5 The epistemology of scientific theorizing

Overview

Suppose we settle the dispute between realism and instrumentalism. The problem still remains of how exactly observation and evidence, the collection of data, etc., actually enable us to choose among scientific theories. On the one hand, that they do so has been taken for granted across several centuries of science and its philosophy. On the other hand, no one has fully explained how they do so, and in this century the challenges facing the explanation of exactly how evidence controls theory have increased.

A brief review of the history of British empiricism sets the agenda for an account of how science produces knowledge justified by experience. Even if we can solve the problem of induction raised by Hume, or show that it is a pseudo-problem, we must face the question of what counts as evidence in favor of a hypothesis. The question seems easy, but it turns out to be a very complex one on which the philosophy of science has shed much light without answering to every one's satisfaction.

Modern science makes great use of statistical methods in the testing of hypotheses. We explore the degree to which a similar appeal to probability theory on behalf of philosophy can be used adequately to express the way data support theory. Just as the invocation of probability in Chapter 2 leads to questions of how we are to understand this notion, invoking it to explain confirmation of hypotheses forces us to choose among alternative interpretations of probability.

Even if we adopt the most widely accepted account of theory confirmation, we face a further challenge: the thesis of underdetermination, according to which even when all the data is in, the data will not by themselves choose among competing scientific theories. Which theory, if any, is the true theory may be underdetermined by the evidence even when all the evidence is in. This conclusion, to the extent it is adopted, not only threatens the empiricist's picture of how knowledge is certified in science but threatens the whole edifice of scientific objectivity altogether, as Chapter 6 describes.

5.1 A brief history of empiricism as science's epistemology

The scientific revolution began in central Europe with Copernicus, Brahe and Kepler, shifted to Galileo's Italy, moved to Descartes's France and ended with Newton in Cambridge, England. The scientific revolution was also a philosophical revolution, and for reasons we have already noted. In the seventeenth century science was "natural philosophy", and figures that history would consign exclusively to one or the other of these fields contributed to both. Thus Newton wrote a good deal of philosophy of science, and Descartes made contributions to physics. But it was the British empiricists who made a self-conscious attempt to examine whether the theory of knowledge espoused by these scientists would vindicate the methods which Newton, Boyle, Harvey, and other experimental scientists employed to expand the frontiers of human knowledge so vastly in their time.

Over a period from the late seventeenth century to the late eighteenth century, John Locke, George Berkeley and David Hume sought to specify the nature, extent and justification of knowledge as founded on sensory experience and to consider whether it would certify the scientific discoveries of their time as knowledge and insulate them against skepticism. Their results were mixed, but nothing would shake their confidence, or that of most scientists, in empiricism as the right epistemology.

Locke sought to develop empiricism about knowledge, famously holding against rationalists like Descartes, that there are no innate ideas. "Nothing is in the mind that was not first in the senses." But Locke was resolutely a realist about the theoretical entities which seventeenth-century science was uncovering. He embraced the view that matter was composed of indiscernible atoms, "corpuscles" in the argot of the time, and distinguished between material substance and its properties on the one hand, and the sensory qualities of color, texture, smell or taste, which matter causes in us. The real properties of matter, according to Locke, are just the ones that Newtonian mechanics tells us it has – mass, extension in space, velocity, etc. The sensory qualities of things are ideas in our heads which the things cause. It is by reasoning back from sensory effects to physical causes that we acquire knowledge of the world, which gets systematized by science.

That Locke's realism and his empiricism inevitably give rise to skepticism, is not something Locke recognized. It was a philosopher of the next generation, George Berkeley, who appreciated that empiricism makes doubtful our beliefs about things we do not directly observe. How could

Locke lay claim to the certain knowledge of the existence of matter or its features, if he could only be aware of sensory qualities, which by their very nature, exist only in the mind? We cannot compare sensory features like color or texture to their causes to see whether these causes are colorless or not, for we have no access to these things. And to the argument that we can imagine something to be colorless, but we cannot imagine a material object to lack extension or mass, Berkeley retorted that sensory properties and non-sensory ones are on a par in this respect: try to image something without color. If you think of it as transparent, then you are adding in the background color and that's cheating. Similarly for the other allegedly subjective qualities that things cause us to experience.

In Berkeley's view, without empiricism we cannot make sense of the meaningfulness of language. Berkeley pretty much adopted the theory of language as naming sensory qualities that was sketched in the last chapter. Given the thesis that words name sensory ideas, realism – the thesis that science discovers truths about things we cannot have sensory experience of – becomes false, for the words that name these things must be meaningless. In place of realism Berkeley advocated a strong form of instrumentalism and took great pains to construct an interpretation of seventeenth- and eighteenth-century science, including Newtonian mechanics, as a body of heuristic devices, calculating rules, and convenient fictions, we employ to organize our experiences. Doing this, Berkeley thought, saves science from skepticism. It did not occur to Berkeley that another alternative to the combination of empiricism and instrumentalism is rationalism and realism. And the reason is that by the eighteenth century, the role of experiment in science was so securely established that no alternative to empiricism seemed remotely plausible as an epistemology for science.

Indeed, it was David Hume's intention to apply what he took to be the empirical methods of scientific inquiry to philosophy. Like Locke and Berkeley he sought to show how knowledge, and especially scientific knowledge, honors the strictures of empiricism. Unable to adopt Berkeley's radical instrumentalism, Hume sought to explain why we adopt a realistic interpretation of science and ordinary beliefs, without taking sides between realism and instrumentalism. But, as we saw in Chapter 3, Hume's pursuit of the program of empiricism led him to face a problem different from that raised by the conflict of realism and empiricism. This is the problem of induction: given our current sensory experience, how can we justify inferences from them and from our records of the past, to the future and to the sorts of scientific laws and theories we seek?

Hume's argument is often reconstructed as follows: there are two and only two ways to justify a conclusion: deductive argument, in which the conclusion follows logically from the premises, and inductive argument, in which the premises support the conclusion but do not guarantee it. A deductive argument is colloquially described as one in which the premises "contain" the conclusion, whereas an inductive argument is often described as one that moves from the particular to the general, as when we infer from observation of 100 white swans to the conclusion that all swans are white. Now, if we are challenged to justify the claim that inductive arguments – arguments from the particular to the general, or from the past to the future – will be reliable in the future, we can do so only by employing a deductive argument or an inductive argument. The trouble with any deductive argument to this conclusion is that at least one of the premises will itself require the reliability of induction. For example, consider the deductive argument below:

1 If a practice has been reliable in the past, it will be reliable in the future.
2 In the past inductive arguments have been reliable.

Therefore:

3 Inductive arguments will be reliable in the future.

This argument is deductively valid, but its first premise requires justification and the only satisfactory justification for the premise would be the reliability of induction, which is what the argument is supposed to establish. Any deductive argument for the reliability of induction will include at least one question-begging premise. This leaves only inductive arguments to justify induction. But clearly, no inductive argument for induction will support its reliability, for such arguments too are question-begging. As we have had occasion to note before, like all such question-begging arguments, an inductive argument for the reliability of induction is like underwriting your promise to pay back a loan by promising that you keep your promises. If your reliability as a promise keeper is what is in question, offering a second promise to assure the first one is pointless. Hume's argument has for 250 years been treated as an argument for skepticism about empirical science, for it suggests that all conclusions about scientific laws, and all predictions science makes about future events, are at bottom unwarranted, owing to their reliance on induction. Hume's own conclusion was quite different. He noted that as a person who acts in the world, he was satisfied that inductive arguments were reasonable; what he thought the argument shows is that we have not yet found the right justification for induction, not that there is no justification for it.

The subsequent history of empiricism shares Hume's belief that there is a justification for induction, for empiricism seeks to vindicate empirical science as knowledge. Throughout the nineteenth century philosophers like John Stuart Mill sought solutions to Hume's problem. In the twentieth century many logical positivists, too, believed that a solution could be found for the problem of induction. One such positivist argument (due to Hans Reichenbach) seeks to show that if any method of predicting the future works, then induction must work. Suppose we wish to establish whether the oracle at Delphi is an accurate predictive device. The only way to do so is to subject the oracle to a set of tests: ask for a series of predictions and determine whether they are verified. If they are, the oracle can be accepted as an accurate predictor. If not, then the future accuracy of the oracle is not to be relied upon. But notice that the form of this argument is inductive. If any method works (in the past), only induction can tell us that it does (in the future). Whence we secure the justification of induction. This argument faces two difficulties. First, at most it proves that if any method works, induction works. But this is a far cry from the conclusion we want: that any method does in fact work. Second, the argument will not sway the devotee of the oracle. Oracle-believers will have no reason to accept our argument. They will ask the oracle whether induction works, and will accept its pronouncement. No attempt to convince oracle-believers that induction supports either their method of telling the future or any other can carry any weight with them. The argument that if any method works, induction works, is question-begging, too.

Other positivists believed that the solution to Hume's problem lay in disambiguating various notions of probability, and applying the results of a century's advance in mathematical logic to Hume's empiricism. Once the various senses of probability employed in science were teased apart, they hoped either to identify the one that is employed in scientific reasoning from data to hypotheses, or to explicate that notion to provide a "rational reconstruction" of scientific inference that vindicates it. Recall the strategy of explicating scientific explanation as the D-N model. The positivists spent more time attempting to understand and explicate the logic of the experimental method – inferring from data to hypotheses – than on any other project in the philosophy of science. The reason is obvious. Nothing is more essential to science than learning from experience; that is what is meant by empiricism. And they believed this was the way to find a solution to Hume's problem.

Some of what Chapter 3 reports about interpretations of probability reflects the work of these philosophers. In this chapter we will encounter more of what they uncovered about probability. What these philosophers and their students discovered about the logical foundations of probability and of the experimental method in general, turned out to raise new problems beyond those which Hume laid before his fellow empiricists.

