CHAPTER 1

The paradox of predictivism

1.1 INTRODUCTION

Suppose that after years of living in genteel poverty, you have inherited a small fortune. Having little financial expertise yourself, you decide you are in need of a financial advisor who will help you invest your money wisely. You consult with two candidates, each of whom endorses a particular investment strategy. Each candidate's strategy is based on an account of the forces that induce the value of various investments to fluctuate. In fact, the two advisors can offer detailed explanations of why the value of these investments have changed in the way they have over the past five years. There is, however, one difference between the two advisors: one offered his account *prior* to the beginning of the five-year period, thus successfully *predicting* the various price changes. The other offered her account *after* the five-year period, and thus proposed to explain the price changes after they occurred. Now the question is whether you have, based on just this information, any reason to prefer one advisor over the other. One might insist that the two advisors are on equal ground: both offer accounts that are consistent with the same body of data. But it seems obvious – to many – that there is reason to prefer the advisor who made successful predictions over the one who didn't. If you agree, you may be inclined to endorse a particular view about how evidence confirms theory, a view known as 'predictivism.'

In philosophical parlance, predictivism asserts that, when E is evidence for T, E supports T more strongly when it is a novel confirmation of T than when it is not. Much ink has been spilled over the nature of novelty – but three primary accounts have been suggested: the temporal account, which claims that E is a novel confirmation of T when E is not known at the time T was proposed (Lakatos 1970), the heuristic account (Zaher 1973) which claims that E is a novel confirmation of T when E is not built to fit T, and the theoretical account (Musgrave 1974) which claims E is a novel confirmation for T when E is not explained by any theory other than T. The temporal account was rejected long ago because it made predictivism into a thesis that struck most philosophers as obviously false – for it is absurd to suggest that the confirming power of E literally depends upon the time at which E was discovered. The theoretical account should be dismissed, in my view, because it has the opposite problem: it renders predictivism into a trivial truth, for it is simply obvious that E will confirm T more strongly if there are no plausible competing explanations for T. This leaves the heuristic account of novelty which renders the predictivist thesis into a thesis that is both plausible but non-obvious – and this account will be assumed in the remainder of this chapter (I propose a new account of novelty in Chapter 2). Predictivism now proclaims that, where E is evidence for T, E confirms T more strongly when T was not built to fit E. Applied to our example, this asserts that the price fluctuations confirm the predictor's account of market forces more strongly than the non-predictors because we believe (or at least suspect) that the non-predictor built her account to fit the data, while the predictor didn't. Is this version of predictivism true? In this chapter I will survey some of the particularly prominent aspects of the long and tangled philosophical literature on this question. I begin with a few examples for the sake of motivation.

1.2 PUMPING UP THE INTUITION THAT PREDICTION MATTERS

Let us stipulate – for the moment – that when T is built to fit E, T 'accommodates' E – when T entails E but was not built to fit E, then T 'predicts' E.[1] The examples I cite fall into three categories, which I call dubious accommodations (in which the fact that T was built to fit E renders T 'fishy' in some clear sense), glorious predictions (in which the fact that T predicts E seems extremely impressive evidence for T), and two popular thought experiments.

1.2.1 Dubious accommodations (also known as ad hoc theory rescues)

French bread

Let hypothesis H be: All bread is nourishing. H is believed to be supported by many cases in which it is known that H has proven true. However, it

[1] I am obviously proposing to use the locution 'T predicts E' in a technical sense that does not correspond to the common meaning of this expression which simply means 'T entails E.'

turns out that bread produced in a particular region in France is not nourishing – it is in fact poisonous. Thus H is modified to H': All bread – except that grown in the relevant region, which is poisonous – is nourishing. The modification of H into H' is an attempt to 'rescue' H from falsification, and H' is of course consistent with all known data. Nonetheless, despite this consistency, there seems to be something fishy about H'. The precise nature of the fishiness is, of course, the point in question here. But there is at least a well-established name for the fishy quality – the modification of H into H' is deemed an 'ad hoc theory adjustment.' It is a straightforward example of a dubious accommodation.

