11 Confirmation, Falsification, Underdetermination

- Overview
- Epistemological Problems of Hypothesis Testing
- Induction as a Pseudo-Problem: Popper's Gambit
- Underdetermination
- Summary
- Study Questions
- Suggested Readings

Overview

Not satisfied with Hume's problem of induction, creative twentieth-century philosophers have created several more fundamental conceptual problems to be overcome by an empiricist epistemology that grounds the general laws and theories characteristic of so much contemporary science. Among these are Hempel's paradoxes of induction and Goodman's "new riddle of induction." Both show how deeply theoretically entrenched hypothesis testing really is.

At least one important twentieth-century philosopher, Karl Popper, thought he had a way around the problem of induction. Indeed, he thought that the whole problem of building up evidence for a theory represents a deep misunderstanding of what science is all about and how it proceeds. Ironically, the pursuit of his approach to theory testing not only failed to solve the problem but raised so formidable a challenge to empiricism as the official epistemology of science that it produced a movement which simply denies that science is controlled by experience, and even threatens the objectivity of science altogether.

Epistemological Problems of Hypothesis Testing

Assume, along with all working scientists, that either we can solve the problem of induction, or that it is no problem at all, but a pseudo-problem, as many philosophers have suggested. Accept that we can acquire knowledge about the future and about laws by experience. Remember this is the claim of the empiricist. It is not a claim about what causes our beliefs about the future and about laws. Every one will grant that it is experience that does so. No one any longer considers knowledge of how the world works to be innate. Empiricism is a thesis about justification, not (merely) about causation. Experience justifies as well as causes those beliefs that count as knowledge.

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 As are Bs—for instance, "All samples of copper are electrical conductors," could be conclusively confirmed because the hypothesis will be about an indefinite number of As 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. What is more, and as we have seen is deeply puzzling, often scientists will rightly embrace a hypothesis as expressing a strict law of nature, true everywhere and always, on the basis of a very small number of experiments or observations. The relation of positive evidence to hypotheses it confirms is obviously complex.

On the other hand, it may seem that such hypotheses could at least be falsified. After all, to show that All As are Bs 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 is a fact of the first importance for our understanding of science.

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, because it doesn't follow that there are any swans at all. Recall that Newton's first law may be vacuously true: there are no bodies in the universe free from all forces, owing to the presence of gravitational forces everywhere. To test the generalization about swans 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 PV = 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. 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 which 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." Suppose we are presented with a white 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 As are Bs" if and only if "All non-Bs are non-As." To see this, consider what would be an exception to "All As are Bs." It would be an A that was not a B. But this would also be the only exception to "All non-Bs are non-As." 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, a.k.a. all swans are white. The black boot is a positive instance of the hypothesis that all swans are white. To many it will seem as though 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.

This puzzle due to Carl G. Hempel is known as "the paradox of confirmation." There are two broad strategies for dealing with this paradox. Hempel's preferred approach was simply to accept that black boots confirm the hypothesis about all swans being white, and to explain the feeling that they don't do so away as a logically unsophisticated attitude we can disregard. The other alternative is to argue that if "all swans are white" is a law, it must express some necessary connection between being a swan and whiteness. Recall, this is the explanation of what the law supports, the counterfactual that if my black boot had been a swan it would have been white. If "all swans are white" is an expression of physical or natural necessity, then it will not be logically equivalent to "all non-white things are non-swans," since this statement obviously lacks any natural or physical necessity. Now a black boot, which is a non-white non-swan, may support this latter general statement, but since it is not equivalent to the law (as it bears no nomic necessity), the black boot won't be a positive instance of the swan-hypothesis. This avoids the problem but at the cost of forcing us to take seriously the nature of nomic or physical necessity, something empiricists, and especially logical positivists like Hempel, were reluctant to do. They and we may have to anyway, once we explore the next of our problems about positive instances.

It is worth noting here that exponents of a Bayesian approach to induction argue that the paradox of confirmation is no problem for 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 conditional probability of seeing a white swan, conditional on all swans being white.

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 that 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."

A predicate's being projectable, a general statement's supporting counterfactuals, a regularity's having explanatory power, a predicate-term naming real property, a universal, and a law's being supported by its positive instances—all of these notions turn out to be far more closely tied together than anyone thought.