5.2 The epistemology of scientific testing

There is a great deal of science to do long before science is forced to invoke unobservable things, forces, properties, functions, capacities and dispositions to explain the behavior of things observable in experience and the lab. Even before we infer the existence of theoretical entities and processes, we are theorizing. A scientific law, even one exclusively about what we can observe, goes beyond the data available, because it makes a claim which if true is true everywhere and always, not just in the experience of the scientist who formulates the scientific law. This of course makes science fallible: the scientific law, our current best-estimate hypothesis may turn out to be, in fact, usually does turn out to be wrong. But it is by experiment that we discover this, and by experiment that we improve on it, presumably getting closer to the natural law we seek to discover.

It may seem a simple matter to state the logical relationship between the evidence that scientists amass and the hypotheses the evidence tests. But philosophers of science have discovered that testing hypotheses is by no means an easily understood matter. From the outset it was recognized that no general hypothesis of the form "All **A**s are **B**s" – for instance, "All samples of copper are electrical conductors" – could be conclusively confirmed because the hypothesis will be about an indefinite number of **A**s and experience can provide evidence only about a finite number of them. By itself a finite number of observations, even a very large number, might be only an infinitesimally small amount of evidence for a hypothesis about a potentially infinite number of, say, samples of copper. At most, empirical evidence supports a hypothesis to some degree. But as we shall see, it may also support many other hypotheses to an equal degree.

On the other hand, it may seem that such hypotheses could at least be falsified. After all, to show that "All **A**s are **B**s" is false, one need only find an **A** which is not a **B**: after all, one black swan refutes the claim that all swans are white. And understanding the logic of falsification is particularly important because science is fallible. Science progresses by subjecting a hypothesis to increasingly stringent tests, until the hypothesis is falsified, so that it may be corrected, improved, or give way to a better hypothesis. Science's increasing approximation to the truth relies crucially on falsifying tests and scientists' responses to them. Can we argue that while general hypotheses cannot be completely confirmed, they can be completely or "strictly" falsified? It turns out that general hypotheses are not strictly falsifiable, and this will be a fact of the first importance in Chapter 6. Strict falsifiability is impossible, for nothing follows from a general law alone. From "All swans are white", it does not follow that there are any white swans; it doesn't even follow that there are any swans at all. To test this generalization we need to independently establish that there is at least one swan and then check its color. The claim that there is a swan, the claim that we can establish its actual color just by looking at it, are "auxiliary hypotheses" or "auxiliary assumptions". Testing even the simplest hypothesis requires "auxiliary assumptions" – further statements about the conditions under which the hypothesis is tested. For example, to test "All swans are white", we need to establish that "this bird is a swan", and doing so requires we assume the truth of other generalizations about swans besides what their color is. What if the grey bird before us is a grey goose, and not a grey swan? No single falsifying test will tell us whether the fault lies with the hypothesis under test or with the auxiliary assumptions we need to uncover the falsifying evidence.

To see the problem more clearly consider a test of $\mathbf{PV} = \mathbf{rT}$. To subject the ideal gas law to test we measure two of the three variables, say, the volume of the gas container and temperature, use the law to calculate a predicted pressure, and then compare the predicted gas pressure to its actual value. If the predicted value is identical to the observed value, the evidence supports the hypothesis. If it does not, then presumably the hypothesis is falsified. But in this test of the ideal gas law, we needed to measure the volume of the gas and its temperature. Measuring its temperature requires a thermometer, and employing a thermometer requires us to accept one or more rather complex hypotheses about how thermometers measure heat, for example, the scientific law that mercury in an enclosed glass tube expands as it is heated, and does so uniformly. But this is another general hypothesis – an auxiliary we need to invoke in order to put the ideal gas law to the test. If the predicted value of the pressure of the gas diverges from the observed value, the problem may be that our thermometer was defective, or that our hypothesis about how expansion of mercury in an enclosed tube measures temperature change is false. But to show that a thermometer was defective, because, say, the glass tube was broken, presupposes another general hypothesis: thermometers with broken tubes do not measure temperature accurately. Now in many cases of testing, of course, the auxiliary hypotheses are among the most basic generalizations of a discipline, like acid turns red litmus paper blue, which no one would seriously challenge. But the logical possibility that they might be mistaken, a possibility that cannot be denied, means that any hypothesis which is tested under the assumption that the auxiliary assumptions are true, can be in principle preserved from falsification, by giving up the auxiliary assumptions and attributing the falsity to these auxiliary assumptions. And sometimes, hypotheses are in practice preserved from falsification. Here is a classic example in which the falsification of a test is rightly attributed to the falsity of auxiliary hypotheses and not the theory under test. In the nineteenth century predictions of the location in the night sky of Jupiter and Saturn derived from Newtonian mechanics were falsified as telescopic observation improved. But instead of blaming the falsification on Newton's laws of motion, astronomers challenged the auxiliary assumption that there were no other forces, beyond those due to the known planets, acting on Saturn and Jupiter. By calculating how much additional gravitational force was necessary and from what direction, to render Newton's laws consistent with the data apparently falsifying them, astronomers were led to the discovery, successively, of Neptune and Uranus.

As a matter of logic, scientific law can neither be completely established by available evidence, nor conclusively falsified by a finite body of evidence. This does not mean that scientists are not justified on the occasions at which they surrender hypotheses because of countervailing evidence, or accept them because of the outcome of an experiment. What it means is that confirmation and disconfirmation are more complex matters than the mere derivation of positive or negative instances of a hypothesis to be tested. Indeed, the very notion of a positive instance turns out to be a hard one to understand.

Consider the hypothesis that "All swans are white". Here is a white bird which is a swan and a black boot. Which is a positive instance of our hypothesis? Well, we want to say that only the white bird is; the black boot has nothing to do with our hypothesis. But logically speaking, we have no right to draw this conclusion. For logic tells us that "All **A**s are **B**s" if and only if "All non-**B**s are non-**A**s". To see this, consider what would be an exception to "All **A**s are **B**s". It would be an **A** that was not a **B**. But this would also be the only exception to "All non-**B**s are non-**A**s". Accordingly, statements of these two forms are logically equivalent. In consequence, all swans are white if and only if all non-white things are non-swans. The two sentences are logically equivalent formulations of the same statement. Since the black boot is a non-white non-swan, it is a positive instance of the hypothesis that all non-white things are non-swans, aka all swans are white. The black boot is a positive instance of the hypothesis that all swans are white. Something has gone seriously wrong here! Surely the way to assess a hypothesis about swans is not to examine boots! At a minimum, this result shows that the apparently simple notion of a "positive instance" of a hypothesis is not so simple, and one we do not yet fully understand.

One conclusion drawn from the difficulty of this problem supports Popper's notion that scientists don't or at least shouldn't try to confirm hypotheses by piling up positive instances. They should try to falsify their hypotheses by seeking counterexamples. But the problem of scientific testing is really much deeper than simply the difficulty of defining a positive instance.

Consider the general hypothesis that "All emeralds are green". Surely a green emerald is a positive instance of this hypothesis. Now define the term "grue" as "green at time **t** and **t** is before 2100 AD or it is blue at **t** and **t** is after 2100 AD". Thus, after 2100 AD a cloudless sky will be grue, and any emerald already observed is grue as well. Consider the hypothesis "All emeralds are grue". It will turn out to be the case that every positive instance so far observed in favor of "All emeralds are green" is apparently a positive instance of "All emeralds are grue", even though the two hypotheses are incompatible in their claims about emeralds discovered after 2100 AD. But the conclusion that both hypotheses are equally well confirmed is absurd. The hypothesis "All emeralds are grue" is not just less well confirmed than "All emeralds are green", it is totally without evidential support altogether. But this means that all the green emeralds thus far discovered are not after all "positive instances" of "All emeralds are grue" – else it would be a well-supported hypothesis since there are very many green emeralds and no non-green ones. But if green emeralds are not positive instances of the grue-hypothesis, then we need to give a reason why they are not.

We could restate the problem as one about falsification, too. Since every attempt to falsify "All emeralds are green" has failed, it has also failed to falsify "All emeralds are grue". Both hypotheses have withstood the same battery of scientific tests. They are equally reasonable hypotheses. But this is absurd. The grue hypothesis is not one we would bother with for a moment, whether our method was seeking to confirm or to falsify hypotheses. So, our problem is not one that demanding science seek only falsification will solve.

One is inclined to respond to this problem by rejecting the predicate "grue" as an artificial, gerrymandered term that names no real property. "Grue" is constructed out of the "real properties" green and blue, and a scientific hypothesis must employ only real properties of things. Therefore, the grue-hypothesis is not a real scientific hypothesis and it has no positive instances. Unfortunately this argument is subject to a powerful reply. Define bleen as "blue at **t** and **t** is earlier than 2100 AD and green at **t** when **t** is later than 2100 AD". We may now express the hypothesis that all emeralds are green as "All emeralds are grue at **t** and **t** is earlier than 2100 AD or bleen at **t** and **t** is later than 2100 AD". Thus, from the point of view of scientific language, "grue" is an intelligible notion. Moreover, consider the definition of "green" as "grue at **t** and **t** is earlier than 2100 AD or bleen at **t** and **t** is later than 2100 AD". What is it that prevents us from saying that green is the artificial, derived term, gerrymandered from "grue" and "bleen"?