The levity of phlogiston

The chemical revolution of the eighteenth century involved the replacement of the phlogiston chemistry, supported by scientists like Priestley and Cavendish, with the "new chemistry" of Lavoisier and others. Proponents of the phlogiston theory had argued that phlogiston was a common component of many substances that was emitted in combustion. The phlogiston theory faced an anomalous result, however, in that some substances actually gained weight when burned. Some proponents of the phlogiston theory, like Gren, responded by suggesting that phlogiston had negative weight – thus the process of emitting phlogiston could explain a corresponding weight gain. This attempt to rescue the phlogiston theory from refutation seemed to many scientists at the time (even to some supporters of the phlogiston theory) to be most unconvincing, despite its ability to reconcile the theory to the data (for discussion see Kitcher 1993: 272–290). Here again we encounter the fishy quality – this time in the form of a hypothesis conjoined with the phlogiston theory to protect that theory from falsification. It is referred to as an 'ad hoc hypothesis' – another form of dubious accommodation.

1.2.2 *Glorious successful predictions*

Psychic

You encounter a self-identified psychic who claims to have knowledge of your personal future. Despite no prior acquaintance with you, she makes a long and precise set of predictions P about your future. P is subsequently fully confirmed – to your utter astonishment. You conclude that the bizarre hypothesis of her clairvoyance is reasonably confirmed. The confirmation of P is an example of what I call a glorious successful prediction – a prediction so bold as to constitute, once confirmed, strong evidence for the hypothesis that predicted it.

Retrograde motion
Nicholas Copernicus proposed a heliocentric theory of the solar system in his 1543 work *De Revolutionibus Orbium Coelestium*. This theory provided a straightforward explanation of the phenomenon of retrograde motion (in which the planets periodically reverse their apparent motion against the fixed stars). By contrast, Ptolemaic astronomy had managed to accommodate retrograde motion only by positing a complex system of epicycles that were built to fit the data. The fact that Copernican astronomy required no elaborate 'fixing up' to accommodate retrograde motion has been equated with the claim that Copernican astronomy predicted retrograde motion (e.g., Scerri and Worrall 2001: 423), and therefore was, by all appearances, more strongly confirmed by it.

1.2.3 *Two thought experiments*

In each of the following two thought experiments there is a theory T and some evidence that supports T – in one scenario T is built to fit the evidence and in the other it isn't. The question is whether the degree of confirmation offered by E to T is the same in the two scenarios.

Connect the dots
The following example is discussed in Howson (1988: 381–382): A coordinate graph represents the relationship between variables x and y – in the first scenario, a set of data points E is plotted on the graph, and the smoothest possible curve is subsequently drawn to fit E – this curve is T. T is a bizarrely irregular curve. In the second scenario, the same curve T is drawn before the accumulation of data – when data is accumulated, E is established, and E falls neatly on T. The question is in which of the two scenarios, if either, is T better confirmed?

Coin flip
This example, due to Maher (1988), also contains two scenarios: in the first scenario, a subject (the accommodator) is presented with E, which is the outcome of a sequence of 99 flips of a single fair coin. E forms an apparently random sequence of heads and tails. The accommodator is then instructed to tell us the outcome of the first 100 flips – he responds by reciting E and then adding the prediction that the 100th toss will be

heads – the conjunction of E and this prediction about the last toss is T. In the other scenario, the subject (the predictor) is asked to predict the first 100 flip outcomes without witnessing any outcomes – the predictor endorses theory T. Thereafter the coin is flipped 99 times, E is established, and the predictor's first 99 predictions are confirmed. The question is in which of these two scenarios is T better confirmed.

These stories all point to the plausibility of predictivism. But before we declare the issue settled, it should be noted there are other examples in which accommodated evidence can provide compelling evidence for a theory – I will call these 'glorious accommodations.' To borrow an example from Mayo (1996: 271), let E contain a list of the SAT scores of all the students in a particular class – a simple averaging technique can be used to generate (T). The average SAT score is x (for some particular value of x). Now clearly T was built to fit E – but E provides compelling evidence for T. Accommodation can, in some cases, provide perfectly compelling evidence for a theory (cf. Hobbes (1993)). Thus the predictivist thesis does not seem to be universally true. Things are turning out to be a bit complex.

