explanation. It is true that only in cases of accommodation can we explain the fit between hypothesis and observation by pointing out that the hypothesis was designed for the purpose, but whether this makes it less likely that the hypothesis is correct is just what we want to know. In contrast, it is clear that the fudging argument competes with the truth explanation. Insofar as we may reasonably infer the explanation that the fit between hypothesis and observation in the case of accommodation is due to fudging, this undermines our confidence in the inference from fit to truth. So the best-explanation account of the difference between accommodation and prediction can be salvaged by replacing the accommodation explanation with the fudging explanation. In cases of accommodation, the inference from fit to truth may be blocked by the inference from fit to the conclusion that the theoretical system is ad hoc, in the pejorative sense.

The fudging argument also gives us an answer to the puzzle of the twin scientists with which this article began and which seemed to show that there could be no relevant difference between prediction and accommodation. We have two scientists who happen to come up with the same hypothesis; where one accommodates evidence, the other predicts. After they compare notes, they should have a common level of confidence in the hypothesis they share. The difficulty was that it seemed impossible to say what level this should be, and how meeting a predictor could be any different for the accommodator than knowing what all accommodators know: that if someone had produced the same hypothesis with less evidence, the prediction of the balance of the data would have been successful.

The answer to this puzzle is now clear. The accommodator ought to worry that she had to do some fudging, but her suspicion is defeasible. One of the things that can defeat it, though not common in the history of science, is meeting a corresponding predictor. If the accommodator meets someone who predicted data she only accommodated, with the same independently produced hypothesis, this shows that the accommodator probably did not fudge to make those accommodations. The predictor, ignorant of the data he predicted, had no motive to fudge the hypothesis to get those results; consequently, the fact that the predictor came up with just the same hypothesis provides independent evidence that the accommodator did not fudge either. It is more likely that neither the predictor nor the accommodator fudged than that they both did. At the same time, merely knowing that the same hypothesis could have been constructed earlier is not relevant, because such a construction might have then required arbitrary and unmotivated fudging. A predictor might have come up instead with a different and better hypothesis. If an actual meeting takes place, however, the twins should leave it sharing the higher confidence of the predictor. As for those of us without scientific siblings, what the fudging argument shows is that we are sometimes justified in being more impressed by predictions than by accommodations.

References and Notes

1. M. Grosser, in *Dictionary of Scientific Biography*, C. C. Gillispie, Ed. (Scribner's, New York, 1970), vol. I, pp. 53–54.
2. J. S. Mill, *A System of Logic* (Longmans, London, ed. 8, 1904), III, xiv, 6.
3. J. M. Keynes, *Collected Writings of John Maynard Keynes Vol. VIII: A Treatise on Probability* (Macmillan, London, 1973).
4. P. Horwich, *Probability and Evidence* (Cambridge Univ. Press, Cambridge, 1982), pp. 108–117.
5. G. Schlesinger, *Australas. J. Philos.* **65**, 33 (1987).
6. S. J. Brush, *Science* **246**, 1124 (1989).
7. K. Popper, *The Logic of Scientific Discovery* (Hutchinson, London, 1959).
8. R. Nozick, in *How Many Questions?*, L. Cauman *et al.*, Eds. (Hackett, Indianapolis, IN, 1983), pp. 105–119.
9. P. Lipton, *Inference to the Best Explanation* (Routledge, London, ed. 2, 2004).
10. C. Glymour, *Theory and Evidence* (Princeton Univ. Press, Princeton, NJ, 1980).
11. C. Howson, P. Urbach, *Scientific Reasoning: The Bayesian Approach* (Open Court, La Salle, IL, 1989).
12. T. Kuhn, *The Copernican Revolution* (Random House, New York, 1957).
13. J. Worrall, in *The Uses of Experiment*, D. Gooding, T. Pinch, S. Schaffer, Eds. (Cambridge Univ. Press, Cambridge, 1989), pp. 135–157.
14. C. Hitchcock, E. Sober, *Br. J. Philos. Sci.* **55**, 1 (2004).

15. Thanks to D. Ruben, C. Howson, D. Papineau, M. Schapiro, T. Williamson, and referees.

10.1126/science.1103024