What we seek is a difference between "green" and "grue" that makes "green" admissible in scientific laws and "grue" inadmissible. Following Nelson Goodman, who constructed the problem of "grue", philosophers have coined the term "**projectable**" for those predicates which are admissible in scientific laws. So, what makes "green" projectable? It cannot be that "green" is projectable because "All emeralds are green" is a well-supported law. For our problem is to show why "All emeralds are grue" is not a well-supported law, even though it has the same number of positive instances as "All emeralds are green". The puzzle of "grue", known as "the new riddle of induction", remains an unsolved problem in the theory of confirmation. Over the decades since its invention philosophers have offered many solutions to the problem, no one of which has gained ascendancy. But the inquiry has resulted in a far greater understanding of the dimensions of scientific confirmation than the logical positivists or their empiricist predecessors recognized. One thing all philosophers of science agree on is that the new riddle shows how complicated the notion of confirmation turns out to be, even in the simple cases of generalizations about things we can observe.

5.3 Induction as a pseudo-problem: Popper's gambit

Sir Karl Popper was among the most influential of twentieth-century philosophers of science, perhaps more influential among scientists, especially social scientists, than he was among philosophers. Popper is famous among philosophers for arguing that Hume's problem of induction is a sort of pseudo-problem, or at least a problem which should not detain either scientists or those who seek to understand the methods of science. The problem of induction is that positive instances don't seem to increase our confidence in a hypothesis, and the new riddle of induction is that we don't even seem to have a good account of what a positive instance is. These are not problems for science, according to Popper, since science is not, and should not be in the business of piling up positive instances that confirm hypotheses. Popper held that as a matter of fact, scientists seek negative evidence against, not positive evidence for, scientific hypotheses, and that as a matter of method, they are correct to do so. If the problem of induction shows anything, it shows that they should not seek to confirm hypotheses by adding to evidence for them. Instead good scientific method, and good scientists, seek only to falsify hypotheses, to find evidence against them, and when they succeed in falsifying, as inevitably they will (until science is "complete" – a state of affairs we won't be able to realize we have attained), scientists do and should go on to frame new hypotheses and seek their falsification, world without end.

Popper's argument for this methodological prescription (and the descriptive claim that it is what scientists actually do) begins with the observation that in science we seek universal generalizations and that as a matter of their logical form, "All **F**s are **G**s", they can never be completely confirmed, established, verified, since the (inductive) evidence is always incomplete; but they can as a matter of logic be falsified by only one counterexample. Of course as we have seen, logically speaking, falsification is no easier than verification, owing to the role of auxiliary assumptions required in the test of any general hypothesis. If Popper did not recognize this fact initially, he certainly came to accept that strict falsification is impossible. His claim that scientists do and should seek to frame hypotheses, "conjectures" he called them, and subject them to falsification, "refutation" he sometimes labeled it, must be understood as requiring something different from strict falsification.

Recall in Chapter 2 the example of one sentence expressing more than a single proposition. Depending on the emphasis the sentence "Why did Mrs R kill Mr R with a knife?" can express three distinct questions. Now consider the sentence, "All copper melts at 1,083 degrees centigrade." If we define copper as the "the yellowish-greenish metal which conducts electricity and melts at 1,083 degrees centigrade", then of course the hypothesis "All copper melts at 1,083 degrees centigrade" will be unfalsifiable owing to the meanings of the words. Now, suppose you define copper in the same way, except that you strike from the definition the clause about melting point, and then test the hypothesis. This will presumably eliminate the unfalsifiability due to meaning alone. Now suppose that for many samples you identify as copper, they either melt well below or well above 1,083 degrees centigrade on your thermometer, and in each case you make *an excuse* for this experimental outcome: the thermometer was defective, or there were impurities in the sample, or it wasn't copper at all, but some similar yellowish-greenish metal, or it was aluminum and illuminated by yellowish-greenish light, or you were suffering from a visual disorder when you read the thermometer, or . . . The ellipses are meant to suggest that an indefinitely large number of excuses can be cooked up to preserve a hypothesis from falsification. Popper argued that such a stratagem – treating a hypothesis as unfalsifiable – is unscientific. Scientific method requires that we envision circumstances which we would count as actually leading us to give up our hypotheses, and that we subject these hypotheses to test under these conditions. Moreover, Popper argued the best science is characterized by framing hypotheses that are highly risky – making claims it is easy to test, testing them, and when they fail these tests (as eventually they must), framing new risky hypotheses. Thus, as noted above, he characterized scientific method as "conjectures and refutations" in a book of that title. Like other philosophers of science, including the logical positivists with whom Popper claimed to disagree on most fundamental issues in philosophy, Popper had nothing much to say about the "conjecture" part of science. Philosophers of science have held by and large that there is no logic of discovery, no recipe for how to come up with significant new scientific hypotheses. But Popper did hold that scientists should advance "risky" hypotheses, ones it would be easy to imagine disconfirming evidence against. And he held that the business of experiment is to seek such disconfirmation.

So Popper's claim about falsifiability may be best treated as a description of the attitudes of scientists towards their hypotheses, and/or a prescriptive claim about what the attitudes of good scientists should be, instead of a claim about statements or propositions independent of attitudes towards their testing. It was on this basis that he famously stigmatized Freudian psychodynamic theory and Marx's dialectical materialism as unscientific, employing the possibility of falsification as a criterion to "demarcate" science from pseudo-science. Despite the pretensions of the exponents of these two "theories", neither could be counted as scientific, for as "true believers" their exponents would never countenance counterexamples to them that require the formulation of new conjectures. Therefore, Popper held their beliefs were not properly to be considered scientific theories at all, not even repudiated ones. At one point Popper also treated Darwin's theory of natural selection as unfalsifiable, owing in part to the proclivity of biologists to define fitness in terms of reproductive rates and so turn the PNS (see Chapter 4, Section 4.5) into a definition. Even when evolutionary theorists are careful not to make this mistake, Popper held that the predictive content of adaptational hypotheses was so weak that falsification of the theory was impossible. Since repudiating Darwin's theory was hardly plausible, Popper allowed that though it was not a scientific theory strictly speaking, it was a valuable metaphysical research program. Of course, Marxian and Freudian theorists would have been able to make the same claim. More regrettably, religiously inspired opponents of the theory of natural selection were only too happy to cloak themselves in the mantle of Popper: they argued that either Christian metaphysics had to share equal time with Darwinian metaphysics in science class-rooms, or the latter should be banished along with the former. It is worth noting for the record that Darwin faced the challenge Popper advances, of identifying circumstances that would falsify his theory, in Chapter 6 of *On the Origin of Species*, entitled "Difficulties of the theory".

This stigmatization of some theories as pseudo-science was subsequently adopted, especially by economic theorists. This may well have been because of Popper's personal influence on them, or owing to his other writings attacking Marxian political economy and political philosophy, with which these social scientists found common cause. The embrace of Popper, by economic theorists particularly, was ironic in two respects. First, their own practice completely belied Popper's maxims. For more than a century economic theorists (including the Popperians among them) have been utterly committed to the generalization that economic agents are rational preference maximizers, no matter how much evidence behavioral, cognitive and social psychologists have built up to disconfirm this generalization. Second, in the last two decades of the twentieth century the persistence in this commitment to economic rationality of consumers and producers despite substantial counterevidence, eventually paid off. The development of game theory, and especially evolutionary game theory, vindicated the economists' refusal to give up the assumption of rationality in spite of alleged falsifications.

What this history shows is that, at least when it comes to economics, Popper's claims seem to have been falsified as descriptions and to have been ill-advised as prescriptions. The history of Newtonian mechanics offers the same verdict on Popper's prescriptions. It is a history in which for long periods scientists were able to reduce narrower theories to broader theories, while improving the predictive precision of the narrower theories, or showing exactly where these narrower theories went wrong, and were only approximately correct. The history of Newtonian mechanics is also the history of data forcing us to choose between "ad hoc" adjustments to auxiliary hypotheses about initial conditions, and falsifying Newtonian mechanics, in which apparently the "right" choice was preserving the theory. Of course sometimes, indeed often, the right choice is to reject a theory as falsified, and frame a new hypothesis. The trouble is to decide in which situation scientists find themselves. Popper's one-size-fits-all recipe, "refute the current theory and conjecture new hypotheses", does not always provide the right answer.

The history of physics also seems to provide counterexamples to Popper's claim that science never seeks, nor should it seek, confirmatory evidence, positive instances, of a theory. In particular, scientists are impressed with "novel" predictions, cases in which a theory is employed to predict a hitherto completely undetected process or phenomenon, and even sometimes to predict its quantitative dimensions. Such experiments are treated not merely as attempts to falsify that fail, but as tests which positively confirm.

Recall the problems physicists and empiricists had with Newton's occult force, gravity. In the early twentieth century Albert Einstein advanced a "General Theory of Relativity" which provided an account of motion that dispensed with gravity. Einstein theorized that there is no such thing as gravity (some of his arguments were methodological, or philosophical). Instead, Einstein's theory holds, space is "curved", and more steeply curved around massive bodies like stars. One consequence of this theory is that the path of photons should be bent in the vicinity of such massive bodies. This is not something Newton's theory should lead us to expect since photons have no mass and so are not affected by gravity – recall the inverse square law of gravitational attraction in which the masses of bodies gravitationally attracting one another effect the force of gravity between them. In 1919 at great expense a British expedition was sent to a location in South America where a total solar eclipse was expected, in order to test Einstein's theory. By comparing the apparent location in the sky of stars the night before the eclipse and their apparent location during the eclipse (when stars are visible as a result of the Moon's blocking the Sun's normal brightness in the same region of the sky), the British team reported the confirmation of Einstein's hypothesis. The result of this test and others was of course to replace Newton's theory with Einstein's.