1.3 THE PARADOX OF PREDICTIVISM

Predictivism is essentially a comparative thesis – it compares the probative weight of E as accommodated with the probative weight of E as predicted. This is why actual historical examples of dubious accommodations and glorious predictions provide only limited evidence for the truth of predictivism, for any actual historical case presents E in only one of the two roles. We are forced to ask, in the case of the levity of phlogiston for example, how well phlogiston theory would have been confirmed if the increased weight of certain burning substances had somehow been predicted by the phlogiston theory. This forces us to construct in the imagination a counterfactual scenario in which this prediction holds – but it is far from clear how we are to imagine the modified phlogiston theory (for there are a variety of ways in which phlogiston theory could be imagined to entail it). Hence the attractive quality of the two thought experiments, in which it seems easy to keep the two scenarios the same except for changing the role of E.

It is high time to note that as an account of how theories are evaluated in science, predictivism is deeply controversial. This is primarily because predictivism makes facts about how theories were constructed relevant to the epistemic assessment of theories. Specifically, predictivism entails that it matters with respect to the assessed probability of T whether it was built to fit any of its supporting evidence. But this makes the assessed probability of T curiously dependent on the mental life of its constructor, specifically on the knowledge and intentions of that theorist to build a theory that accommodated certain data rather than others. This means that scientists who want to assess the probability of some theory need to do more than simply know all the relevant evidence and criteria of theory assessment and apply them to the theory – they need to know facts about the biography of the theorist who constructed the theory. In what follows I will use the term 'biographicalism' to denote the view that facts about the life stories of scientists are epistemically relevant to the assessment of theories. Predictivism as defined here entails the truth of biographicalism, and this is why many philosophers reject it.

Philosophical literature on this subject abounds with passionate rejections of biographicalism. For one thing, if it were the case that biographical information were crucial for the proper evaluation of theories, then such proper evaluation would be difficult if not impossible for many working scientists to perform, given that the relevant biographical information (about what the scientist who constructed a theory knew and intended) is not always provided in the scientific literature that such scientists read, and can be unearthed only by subsequent historians of science who study the archives of theorists. Clearly "the fruits of such historical research are usually unavailable to the scientists involved in an ongoing rivalry between research programmes. The decisions of most of them about which programme to work on or believe in cannot, of course, be influenced by discoveries made by historians examining unpublished (sometimes oral) information – many years, perhaps after the rivalry has ended" (Gardner 1982: 6). Furthermore biographical facts seem to be widely ignored in scientific literature, except where issues of priority are concerned (Thomason 1992: 195). But even more to the point, the very idea of biographicalism flies in the face of what appear to be obvious facts about the objectivity of scientific method. "The extent to which a given postulate is confirmed by the evidence at hand is a function of the precise contents of that postulate and the nature of that evidence, and definitely not of such fortuitous factors such as the time when someone happened to hit upon the idea of formulating that postulate, or the stage at which it has in practice been employed and for what purpose" (Schlesinger 1987: 33). Leplin similarly points to the absurdity of biographicalism: "The theorist's hopes, expectations, knowledge, intentions, or whatever, do not seem to relate to the epistemic standing of his theory in a way that can sustain a pivotal role for them . . ." (1997: 54). (See also Collins (1994) and Achinstein (2001: 210–230).)

A paradox is a statement that appears as though it cannot be true and yet somehow seems as though it must be true. Predictivism is genuinely paradoxical: it seems too obvious to question when we consider examples like those adduced in the previous section, but it seems too absurd to even consider when we consider its commitment to biographicalism as an account of scientific method. In the next section we consider some of the key moments in the history of the vast philosophical literature on predictivism with the hope that the paradox may be illuminated.