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.

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 that 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 frame substantial conjectures that make strong claims about experience and then to try as hard as they can to falsify these conjectures, to find evidence against them, and when they succeed in falsifying them, 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 Fs are Gs," 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 3 the example of one sentence expressing more than a single proposition. Depending on the emphasis the sentence "Why did Ms. 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 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 pseudoscience. 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 9) 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 classrooms, 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."

Stigmatizing some theories as pseudoscience 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 that attacked Marxist political economy and political philosophy. Many social scientists made common cause with Popper on this account. 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 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 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 effected 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 (a) makes any prediction about future tests of the theory, or (b) 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 shortsighted 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.

Underdetermination

The testing of claims about unobservable things, states, events, and 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 disturbing "**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, and experiment. The objectivity of science is held to rest on the role that 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. If it is empirical data and only empirical data that test theory, and if empirical data do not point to where theories that are disconfirmed need to be changed, then over time, theories in the sciences should continually proliferate. But this does not appear to be the case, especially in the physical sciences. We return this point and its implications below.

The problem of empirically equivalent but logically incompatible theories becomes especially serious as science becomes more theoretical. A famous example of van Fraassen's illustrates this point. Recall van Fraassen's view, "constructive empiricism," which urges that we be agnostic about the truth claims of the theoretical "parts" of theories. One of his arguments rests on the following possibility: compare Newtonian mechanics—the four laws we discussed in Chapter 7 that explain so much in physics. Add to these four laws the further axiom that the universe and everything in it is moving at 100 kilometers per hour along a vector whose direction is given by extending a line from the Earth to the North Star, Polaris. These two theories will be equally powerful in explanation and prediction, there will be no empirical way to tell the difference between them. They are empirically equivalent. Which we choose is underdetermined by observations.

Suppose one objects to this example on the ground, well-established by Einstein's special theory, and before that argued by several philosophers from Leibniz and Berkeley in the eighteenth century onwards, that there is no such thing as motion in absolute space. Since there is empirical evidence that falsifies this assumption, we cannot add to Newton's laws to produce an empirically equivalent theory. This line of argument has several problems. First, if it relies on the strict falsification of the assumption of motion in a direction in absolute space, then it begs the question against underdetermination to begin with. Second, it is only retrospectively that we know the additional assumption to be factually false. Since we cannot know which are the false parts of theories that are empirically equivalent with respect to evidence available at a given time, as opposed to later, we cannot use this objection to deny the possibility of underdetermination at any time. Third, the example van Fraassen constructed is meant to illustrate a possibility that for all we know could have obtained in the past and may well obtain in the future.

Imagine, now, the "end of inquiry" when all the data on every subject are 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 about everything, our two total "systems of the world" must disagree somewhere, they must be incompatible, and therefore cannot both be true. We cannot either remain agnostic about whether one is right or ecumenical about embracing both. Yet it appears that observation would not be able to decide between these theories.

Contemporary debates in cosmology may illustrate the possibility we contemplate here. Worse, they may exemplify its actuality. There are several versions of a superstring "theory of everything," which combine both quantum mechanics and the general theory of relativity, and between which it may be impossible to choose empirically, since the observations needed to do so require at least as much energy as there is in the universe. In addition there are non-string theory alternatives, so-called quantum-loop gravity theories, which are also at present empirically equivalent, and may for all we know be permanently empirically equivalent.

If it is possible for there to be, at the actual or merely hypothetical "end of inquiry" more than one theory equally empirically adequate, equally compatible with all the evidence, equally general in explanatory domain and equally precise in predictive power, then clearly empiricism is in serious trouble. As total theories, they cannot both be true for they will disagree about something theoretical. Yet as empirically equivalent, no data will be able to choose between them. If there is a fact of the matter about which one is true, it will not be accessible on any empiricist epistemology.