Many scientists treated the outcome of this expedition's experiment as strong confirmation of the General Theory of Relativity. Popper would of course have to insist that they were mistaken. At most, the test falsified Newton's theory, while leaving Einstein's unconfirmed. One reason many scientists would reject Popper's claim is that in the subsequent 80 years, as new and more accurate devices became available for measuring this and other predictions of Einstein's theory, its consequences for well-known phenomena were confirmed to more and more decimal places, and more important, its novel predictions about phenomena no one had ever noticed or even thought of, were confirmed. Still, Popper could argue that scientists are mistaken in holding the theory to be confirmed. After all, even if the theory does make more accurate predictions than Newton's, they don't match up 100 percent with the data, and excusing this discrepancy by blaming the difference on observational error or imperfections in the instruments, is just an *ad hoc* way of preserving the theory from falsification. One thing Popper could not argue is that the past fallibility of physics shows that probably Einstein's General Theory of Relativity is also at best an approximation and not completely true. Popper could not argue this way, for this is an inductive argument, and Popper agrees with Hume that such arguments are ungrounded.

What can Popper say about theories that are repeatedly tested, whose predictions are borne out to more and more decimal places, which make novel striking predictions that are in agreement with (we can't say "confirmed by") new data? Popper responded to this question by invoking a new concept: "corroboration". Theories can never be confirmed, but they can be corroborated by evidence. How does corroboration differ from confirmation?

It is a quantitative property of hypotheses which measures their content and testability, their simplicity and their previous track-record of success in standing up to attempts to falsify them in experiments. For present purposes the details of how corroboration differs from confirmation is not important, except that corroboration cannot be a relationship between a theory and already available data that either makes any prediction about future tests of the theory, or gives us any positive reason at all to believe that the theory is true or even closer to the truth than other theories. The reason is obvious. If corroboration had either of these properties, it would be at least in part a solution to the problem of induction, and this is something Popper began by dispensing with.

If hypotheses and theories are the sorts of things that people can believe to be true, then it must make sense to credit some of them with more credibility than others, as more reasonable to believe than others. It may well be that among the indefinitely many possible hypotheses, including all the ones that never have and never will occur to anyone, the theories we actually entertain are less well supported than others, are not even approximately true and are not improving in approximate truth over their predecessors. This possibility may be a reason to reject increasing confirmation as merely short-sighted speculation. But it is an attitude difficult for working scientists to take seriously. As between competing hypotheses they are actually acquainted with, the notion that none is more reasonable to believe than any other doesn't seem attractive. Of course, an instrumentalist about theories would not have this problem. On the instrumentalist view, theories are not to be believed or disbelieved, they are to be used when convenient, and otherwise not. Instrumentalists may help themselves to Popper's rejection of induction in favor of falsification. But, ironically, Popper was a realist about scientific theories.

5.4 Statistics and probability to the rescue?

At some point the problems of induction will lead some scientists to lose patience with the philosopher of science. Why not simply treat the puzzle of grue and bleen as a philosopher's invention, and get on with the serious but perhaps more soluble problem of defining the notion of empirical confirmation? We may grant the fallibility of science, the impossibility of establishing the truth or falsity of scientific laws once and for all, and the role which auxiliary hypotheses inevitably play in the testing of theories. Yet we may still explain how observation, data collection and experiment test scientific theory by turning to statistical theory and the notion of **probability**. The scientist who has lost patience with the heavy weather which philosophers make of how data confirm hypotheses will also insist that this is a problem for statistics, not philosophy. Instead of worrying about problems like what a positive instance of a hypothesis could be, or why positive instances confirm hypotheses we actually entertain and not an infinitude of alternative possibilities we haven't even dreamed up, we should leave the nature of hypothesis-testing to departments of probability and statistics. This is advice philosophers have resolutely tried to follow. As we shall see, it merely raises more problems about the way experience guides the growth of knowledge in science.

To begin with, there is the problem of whether the fact that some data raise the probability of a hypothesis makes that data positive evidence for it. This may sound like a question trivially easy to answer, but it isn't. Define **p(h, b)** as the probability of hypothesis **h**, given auxiliary hypotheses **b**, and **p(h, e** and **b)** as the probability of **h** given the auxiliary hypotheses, **b**, and some experimental observations **e**. Suppose we adopt the principle that

> **e** is positive evidence for hypothesis **h** if and only if **p(h, e** and **b)** > **p(h, b)**

So, in this case, **e** is "new" data that count as evidence for **h** if they raise the probability of **h** (given the auxiliary assumptions required to test **h**). For example, the probability that the butler did it, **h**, given that the gun found at the body was not his, **b**, and the new evidence **e** that the gun carried his fingerprints, is higher than the hypothesis that the butler did it, given the gun found at the body, and no evidence about fingerprints. It is the fingerprints that raise the probability of **h**. That's why the prints are "positive evidence".

It is easy to construct counterexamples to this definition of positive evidence which shows that increasing probability is by itself neither necessary nor sufficient for some statement about observations to confirm a hypothesis. Here are two:

This book's publication increases the probability that it will be turned into a blockbuster film starring Nicole Kidman. After all, were it never to have been published, the chances of its being made into a film would be even smaller than they are. But surely the actual publication of this book is not positive evidence for the hypothesis that this book will be turned into a blockbuster film starring Nicole Kidman. It is certainly not clear that some fact which just raises the probability of a hypothesis thereby constitutes positive evidence for it. A similar conclusion can be derived from the following counterexample, which invokes lotteries, a useful notion when exploring issues about probability. Consider a fair lottery with 1,000 tickets, 10 of which are purchased by Andy and 1 is purchased by Betty. **h** is the hypothesis that Betty wins the lottery. **e** is the observation that all tickets except those of Andy and Betty are destroyed before the drawing. **e** certainly increases the probability of **h** from 0.001 to 0.1. But it is not clear that **e** is positive evidence that **h** is true. In fact, it seems more reasonable to say that **e** is positive evidence that **h** is untrue, that Andy will win. For the probability that **he** wins has gone from 0.01 to 0.9. Another lottery case suggests that raising probability is not necessary for being positive evidence; indeed a piece of positive evidence may lower the probability of the hypothesis it confirms. Suppose in our lottery Andy has purchased 999 tickets out of 1,000 sold on Monday. Suppose **e** is the evidence that by Tuesday 1,001 tickets have been sold, of which Andy purchased 999. This **e** lowers the probability that Andy will win the lottery from 0.999 to 0.998 . . . But surely **e** is still evidence that Andy will win after all.

One way to deal with these two counterexamples is simply to require that **e** is positive evidence for **h** if **e** makes **h**'s probability high, say above 0.5. Then, in the first case, since the evidence doesn't raise the probability of Betty's winning anywhere near 0.5, and in the second case the evidence does not lower the probability of Andy's winning much below 0.999, these cases don't undermine the definition of positive evidence when so revised. But of course, it is easy to construct a counterexample to this new definition of positive evidence as evidence that makes the hypothesis highly probable. Here is a famous case: **h** is the hypothesis that Andy is not pregnant, while **e** is the statement that Andy eats Weetabix breakfast cereal. Since the probability of **h** is extremely high, **p(h, e)** – the probability of **h**, given **e**, is also extremely high. Yet **e** is certainly no evidence for **h**. Of course we have neglected the background information, **b**, built into the definition. Surely if we add the background information that no man has ever become pregnant, then **p(h, e and b)** – the probability of **h**, given **e** and **b** – will be the same as **p(h, e)**, and thus dispose of the counterexample. But if **b** is the statement that no man has ever become pregnant, and **e** is the statement that Andy ate Weetabix, and **h** is the statement that Andy is not pregnant, then **p(h, e and b)** will be very high, indeed about as close to 1 as a probability can get. So, even though **e** is not by itself positive evidence for **h**, **e** plus **b** is, just because **b** is positive evidence for **h**. We cannot exclude **e** as positive evidence, when **e** plus **b** is evidence, just because it is a conjunct which by itself has no impact on the probability of **h**, because sometimes positive evidence only does raise the probability of a hypothesis when it is combined with other data. Of course, we want to say that in this case, **e** could be eliminated without reducing the probability of **h**, **e** is probabilistically irrelevant and that's why it is not positive evidence. But providing a litmus test for probabilistic irrelevance is no easy task. It may be as difficult as defining positive instance. In any case, we have an introduction here to the difficulties of expounding the notion of evidence in terms of the concept of probability.

Philosophers of science who insist that probability theory and its interpretation suffice to enable us to understand how data test hypotheses will respond to these problems that they reflect the mis-fit between probability and our common-sense notions of evidence. Our ordinary concepts are qualitative, imprecise, and not the result of a careful study of their implications. Probability is a quantitative mathematical notion with secure logical foundations that enables us to make distinctions ordinary notions cannot draw, and to explain these distinctions. Recall the logical empiricists who sought rational reconstructions or explications of concepts like explanation that provide necessary and sufficient conditions in place of the imprecision and vagueness of ordinary language. Likewise, many contemporary students of the problem of confirmation seek a more precise substitute for the ordinary notion of evidence in the quantifiable notion of probability; for them counterexamples such as the ones adduced above simply reflect the fact that the two concepts are not identical. They are no reason not to substitute "probability" for "evidence" in our inquiry about how data test theory. Some of these philosophers go further and argue that there is no such thing as evidence confirming or disconfirming a hypothesis by itself. Hypothesis testing in science is always a comparative affair: it only makes sense to say hypothesis $\mathbf{h}_1$ is more or less well confirmed by the evidence than is hypothesis $\mathbf{h}_2$, not that $\mathbf{h}_1$ is confirmed by **e** in any absolute sense.

These philosophers hold that the mathematical theory of probability holds the key to understanding the confirmation of scientific theory. And this theory is extremely simple. It embodies only three very obvious assumptions:

1 Probabilities are measured in numbers from 0 to 1.
2 The probability of a necessary truth (like "4 is an even number") is 1.
3 If hypotheses **h** and **j** are incompatible, then **p(h** or **j) = p(h) + p(j)**.