1.4 A SKETCHY HISTORY OF PREDICTIVISM

1.4.1 John Herschel, William Whewell, and the method of hypothesis[2]

There was in the eighteenth century a passionate debate about scientific method. Newton's *Principia* (1687) had famously repudiated the use of hypotheses in science, and many subsequently followed Newton in holding that the only theories worth considering were those that did not postulate unobservable entities but were straightforwardly 'deduced from the phenomena.' But other scientists defended the method of hypothesis by which theories about unobservables were postulated and shown to save the phenomena. One important proponent of the method of hypothesis was David Hartley, whose *Observations of Man* (which first appeared in (1749)) claimed that the brain and nervous system are filled with a 'subtle fluid' – or aether – which transmits vibrations from one point in the perceptual system to another. Hartley used this basic picture to concoct explanations of a vast array of phenomena, including sleep, the generation of simple and complex ideas, paralysis, taste, sexual desire, memory, and many more. Hartley defended his theory by pointing to the great number of phenomena which it could explain. He writes:

> Let us suppose the existence of the aether, with these its properties, to be destitute of all direct evidence, still, *if it serves to explain a great variety of phenomena, it will have an indirect evidence in its favour by this means*. (*Observations of Man*, London, 1791, vol. I, p. 15, quoted in Laudan 1981: 115, italics in Laudan)

Hartley defends the method of hypothesis using a number of arguments, including appeals to its heuristic value (it can lead to the discovery of new phenomena and generate new experiments) and the fact that the method

[2] For my account in this section I have relied heavily on Laudan (1981: 111–140).

can produce hypotheses which, though not fully proven, are nonetheless 'the best science can do' in certain contexts.

Hartley's defense of the method of hypothesis met with scathing criticism. For as his critics pointed out "there were many rival systems of natural philosophy which – after suitable *ad hoc* modifications – could be reconciled with all the known phenomena. The physics of Descartes, the physiology of Galen, and the astronomy of Ptolemy would all satisfy Hartley's criterion. There was, in Hartley's approach to the epistemology of science, nothing which would discredit the strategy of saving a discarded hypothesis by cosmetic surgery or artificial adjustments to it" (Laudan 1981: 117). This had, after all, been the point of Newton's repudiation of the method of hypothesis.

But something fundamental happened in the early nineteenth century to change all this. The story involves a glorious prediction. At this time most physicists held the 'emission theory' of light which claimed that light consisted of particles. However in 1819 the young French physicist Augustin Fresnel wrote an essay propounding a version of the wave theory of light and showed that various known facts supported his theory. He submitted his essay to a science competition held by the French Academy. One of the prize commission's judges, Poisson, demonstrated that Fresnel's wave theory of light had an obviously absurd consequence: it implied that if a small circular disk were used to create a shadow cast by light from a small hole, the center of the disk's shadow would contain a spot of bright light. This consequence was tested by Arago, who with much amazement found the bright spot to be exactly where Fresnel's theory entailed it should be. Fresnel won the prize in question. According to a common version of the story, this glorious successful prediction proved a turning point in the physics community with respect to the nature of light, and by the mid to late 1820s the wave theory was generally accepted in the scientific community.[3]

Light waves, of course, are unobservable entities – and thus the wave theory of light had been formulated by an application of the method of hypothesis. But in this case, unlike that of Hartley's aether theory of the nervous system, the theory had led to an astonishing successful prediction. This led individuals like John Herschel and William Whewell to argue that the method of hypothesis was legitimate when it produced hypotheses

[3] For discussion of relevant points, including the shortcomings of this 'general version' of the story, see Worrall (1989). See also Achinstein (1991 Part 1), which argues that the methods used by proponents and opponents of the Method of Hypothesis actually had many commonalities.

that led to the successful prediction of previously unknown phenomena. Successful prediction thus served to distinguish the arbitrary and ad hoc products of hypothesizing like Hartley's aetherial account of the nervous system from legitimate products like the wave theory of light. In this context Whewell and Herschel embraced predictivism. Insofar as the wave theory was widely accepted by the scientific community, it seemed as though Whewell's and Herschel's predictivism was also widely accepted.