Even if we find a way to rule out incompatible, empirically equivalent, equally powerful total theories, we still face a severe problem of the underdetermination by observation of the actual theories that scientists have adopted. And yet science does not show the sort of proliferation of theory that this empirical underdetermination would lead one to expect. The kind of intractable, irresolvable theoretical disputes that underdetermination makes possible are almost never actual. The more we consider reasons why this sort of underdetermination does not manifest itself, the more problematical 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, explanatory unification, precision in prediction, and consistency with other already adopted theories. In theory choice we are not limited merely to the derivation of predictions for observation that we can test. When observations disconfirm a body of hypotheses, there are methodological guidelines which enable us to design new experiments and tests, which may enable us to point the finger more precisely at one or another of the components of our original package of hypotheses under test. And here too, considerations of simplicity, explanatory unification, precision of prediction, amount of allowable experimental error, and consistency with other established theories apply again. Theory choice is a continual process of iterative application of this same toolbox of considerations in order to assess the implications of empirical observation in making theory choices.

Theory choice sounds like it is controlled by empirical observation after all, though ones that are guided by broad rules and regulations of empirical inquiry. But what are the grounds of these rules and regulations? Two obvious answers suggest themselves, neither satisfactory. First, the methodological principles we add to observation in order to eliminate the threat of underdetermination could be vouched safe to us by some sort of *a priori* considerations. Rationalists like Kant thought there were such criteria for theory choice that guaranteed the truth of Newton's theory against skepticism. Clearly, this source of assurance against the threat of underdetermination won't be tolerated by empiricism. But the alternative looks question-begging.

Suppose empiricists argue that the extra-observational criteria of theory choice have themselves been vindicated by observation and experiment. But then these criteria would be guilty of simply invoking 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. An empiricist justification for the methodological rules we employ in theory choice is circular as an argument against the threat of underdetermination.

Having excluded both rationalist and empiricist sources for the consensus that science has shown in the face of underdetermination, over the last 400 years especially, philosophers of science face a serious problem.

There is one alternative source that almost all philosophers of science are strongly disinclined to accept: 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

In this chapter we started out with what looked like small problems, cute paradoxes, clever philosophical conundrums about white swans and black boots and funny properties that seem to involve changing colors. But these matters ended up making very serious trouble for empiricism.

Attempts like Popper's to entirely short-circuit the inquiry about how evidence supports theory and replaces it with the question of how it falsifies theory, seem to boomerang into the persistent possibility of underdetermination, and the global threat that it may not be observation, experiment and data collection that really control inquiry. In other words, a series of problems confronting empiricism has become increasingly formidable. This is especially so when we consider that these problems for empiricism have in fact surfaced and been taken seriously, largely by empiricists seeking to ground ever more firmly their epistemology.

As we have seen, 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 needed 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 is 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, and the desire for fame or at least security, 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. Defend or criticize: "We all know a positive instance when we see one. No one need worry about the paradox of confirmation. The same goes for telling projectable predicates like 'green' from unprojectable ones like 'grue.' "
2. What's wrong with this claim: "It's always clear in science which statement is falsified by a disconfirmed prediction, and it's never the claims about the apparatus. 'A poor carpenter always blames his tools' is a sound maxim in experimental science."
3. Why can't we always claim that two equally well confirmed total theories that appear to be incompatible are only disguised terminological variants of one another?
4. Exactly why is underdetermination a real threat to empiricism, and to the objectivity of science?

Suggested Readings

The paradoxes of confirmation first broached by Hempel in essays on confirmation theory collected in his *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. Hempel's and Goodman's papers are anthologized in Lange. 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.

Popper, "Science: Conjectures and Refutations," from the book of the same name is reprinted in Curd and Cover, as his attack on "The Problem of Induction."

The possibility of underdetermination was recognized early in the twentieth century by the French philosopher, Pierre Duhem, in *The Aim and Structure of Physical Theory*, and later made central to the work of Quine, *Word and Object*. His work is discussed at length in Chapter 14. The problem of underdetermination was subject to sustained critical scrutiny over the succeeding half-century. A paper by Duhem is reprinted in Curd and Cover. For an important example of this criticism, see J. Leplin and L. Laudan, "Empirical Equivalence and Underdetermination," *Journal of Philosophy* 88(1991): 449–472; and C. Hoefer and A. Rosenberg, "Empirical Equivalence, Underdetermination and Systems of the World," *Philosophy of Science* 61(1994): 592–607, respond to their denial of underdetermination. Laudan, "Demystifying Underdetermination," is reprinted in Curd and Cover.