It's easy to illustrate these axioms with a deck of normal playing cards. The probability of any one card being drawn from a complete deck is between 0 and 1. In fact it's 1/52. The probability that a card will be red or black (the only two possibilities) is 1 (it's a certainty), and if drawing an ace of hearts is incompatible with drawing a jack of spades, then the probability of drawing one of them is 1/52 + 1/52, or 1/26, about 0.038461 . . .

From these simple and straightforward assumptions (plus some definitions) the rest of the mathematical theory of probability can be derived by logical deduction alone. In particular, from these three axioms of the theory of probability, we can derive a theorem, first proved by a British theologian and amateur mathematician in the eighteenth century, Thomas Bayes, which has bulked large in contemporary discussions of confirmation. Before introducing this theorem, we need to define one more notion, the conditional probability of any one statement, assuming the truth of another statement. The conditional probability of a hypothesis, **h**, on a description of data, **e**, written **p(h/e)**, is defined as the ratio of the probability of the truth of both **h** and **e** to the probability of the truth of **e** alone:

$$\mathbf{p(h/e)} = \frac{\mathbf{df\ p(h\ \text{and}\ e)}}{\mathbf{p(e)}}$$

Roughly, "the conditional probability of **h** on **e**" measures the proportion of the probability that **e** is true, which "contains" the probability that **h** is also true. Adapting an expository idea of Martin Curd and Jan Cover, we can illuminate this definition with a few diagrams. Suppose we are shooting darts at a board on which two overlapping circles are drawn in the shape of a Venn diagram:

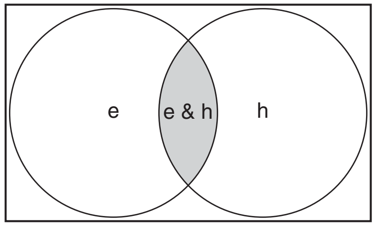

Figure 1 Circles **e** and **h** are the same size, and between them cover most of the rectangle, to suggest that the probability of a dart hitting one of them (and not the other) is large and about the same.

If a dart lands inside circle **e**, what is the probability that it will also land inside circle **h**, i.e. the probability of landing in **h**, on the condition that it lands in **e**, the conditional probability, **p**(**h**/**e**)? That depends on two things: the area of overlap between circle **e** and circle **h** (the intersection **e** and **h**), relative to the area of **e**, and the size of **e** compared to the size of **h**. To see this, compare the two following diagrams. In this one, **e** is very large compared to the size of **h**, so the chance that a dart thrown inside **e** also lands in **h** is low. But it would be higher if more of **h** were inside **e**. On the other hand, the chance that a dart which lands in **h** also lands in **e** is much higher, and increases as the proportion of **h** inside **e** grows.

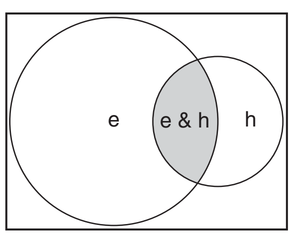

Figure 2 Circle **e** is much larger than circle **h**, so the probability of the dart hitting **e** is much higher than the probability of the dart hitting **h**. The shaded intersection **e** & **h** is much smaller than **e**, and a relatively large proportion of **h**. Thus **p**(**h**/**e**) is low, and **p**(**e**/**h**) is much higher than **p**(**h**/**e**).

By contrast, consider the following diagram. Here **e** is small and **h** is large. In this case the chance of a dart which lands in **e** also landing in **h** is higher than in the previous case, and becomes even higher the more of **e** is inside **h**. Again, the conditional probability of **e** on **h** is of course much lower the smaller the **h** circle is and the less it overlaps **e**.

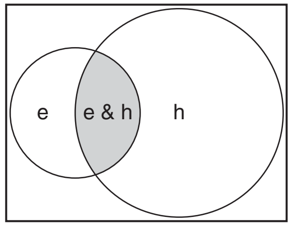

Figure 3 Circle **h** is much larger than circle **e**, so the probability of the dart hitting **h** is much higher than the probability of the dart hitting **e**. The shaded intersection **e & h** is much smaller than **h**, and is a relatively large proportion of **e**. Thus **p(h/e)** is high, and **p(e/h)** is much lower than **p(h/e)**.

The definition of conditional probability incorporates these two factors on which conditional probability depends. The numerator reflects the size of the overlap of **e** and **h** relative to the sizes of **e** and **h**, and the denominator measures that size in units of **e**'s size.

Now if **h** is a hypothesis and **e** is a report of data, Bayes' theorem allows us to calculate the conditional probability of **h** on **e**, **p(h/e)**. In other words, Bayes' theorem gives us a mathematical formula for calculating how much more or less probable a bit of evidence, **e**, makes any hypothesis, **h**. The formula is as follows:

$$\text{Bayes' theorem: } \mathbf{p(h/e)} = \frac{\mathbf{p(e/h)} \times \mathbf{p(h)}}{\mathbf{p(e)}}$$

Bayes' theorem tells us that once we acquire some data, **e**, we can calculate how the data **e** change the probability of **h**, raising or lowering it, provided we already have three other numbers:

p(e/h) – the probability that **e** is true assuming that **h** is true (as noted above, not to be confused with **p(h/e)**, the probability that **h** is true, given **e**, which is what we are calculating). This number reflects the degree to which our hypothesis leads us to expect the data we have gathered. If the data are just what the hypothesis predicts, then of course **p(e/h)** is very high. If the data are nothing like what the hypothesis predicts, **p(e/h)** is low.

p(h) – the probability of the hypothesis independent of the test which the data described by **e** provides. If **e** reports new experimental data, then **p(h)** is just the probability the scientist assigned to **h** before the experiment was conducted.

p(e) – the probability that the statement describing the data is true

> independent of whether **h** is true or not. Where **e** is a surprising result which previous scientific theory and evidence (independent of **h**) does not lead us to expect, **p(e)** will be low.

To see how easily Bayes' theorem follows from the axioms of probability and our definition of conditional probability, return to any of the dartboard diagrams above. If we can calculate **p(e/h)** by comparing the relative sizes of the circles and the ratio of their intersections to their sizes, we can also calculate **p(h/e)** the same way. Of course the figures for each conditional probability will be different (as each of the diagrams illustrates).

By drawing **e**- and **h**-circles and intersections of them of different sizes, it is easy to see that the probability of a dart which hits the **e**-circle also hitting the **h**-circle, **p(h/e)**, will vary directly as the ratio of the intersection of the two circles to the size of the **e**-circle, and inversely as the ratio of the sizes of the **e**-circle to the size of the **h**-circle. And this is exactly what Bayes' theorem says: it makes **p(h/e)** equal to **p(e/h)** – the ratio of the intersection of **e** and **h** to the size of **e** – times the fraction **p(h)/p(e)**, which is the ratio of the size of **h** to the size of **e**.

Two simple examples may help us see how Bayes' theorem is supposed to work: Consider how data on the observed position of Halley's comet provide a test for Newton's laws. Suppose, given prior observations, that **p(e)**, the probability that Halley's comet will be observed in a particular location of the night sky, is 0.8. This allows for imperfections in the telescope, atmospheric irregularities, all the factors that eventually led astronomers to take many photographs of the stars and planets and to average their positions to make estimates of their expected positions in the sky. **p(e/h)** is also high, the expected position of Halley's comet in the night sky is very close to what the theory predicts it would be. Let's set **p(e/h)** at 0.95. Let's assume that prior to the acquisition of **e**, the new data about Halley's comet, the probability that Newton's laws are true is, say, 0.8. Thus, if Halley's comet appears where expected, **p(h/e)** = (0.95 . . .) × (0.8)/(0.8) = 0.95. Thus, the evidence as described by **e** has raised the probability of Newton's laws from 0.8 to 0.95.

But now, suppose we acquire new data about, say, the precession of the perihelion of Mercury – that is, data which shows that the elliptical orbit of Mercury around the Sun is itself swinging so that the closest point between Mercury and the Sun keeps shifting. Suppose, as was indeed the case, that the figure turns out to be much higher than Newton's laws (and the auxiliary hypotheses used to apply them) would lead us to expect, so that **p(e/h)** is low, say, 0.3. Since Newton's laws did not lead us to expect this data, the **prior probability** of **e** must be low, so let's let **p(e)** be low, say, 0.2; and the prior probability of such unexpected data, given Newton's laws plus auxiliary hypotheses, will also be quite low, say, **p(e/h)** is 0.1. If **p(h)** for Newton's laws plus auxiliaries is 0.95, then Bayes' theorem tells us that for the new **e**, the precession data for Mercury, the **p(h/e)** = (0.1) × (0.95)/(0.2) = 0.475, a significant drop from 0.95. Naturally, recalling the earlier success of Newton's laws in uncovering the existence of Neptune and Uranus, the initial blame for the drop was placed on the auxiliary hypotheses. Bayes' theorem can even show us why. Though the numbers in our example are made up, in this case, the auxiliary assumptions were eventually vindicated, and the data about the much greater than expected precession of the perihelion of Mercury undermined Newton's theory, and (as another application of Bayes' theorem would show), increased the probability of Einstein's alternative theory of relativity.

Philosophers and many statisticians hold that the reasoning scientists use to test their hypotheses can be reconstructed as inferences in accordance with Bayes' theorem. These theorists are called Bayesians. Some philosophers and historians of science among them seek to show that the history of acceptance and rejection of theories in science honors Bayes' theorem, thus showing that in fact, theory testing has been on firm footing all along. Other philosophers, and statistical theorists, attempt to apply Bayes' theorem actually to determine the probability of scientific hypotheses when the data are hard to get, sometimes unreliable, or only indirectly relevant to the hypothesis under test. For example, they seek to determine the probabilities of various hypotheses about evolutionary events like the splitting of ancestral species from one another, by applying Bayes' theorem to data about differences in the polynucleotide sequences of the genes of currently living species.