There was just one problem with the status of predictivism in the early nineteenth century: according to Laudan, neither Whewell nor Herschel offered any logical or epistemological explanation of why it held true (1981: 131, 134). There was no attempt on their part, e.g., to explain the oddly biographical picture of scientific method it entailed. It was simply offered as a rule that scientists should follow.[4] Whewell and Herschel thus advocated predictivism while ignoring the obvious problems it raised. John Stuart Mill in his *System of Logic* responded to Whewell by emphatically denying that any special significance should be assigned to the fact that an established consequence of a theory was predicted rather than accommodated (1961: 328–329). The paradox of predictivism had emerged – but no clues about its resolution were yet to be found.

1.4.2 Popper, Lakatos, and the fallacy of pure consequentialism

Karl Popper is probably the most famous proponent of the preference for prediction in the entire history of philosophy. In his lecture "Science: Conjectures and Refutations" Popper recounts his boyhood attempt to grapple with the question '*When should a theory be ranked as scientific?*' (Popper 1963: 33–65). Popper had become convinced that certain popular theories of his day, including Marx's theory of history and Freudian psychoanalysis, were pseudosciences. Though they superficially resembled genuinely scientific theories, like Einstein's theory of relativity, they fell short of scientific status. Popper deemed the problem of distinguishing scientific from pseudoscientific theories the 'demarcation problem.' His solution to the demarcation problem, as everyone knows, was to identify the quality of falsifiability (or 'testability') as the mark of the scientific theory.

[4] Worrall (1985: 324) responds to Laudan by claiming that Whewell did purport to give a rationalization of predictivism along the following lines: predictions carry special weight because a theory that correctly predicts a surprising result cannot have done so by chance, and thus must be true. Insofar as this counts as an argument at all it appears to be a version of what I call the miracle argument for strong predictivism – I discuss this argument and argue that it is fallacious in Chapter 4.

The pseudosciences were marked, Popper claimed, by their vast explanatory power. They could explain not only all the relevant actual phenomena the world presented, they could explain any *conceivable* phenomena that fell within their domain. This was because the explanations offered by the pseudosciences were sufficiently malleable that they could always be adjusted ex post facto to explain anything. Thus the pseudosciences never ran the risk of being inconsistent with the data. By contrast, a genuinely scientific theory – such as Einstein's theory of relativity – made specific predictions about what should be observed and thus ran the risk of falsification. Popper emphasized that what established the scientific character of relativity theory was that it 'stuck its neck out' in a way that pseudosciences never did.

At one level all of this should strike us as familiar. The theories Popper identifies as pseudoscientific are strikingly reminiscent of Hartley's aetherial theory discussed above, which likewise made no specific predictions but simply offered ex post facto explanations of phenomena it was built to fit. Like Whewell and Herschel, Popper appeals to the predictions a theory makes as a way of separating the illegitimate uses of the method of hypothesis from its legitimate uses. But there was a big difference as well. Whewell and Herschel pointed to predictive success as a necessary condition for the acceptability of a theory that had been generated by the method of hypothesis. Popper by contrast focuses in his solution to the demarcation problem not on the success of a prediction but on the fact that the theory made the prediction at all – as noted above, what marked a theory as scientific was not that its detailed predictions were confirmed, but simply that it made specific predictions which established a risk of falsification. This criterion of scientific status would have worked to disqualify Hartley's aetherial theory as well as it did to disqualify Freudian psychoanalysis.[5] Of course, there was for Popper an important difference between scientific theories whose predictions were confirmed and those whose predictions were falsified. Falsified theories were to be rejected, whereas theories that survived such testing were to be 'tentatively accepted' until such time as they might be falsified. Popper did not hold, with Whewell and Herschel, that successful predictions could constitute legitimate proof of a theory – in fact, Popper held that it was impossible to show that a theory was even probable on the basis of the evidence, for he embraced Hume's critique of inductive logic which made evidential

[5] Popper's argument that Freudian psychoanalysis is unfalsifiable has been questioned by Grunbaum (1984).