How much understanding of the nature of empirical testing does **Bayesianism** really provide? Will it reconcile science's empiricist epistemology with its commitment to unobservable events and processes that explain observable ones? Will it solve Hume's problem of induction? To answer these questions, we must first understand what the probabilities are that all these **p**s symbolize and where they come from. We need to make sense of **p(h)**, the probability that a certain proposition is true. There are at least two questions to be answered: First, there is the "metaphysical" question of what fact is it about the world, if any, that makes a particular probability value, **p(h)**, for a hypothesis, **h**, the true or correct one? Second, there is the epistemological question of justifying our estimate of this probability value. The first question may also be understood as a question about the meaning of probability statements, and the second about how to justify inductive conclusions about general theories and future eventualities.

Long before the advent of Bayesianism in the philosophy of science, the meaning of probability statements was already a vexed question. There are some traditional interpretations of probability we can exclude as unsuitable interpretations for the employment of Bayes' theorem. One such is the interpretation of probability as it is supposed to figure in fair games of chance like roulette or black jack. In a fair game of roulette the chance of the ball landing in any trap is exactly 1/37 or 1/38 because there are 37 (or in Europe 38) traps into which the ball can land. Assuming it is a fair roulette wheel, the probability of the hypothesis that the ball will land on number 8 is exactly 1/37 or 1/38 and we know this *a priori* – without experience, because we know *a priori* how many possibilities there are and that each is equally probable (again, assuming the roulette wheel is fair, a bit of knowledge we could never have acquired *a priori* any way!). Now, when it comes to hypotheses that can account for a finite body of data, there is no limit to the number of possibilities and no reason to think that each of them has the same probability. Accordingly, the probabilities of a hypothesis about, say, the number of chromosomes in a human nucleus, will not be determinable *a priori*, by counting up possibilities and dividing 1 by the number of possibilities.

Another interpretation of probabilities involves empirical observations, for example, coin flips. To establish the frequency with which a coin will come up heads, one flips it several times and divides the number of times it comes up heads by the number of times it was flipped. When will this frequency be a good estimate of the probability of heads? When the number of coin flips is large, and the frequencies we calculate for finite numbers of coin flips converge on one value and remain near that value no matter how many times we continue flipping. We can call this value, if there is one, the **long-run relative frequency** of heads. And we treat it as a measure of the probability the coin comes up heads. But is the long-run relative frequency of heads identical to the probability it will come up heads? This sounds like a silly question, until you ask what the connection is between the long-run relative frequency's being, say 50 percent and the chance that the very next toss will be heads. Notice that a long-run relative frequency of 50 percent is compatible with a run of ten, or a hundred, or a million heads in a row, just so long as the total number of tosses is very large, so large that a million is a small number in comparison to the total number of tosses. If this is right, the long-run relative frequency is compatible with any finite run of all heads, or all tails, and of course perfectly compatible with the coin's coming up tails on the next toss. Now, suppose we want to know what the probability is that the coin will come up heads on the next toss. If the probability that the coin will come up heads on the next toss is a property of that particular toss, it is a different thing from the long-run relative frequency of heads (which is perfectly compatible with the next 234,382 tosses all being tails). We need some principle that connects the long-run to the next toss. One such principle which gets us from the long-run relative frequency to the probability of the next toss being heads is to assume that coins do in any finite run what they do in the long run. But this principle is just false. A better principle for connecting long-run relative frequencies to the probability of the next occurrence is something like this: If you know the long-run relative frequency, then you know how to bet on whether the coin will land heads or tails, and if you take all bets against heads at odds greater than even money, you will win. But notice this is a conclusion about what you should do as a gambler, not a conclusion about what the coin will in fact do. We will come back to this insight.

Could long-run relative frequencies provide the probability values for a hypothesis without a track record? It is hard to see how. Compare a novel hypothesis to a shiny new penny about to be flipped. Long-run relative frequencies data provide some reason to ascribe a probability of 50 percent to the chances of heads on the new penny. Is there a track record of previous hypotheses relevant to the new one? Only if we can compare it to the right class of similar hypotheses the way we can compare new pennies to old ones. But hypotheses are not like pennies. Unlike pennies, they differ from one another in ways we cannot quantify as we would have to were we to grade them for similarity to one another. Even if we could identify the track record of truth and falsity for similar hypotheses formulated over the past history of science, we would have the problems of (a) justifying the inference from a finite actual sequence to a long-run relative frequency; and (b) justifying the inference from a long-run relative frequency to the next case, the new hypothesis. Recall that in the case of coin flipping, the only connection appears to be that relative frequencies are our best guide to how to lay our bets about the next toss. Perhaps the kind of probability which theory testing invokes is the gambler's kind, what has come to be called "subjective probability". "Subjective" because it reflects facts about the gambler, and what the gambler believes about the past and the future, and "probability" because the bets the gambler makes should honor the axioms of probability.

It is the claim that in scientific testing, the relevant probabilities are subjective probabilities, gambler's odds, that is the distinctive mark of the Bayesian. A Bayesian is someone who holds that at least two of the three probabilities we need to calculate **p(h/e)** are just a matter of betting odds and that within certain weak constraints they can take on any values at all. You and I may think that the best betting odds are those which mirror our previous experience of actual frequencies or our estimate of long-run relative frequencies, but this is no part of Bayesianism. The Bayesian holds that in the long run it doesn't matter what values they start with, Bayes' theorem will lead the scientist inexorably to the (available) hypothesis best supported by the evidence. These remarkable claims demand explanation and justification.

Calculating the value of **p(e/h)** is a matter of giving a number to the probability that **e** obtains if **h** is true. This is usually easy to do. If **h** tells us to expect **e**, or data close to **e**, then **p(e/h)** will be very high. The problem is that using Bayes' theorem also requires we calculate input values, so-called "prior probabilities", **p(h)** and **p(e)**. **p(h)** is especially problematical: after all, if **h** is a new theory no one has ever thought of, why should there be any particular right answer to the question of with what probability it is true? And assigning a value to **p(e)**, the probability that our data description is correct may involve so many auxiliary assumptions, that even if there is a correct number, it is hard to see how we could figure out what it is. The Bayesian asserts that these are not problems. Both values, **p(h)** and **p(e)** (and **p(e/h)** for that matter), are simply degrees of belief, and degrees of belief are simply a matter of what betting odds the scientist would take or decline on whether their beliefs are correct. The higher the odds one takes, the stronger the degree of belief. Here the Bayesian takes a page from economists and others who developed the theory of rational choice under uncertainty. The way to measure a degree of belief is to offer the believer wagers against the truth of his or her belief. Other things being equal, if you are rational, and you are willing to take a bet that **h** is true at odds of 4:1, then your degree of belief that **h** is true is 0.8. If you are willing to take a 5:1, then your degree of belief is 0.9. Probabilities are identical to degrees of belief. The other things that have to be equal for this way of measuring the strength of your beliefs to work are (a) that you have enough money so that you are not so averse to the risk of losing that it swamps your attraction to the prospect of winning; (b) that the degrees of belief you assign to your beliefs obey the rules of logic and the three laws of probability above. So long as your degrees of belief, aka probability assignments, honor these two assumptions, the Bayesian says, the initial values or "prior probabilities" you assign to them can be perfectly arbitrary, in fact may be arbitrary, but it doesn't really matter. In the parlance of the Bayesians, as more and more data come in, the prior probabilities will be "swamped", that is, when we use Bayes' theorem to "update" prior probabilities, i.e. feed new **p(e)**s into the latest values for **p(e/h)** and **p(h/e)**, the successive values of **p(h/e)** will converge on the correct value, no matter what initial values for these three variables we start with! Prior probabilities are nothing but measures of the individual scientists' purely subjective degrees of belief before applying Bayes' theorem. In answer to our metaphysical question about what facts about the world probabilities report, prior probabilities report no facts about the world, or at least none about the world independent of our beliefs. In answer to the epistemological question of what justifies our estimates of probabilities, when it comes to prior probabilities, no more justification is needed or possible than that our estimates obey the axioms of probability.

There is no right answer or wrong answer as to what the prior probabilities of **p(h)** or **p(e)** are, so long as the values of these probabilities obey the rules of probability and logical consistency on betting. Logical consistency simply means that one places one's bets – that is, assigns strengths to one's degrees of belief in such a way that bookies can't use you for a money pump: that is, make bets with you so that no matter which propositions come out true or false you lose money. What is more, another theorem of the probability theory shows that if we apply Bayes' theorem relentlessly to "update" our prior probabilities as new evidence comes in, the value of **p(h)** all scientists assign will converge on a single value no matter where each scientist begins in his or her original assignment of prior probabilities. So not only are prior probabilities arbitrary but it doesn't matter that they are! Some scientists may assign prior probabilities on considerations like simplicity or economy of assumptions, or similarity to already proven hypotheses, or symmetry of the equations expressing the hypothesis. Other scientists will assign prior probabilities on the basis of superstition, aesthetic preference, number worship, or by pulling a ticket out of a hat. It doesn't matter, so long as they all conditionalize on new evidence via Bayes' theorem.

It is not much of an objection to this account of scientific testing that scientists actually offer good reasons for their methods of assigning prior probabilities. To begin with, Bayesianism doesn't condemn these reasons, at worst, it is silent on them. But if features like the simplicity of a hypothesis or the symmetry of its form do in fact increase its prior probability, this will be because a hypothesis having features like this will, via Bayes' theorem, acquire a higher posterior probability than other hypotheses with which it is competing that lack these features. More important, attempts to underwrite the reasoning of scientists who appeal to considerations like economy, simplicity, symmetry, invariance, or other formal features of hypotheses, by appeal to the claim that such features increase the objective probability of a hypothesis, come up against the problem that the only kind of probability that seems to make any sense for scientific testing is Bayesian subjective probability.

Furthermore, so understood, some Bayesians hold that probabilities can after all deal with some of the traditional problems of confirmation. Recall the black boot/white swan positive instance-puzzle discussed above, according to which a black boot is positive evidence for "All swans are white". Not on Bayesianism. After all, the prior conditional probability of a boot being black, conditional on all swans being white, is lower than the prior probability of the next swan we see being white, conditional on all swans being white. When we plug these two priors into Bayes' theorem, if the prior probabilities of seeing a white swan and a black boot are equal, the probability of "All swans are white" is raised much more by the latter conditional probability.

One of the major problems confronting Bayesianism, and perhaps other accounts of how evidence confirms theory, is the "problem of old evidence". It is not uncommon in science for a theory to be strongly confirmed by data already well known long before the hypothesis was formulated. Indeed, as we saw, this is an important feature of situations in which scientific revolutions take place: Newton's theory was strongly confirmed by its ability to explain the data on which Galileo's and Kepler's theories were based. Einstein's general theory of relativity explained previously recognized but highly unexpected data such as the invariance of the speed of light and the precession of the perihelion of Mercury. In these two cases $\mathbf{p(e)} = 1$, $\mathbf{p(e/h)}$ is very high. Plugging these values into Bayes' theorem gives us

$$\mathbf{p(h/e)} = \frac{1 \times \mathbf{p(h)}}{1} = \mathbf{p(h)}$$

In other words, on Bayes' theorem the old evidence does not raise the posterior probability of the hypothesis – in this case Newton's laws, or the special theory of relativity – at all. Bayesians have gone to great lengths to deal with this problem. One stratagem is to "bite the bullet" and argue that old evidence does not in fact confirm a new hypothesis. This approach makes common cause with the well-established objection to hypotheses which are designed with an eye to available evidence. Scientists who construct hypotheses by intentional "curve fitting" are rightly criticized and their hypotheses are often denied explanatory power on the grounds that they are *ad hoc*. The trouble with this strategy is that it doesn't so much solve the original Bayesian problem of old evidence as combine it with another problem: how to distinguish cases like the confirmation of Newton's and Einstein's theories by old evidence from cases in which old evidence does not confirm a hypothesis because it was accommodated to the old evidence. The alternative approach to the problem of old evidence is to supplement Bayes' theorem with some rule that gives **p(e)** a value different from 1. For example, one might try to give **p(e)** the value it might have had before **e** was actually observed in the past, or else try to rearrange one's present scientific beliefs by deleting **e** from them and anything which **e** makes probable; then go back and assign a value to **p(e)**, which presumably will be lower than 1. This strategy is obviously an extremely difficult one to actually adopt. And it is (subjectively) improbable that any scientist consciously thinks this way.

Many philosophers and scientists who oppose Bayesianism do so not because of the difficulties which are faced by the program of developing it as an account of the actual character of scientific testing. Their problem is with the approach's commitment to subjectivism. The Bayesian claim that no matter what prior probabilities the scientist subjectively assigns to hypotheses, their subjective probabilities will converge on a single value, is not sufficient consolation to opponents. Just for starters, values of **p(h)** will not converge unless we start with a complete set of hypotheses that are exhaustive and exclusive competitors. This seems never to be the case in science. Moreover, objectors argue, there is no reason given that the value on which all scientists will converge by Bayesian conditionalization is the *right* value for **p(h)**. This objection of course assumes there is such a thing as the right, i.e. the objectively correct, probability, and so begs the question against the Bayesian. But it does show that Bayesianism is no solution to Hume's problem of induction, as a few philosophers hoped it might be.

And the same pretty much goes for other interpretations of probability. If sequences of events reveal long-run relative frequencies that converge on some probability value and stay near it forever, then we could rely on them at least for betting odds. But to say that long-run relative frequencies will converge on some value is simply to assert that nature is uniform, that the future will be like the past and so begs Hume's question. Similarly, hypothesizing probabilistic propensities that operate uniformly across time and space also begs the question against Hume's argument. In general, probabilities are useful only if induction is justified, not vice versa. Still, as noted, only a handful of philosophers have sought explicitly to solve Hume's problem by appeal to probabilities.

There is a more severe problem facing Bayesianism. It is the same problem that we came up against in the discussion of how to reconcile empiricism and explanation in theoretical science. Because empiricism is the doctrine that knowledge is justified by observation, in general, it must attach the highest probability to statements which describe observations, and lower probability to those which make claims about theoretical entities. Since theories explain observations, we may express the relation between theory and observation as (**t** and **t**→**h**), where **t** is the theory and **t**→**h** reflects the explanatory relation between the theoretical claims of the theory, **t**, and an observational generalization, **h**, describing the data that the theory leads us to expect. The relation between **t** and **h** may be logically deductive, or it may be some more complex relation. But **p(h)** must never be lower than **p(t** and **t**→**h)**, just because the antecedent of the latter is a statement about what cannot be observed whose only consequence for observation is **h**. Bayesian conditionalization on evidence will never lead us to prefer (**t** and **t**→**h**) to **h** alone. But this is to say that Bayesianism cannot account for why scientists embrace theories at all, instead of just according high subjective probability to the observational generalizations that follow from them. Of course, if the explanatory power of a theory were a reason for according it a high prior probability, then scientists' embracing theories would be rational from the Bayesian point of view. But to accord explanatory power such a role in strengthening the degree of belief requires an account of explanation. And not just any account. It cannot, for example, make do with the D-N model, for the principal virtue of this account of explanation is that it shows that the *explanandum* phenomenon could be expected with at least high probability. In other words, it grounds explanatory power on strengthening probability, and so cannot serve as an alternative to probability as a source of confidence in our theories. To argue, as seems tempting, that our theories are explanatory in large part because they go beyond and beneath observations to their underlying mechanisms is something the Bayesian cannot do.

5.5 Underdetermination

The testing of claims about unobservable things, states, events or processes is evidently a complicated affair. In fact the more one considers how observations confirm hypotheses and how complicated the matter is, the more one is struck by a certain inevitable and quite disturbingly "underdetermination" of theory by observation.

As we have noted repeatedly, the "official epistemology" of modern science is empiricism – the doctrine that our knowledge is justified by experience – observation, data collection, experiment. The objectivity of science is held to rest on the role which experience plays in choosing between hypotheses. But if the simplest hypothesis comes face to face with experience only in combination with other hypotheses, then a negative test may be the fault of one of the accompanying assumptions, a positive test may reflect compensating mistakes in two or more of the hypotheses involved in the test that cancel one another out. Moreover, if two or more hypotheses are always required in any scientific test, then when a test-prediction is falsified there will always be two or more ways to "correct" the hypotheses under test. When the hypothesis under test is not a single statement like "All swans are white" but a system of highly theoretical claims like the kinetic theory of gases, it is open to the theorist to make one or more of a large number of changes in the theory in light of a falsifying test, any one of which will reconcile the theory with the data. But the large number of changes possible introduces a degree of arbitrariness foreign to our picture of science. Start with a hypothesis constituting a theory that describes the behavior of unobservable entities and their properties. Such a hypothesis can be reconciled with falsifying experience by making changes in it that cannot themselves be tested except through the same process all over again – one which allows for a large number of further changes in case of falsification. It thus becomes impossible to establish the correctness or even the reasonableness of one change over another. Two scientists beginning with the same theory, subjecting it to the same initial disconfirming test, and repeatedly "improving" their theories in the light of the same set of further tests will almost certainly end up with completely different theories, both equally consistent with the data their tests have generated.

Imagine, now, the "end of inquiry" when all the data on every subject is in. Can there still be two distinct equally simple, elegant, and otherwise satisfying theories equally compatible with all the data, and incompatible with one another? Given the empirical slack present even when all the evidence appears to be in, the answer seems to be that such a possibility cannot be ruled out. Since they are distinct theories, our two total "systems of the world" must be incompatible, and therefore cannot both be true. We cannot remain agnostic about whether one is right nor ecumenical about embracing both. Yet it appears that observation would not be able to decide between these theories.

In short, theory is underdetermined by observation. And yet science does not show the sort of proliferation of theory and the kind of unresolvable theoretical disputes that the possibility of this underdetermination might lead us to expect. But the more we consider reasons why this sort of underdetermination does not manifest itself, the more problematic becomes the notion that scientific theory is justified by objective methods that make experience the final court of appeal in the certification of knowledge. For what else besides the test of observation and experiment could account for the theoretical consensus characteristic of most natural sciences? Of course, there are disagreements among theorists, sometimes very great ones, and yet over time these disagreements are settled to almost universal satisfaction. If, owing to the ever-present possibility of underdetermination, this theoretical consensus is not achieved through the "official" methods, how is it achieved?

Well, besides the test of observation, theories are also judged on other criteria: simplicity, economy, consistency with other already adopted theories. But these criteria simply invoke observations, albeit somewhat indirectly. A theory's consistency with other already well-established theories confirms that theory only because observations have established the theories it is judged consistent with. Simplicity and economy in theories are themselves properties that we have observed nature to reflect and other well-confirmed theories to bear, and we are prepared to surrender them if and when they come into conflict with our observations and experiments. One alternative source of consensus philosophers of science are disinclined to accept is the notion that theoretical developments are epistemically guided by non-experimental, non-observational considerations, such as *a priori* philosophical commitments, religious doctrines, political ideologies, aesthetic tastes, psychological dispositions, social forces or intellectual fashions. Such factors we know will make for consensus, but not necessarily one that reflects increasing approximation to the truth, or to objective knowledge. Indeed, these non-epistemic, non-scientific forces and factors are supposed to deform understanding and lead away from truth and knowledge.

The fact remains that a steady commitment to empiricism coupled with a fair degree of consensus about the indispensability of scientific theorizing strongly suggests the possibility of a great deal of slack between theory and observation. But the apparent absence of arbitrariness fostered by underdetermination demands explanation. And if we are to retain our commitment to science's status as knowledge *par excellence*, this explanation had better be one we can parlay into a justification of science's objectivity as well. The next chapter shows that prospects for such an outcome are clouded with doubt.

Summary

Empiricism is the epistemology which has tried to make sense of the role of observation in the certification of scientific knowledge. Since the eighteenth century, if not before, especially British philosophers like Hobbes, Locke, Berkeley and Hume have found inspiration in science's successes for their philosophies, and sought philosophical arguments to ground science's claims. In so doing, these philosophers and their successors set the agenda of the philosophy of science and revealed how complex the apparently simple and straightforward relation between theory and evidence is.

In the twentieth century the successors of the British empiricists, the "logical positivists" or logical empiricists as some of them preferred, sought to combine the empiricist epistemology of their predecessor with advances in logic, probability theory, and statistical inference, to complete the project initiated by Locke, Berkeley and Hume. What they found was that some of the problems seventeenth- and eighteenth-century empiricism uncovered were even more resistant to solution when formulated in updated logical and methodological terms. "Confirmation theory", as this part of the philosophy of science came to be called, has greatly increased our understanding of the "logic" of confirmation, but has left as yet unsolved Hume's problem of induction, the further problem of when evidence provides a positive instance of a hypothesis, and the "new riddle of induction" – Goodman's puzzle of "grue" and "bleen".

Positivists and their successors have made the foundations of probability theory central to their conception of scientific testing. Obviously much formal hypothesis testing employs probability theory. One attractive late twentieth-century account that reflects this practice is known as Bayesianism: The view holds that scientific reasoning from evidence to theory proceeds in accordance with Bayes' theorem about conditional probabilities, under a distinctive interpretation of the probabilities it employs.

The Bayesians hold that scientists' probabilities are subjective degrees of belief, betting-odds. By contrast with other interpretations, according to which probabilities are long-run relative frequencies, or distributions of actualities among all logical possibilities, this frankly psychological interpretation of probability is said to best fit the facts of scientific practice and its history.

The Bayesian responds to complaints about the subjective and arbitrary nature of the probability assignment it tolerates by arguing that, no matter where initial probability estimates start out, in the long-run using Bayes' theorem on all possible alternative hypotheses will result in their convergence on the most reasonable probability values, if there are such values. Bayesianism's opponents demand it substantiate the existence of such "most reasonable" values and show that all alternative hypotheses are being considered. To satisfy these demands would be tantamount to solving Hume's problem of induction. Finally, Bayesianism has no clear answer to the problem which drew our attention to hypothesis testing: the apparent tension between science's need for theory and its reliance on observation.

This tension expresses itself most pointedly in the problem of underdetermination. Given the role of auxiliary hypotheses in any test of a theory, it follows that no single scientific claim meets experience for test by itself. It does so only in the company of other, perhaps large numbers of, other hypotheses' need to effect the derivation of some observational prediction to be checked against experience. But this means that a disconfirmation test, in which expectations are not fulfilled, cannot point the finger of falsity at one of these hypotheses and that adjustments in more than one may be equivalent in reconciling the whole package of hypotheses to observation. As the size of a theory grows, and it encompasses more and more disparate phenomena, the alternative adjustments possible to preserve or improve it in the face of recalcitrant data increase. Might it be possible, at the never-actually-to-be-reached "end of inquiry", when all the data are in, that there be two distinct total theories of the world, both equal in evidential support, simplicity, economy, symmetry, elegance, mathematical expression or any other desideratum of theory choice? A positive answer to the question may provide powerful support for an instrumentalist account of theories. For apparently there will be no fact of the matter accessible to inquiry that can choose between the two theories.

And yet, the odd thing is that underdetermination is a mere possibility. In point of fact, it almost never occurs. This suggests two alternatives. The first alternative, embraced by most philosophers of science, is that observation really does govern theory choice (else there would be more competition among theories and models than there is); it's just that we simply haven't figured it all out yet. The second alternative is more radical, and is favored by a generation of historians, sociologists of science and a few philosophers who reject both the detailed teachings of logical empiricism, and also its ambitions to underwrite the objectivity of science. On this alternative, observations underdetermine theory, but it is fixed by other facts – non-epistemic ones, like bias, faith, prejudice, the desire for fame or at least security, or power-politics. This radical view, that science is a process, like other social processes, and not a matter of objective progress, is the subject of the next two chapters.

Study questions

1. Discuss critically: "Lots of scientists pursue science successfully without any regard to epistemology. The idea that science has an 'official one', and that empiricism is it, is wrong-headed."
2. Why would it be correct to call Locke the father of modern scientific realism and Berkeley the originator of instrumentalism? How would Berkeley respond to the argument for realism as an inference to the best explanation of science's success?
3. We have defined grue and bleen by way of the concepts of green and blue. Construct a definition of green and blue which starts with grue and bleen. What does this show about the projectability of green and blue?
4. What advantages do riskier hypotheses have over less risky ones in science?
5. Give examples, preferably from science, in which all three concepts of probability are used: subjective, relative frequency and probabilistic propensity. Hint: think of weather reports.
6. Argue against the claim that two equally well-confirmed total theories which appear to be incompatible are only disguised terminological variants of one another.

Suggested reading

The relationship between science and philosophy, and especially the role of science in the dispute between empiricism and rationalism during that period are treated in E.A. Burtt, *Metaphysical Foundations of Modern Science*. John Locke's *Essay on Human Understanding* is a long work, George Berkeley's *Principles of Human Knowledge* is brief but powerful. The last third develops an explicitly instrumental conception of science which he contrasts to Locke's realism. Berkeley argued for idealism – the thesis that only what is perceived exists, that the only thing we perceive is ideas, that therefore only ideas exist. His argument turns on the very same theory of language which the logical empiricists initially embraced: the meaning of every term is given by the sensory idea it names. About this work, Hume wrote, "it admits no refutation, and carried no conviction", in his *Inquiry Concerning Human Understanding*. In this work he develops the theory of causation discussed in Chapter 2, the theory of language common to empiricists from Berkeley to the logical positivists, and the problem of induction. Bertrand Russell's famous paper, "On Induction", reprinted in Balashov and Rosenberg, *Philosophy of Science: Contemporary Readings*, brought Hume's argument to central stage in twentieth-century analytical philosophy.

J.S. Mill, *A System of Logic*, carried the empiricist tradition forward in the nineteenth century, and proposed a canon for experimental science still widely employed under the name, Mill's methods of induction. The physicist Ernst Mach, *The Analysis of Sensation*, embraced Berkeley's attack on theory as empirically unfounded against Ludwig Boltzman's atomic theory. This work was greatly influential on Einstein. In the first half of the twentieth century logical empiricists developed a series of important theories of confirmation, R. Carnap, *The Continuum of Inductive Methods*, H. Reichenbach, *Experience and Prediction*. Their younger colleagues and students wrestled with these theories and their problems. Essays on confirmation theory in Hempel, *Aspects of Scientific Explanation* are of special importance, as is N. Goodman, *Fact, Fiction and Forecast*, where the new riddle of induction is introduced along with Goodman's path-breaking treatment of counterfactuals. Peter Achinstein's paper, "The Grue Paradox", which appears in print initially in Balashov and Rosenberg, is an invaluable exposition of Goodman's new riddle, and a novel solution.

W. Salmon, *Foundations of Scientific Inference*, is a useful introduction to the history of confirmation theory from Hume through the positivists and their successors. D.C. Stove, *Hume, Probability and Induction*, attempts to solve the problem of induction probabilistically.

Objection to the logical empiricist theory of testing was early advanced by Karl Popper, *The Logic of Scientific Discovery*, first published in German in 1935. In that work, and in *Conjectures and Refutations* (1963), Popper advanced a quite striking thesis about which theories to accept and why. An excellent critical discussion of Popper's views is to be found in W. Newton Smith, *The Rationality of Science*. Balashov and Rosenberg reprint a portion of Popper's *Conjectures and Refutations*, along with his attack on the theory of natural selection, "Darwinism as a Metaphysical Research Program", and the relevant portion of Darwin's *On the Origin of Species*.

The arguments against strict falsification of W.V.O. Quine, *From a Logical Point of View*, and *Word and Object*, followed a much earlier work, P. Duhem, *The Aim and Structure of Physical Theory*. The recognition that the role of auxiliary hypotheses makes strict falsification impossible limited the influence of Popper's views among philosophers.

L. Savage, *Foundations of Statistics*, provides a rigorous presentation of Bayesianism, as does R. Jeffrey, *The Logic of Decision*. A philosophically sophisticated presentation is P. Horwich, *Probability and Evidence*. An introduction to Bayesianism is to be found in Salmon's *Foundations of Scientific Inference*. Salmon defends the application of the theorem to cases from the history of science in "Bayes' Theorem and the History of Science", reprinted in Balashov and Rosenberg.

The problem of old evidence, among other issues, has led to dissent from Bayesianism by C. Glymour, *Theory and Evidence*.

Peter Achinstein, *The Book of Evidence*, anthologizes several papers that reflect the complexities of inference from evidence to theory.

The possibility of underdetermination is broached first in Quine, *Word and Object*. It has been subject to sustained critical scrutiny over the succeeding half-century. For an important example of this criticism, see J. Leplin and L. Laudan, "Empirical Equivalence and Underdetermination". C. Hoefer and A. Rosenberg, "Empirical Equivalence, Underdetermination and Systems of the World", respond to their denial of underdetermination.