CHAPTER EIGHT

Scientific Methodology

What is the foundation of all conclusions from experience? This implies a new question, which may be of more difficult solutions and explication.

—David Hume (1748/1974)

Many of us learned in grade school a rather simple and stilted process labeled the scientific method. A science education website aimed at schoolchildren still proposes such a formulation in six steps:

1. Ask a question
2. Do background research
3. Construct a hypothesis
4. Test your hypothesis by doing an experiment
5. Analyze your data and draw a conclusion
6. Communicate your results (Steps of the Scientific Method, 2009)

For schoolchildren, this may be a good introduction, but it does not begin to hint at the controversies that the presumptions underlying this "method" have sparked. This method has been challenged both in theory and in practice. Philosophers of science have pointed to theoretical problems underlying this method; and historians of science (and some scientists themselves—see Bauer, 1992) have challenged the presumption that science is always, normally, ideally, or even ever performed according to this rubric. We begin with the most fundamental problem underlying this method and continue with various attempts to resolve this problem.

THE PROBLEM OF INDUCTION

An inductive generalization is the logical process of attributing a quality to a whole class of objects based on our limited experience of some sample of that class. The interesting, and perhaps troublesome, aspect of this logical process is that it is common not just in scientific methodology but in daily life as well. We each make these types of inferences every day, often without even noting or realizing it. Merely opening your front door requires what seems to be an implicit inductive inference that turning the knob has opened the door in the past and will do so today. Scientific claims, laws, and theories are largely based on inductive inferences. Isaac Newton's law of inertia was inductively inferred from the observation of many individual objects that manifested this quality. From the continual confirmation of the hypothesis that all matter has the quality of inertia, the inference is made that the claim of inertia as a property of all matter is in fact a law about matter. In this example, we see the direct correlation between induction and empiricism. Empiricism holds that knowledge is attained through empirical observation; observation of the eyes, ears, and so forth. Empirical observation is always observation of particulars. That is, we observe particular dogs, not the species dog or the concept of dog. We observe particular instances of matter manifesting inertia, not matter itself or inertia itself as general concepts or realities. Logical positivists/empiricists, being rather strict empiricists, noted this aspect of observation and saw at least part of the method of science to take these particular instances and generalize from them inductive inferences as a product of knowledge. The particular observations in themselves are not real or strong knowledge. The inferences based on these observations are bolder, more useful forms of knowledge. The claim that a particular billiard ball has the property of inertia is a less bold and less useful claim than the claim that all matter has this property. The claim that a new antibiotic will kill a certain range of bacteria in one patient is a less bold and less useful claim than the claim that this antibiotic will kill such bacteria in all patients. However, the stronger claim in both instances is more difficult to establish and much riskier in terms of being knowledge. As we learned in Chapter 4, inductive reasoning only leads to probabilistically supported conclusions. Thus, all inductively reached scientific claims, laws, and theories are true only to a degree of probability and always carry with them the logical possibility of being proven false. Yet there is an even deeper problem underlying this limitation. According to the classical empiricist David Hume (a person clearly admired by both logical positivists and logical empiricists), induction is not only limited by probability but is not even "founded on reasoning, or any form of understanding" (Hume, 1748/1974, p. 328).

Let us take a simple example of inductive reasoning. Most people accept, without giving it much thought, that the sun will rise tomorrow.[1] Why do they believe this? The simplest explanation is that the sun has risen every day in their lives and in recorded history in the past. So, the inference might be expressed in the following simple argument:

The sun has risen every day in the past.
Therefore, the sun will rise tomorrow.

At first blush this may seem like a clear argument. Yet, there is an important premise missing:

The sun has risen every day in the past.
The future will continue to be like the past.
Therefore, the sun will rise tomorrow.

This new version provides a general principle (the italicized premise) that better connects the original premise with the conclusion, more clearly connecting the past to the future. But this new claim is clearly open to doubt and in need of evidentiary support. Being a general principle, it is not something known through direct experience, as explained previously, but reached through some form of inference or as a self-evident, analytic claim. The denial of an analytic claim is a self-contradiction: for example, "All bachelors are unmarried." The denial of that claim would be "It is not the case that all bachelors are unmarried," or more clearly, "Some bachelors are married." But to claim that any bachelor is married is a contradiction of the concept of "bachelor." However, the denial of the claim "The future will continue to be like the past" does not result in a self-contradictory claim. To say that the future will *not* continue to be like the past is an empirical and speculative claim. It can only be proven true by observing the future. So, this principle is not an analytic claim and cannot be said to be true on that basis. Thus, it must be an empirical claim and require empirical support. One might attempt to provide support for this claim by adding the premise "In the past the future has been like the past." Yet to use this claim to support the "The future will continue to be like the past" claim would be to employ once again the type of reasoning we are trying to prove—to assume once again that the future will be like the past by again being like the past. This is a type of logical error (fallacious reasoning) philosophers refer to as circular reasoning or begging the question: employing what is intended to be proven in order to prove it.

CAUSALITY

Another form of inductive reasoning, called causality, is subject to the same type of criticism. Indeed, a causal inference may be understood as a type of inductive generalization. When one makes a causal claim, which is common in science, one asserts that some event A causes some second event B. This is an empirical, not an analytic, claim. For once again to deny it does not result in a contradiction. For example, one might assert that HIV causes AIDS. To deny this claim is only contradictory if we presume the definition of HIV as that which causes AIDS, but that again results in a question-begging argument. Indeed, if this were an analytic claim, it would not have been so difficult to reach. The early history of AIDS research shows the proposal of many possible causes of AIDS: amyl nitrate, a weakening of the immune system by a succession of (especially sexually transmitted) infectious diseases, and so forth. And the fact that HIV as the cause of AIDS is so widely accepted and so strongly supported by research now that today those who will deny this claim are almost seen as maintaining a contradiction, they in fact are not. As small a possibility that this causal claim is false that there is, there still is that possibility. It is a claim that can only be justified through experience—in this case experience in terms of scientific observation and research—and inductive logic. In saying that A causes B, we are referring to two conceptually separable events. When we analytically say that all bachelors are unmarried, the two concepts (bachelors and being unmarried) are not conceptually separable because the concept of "being unmarried" is included in the concept of "bachelor." To say that HIV causes AIDS is again to synthetically connect these conceptually distinct events or phenomena. For a simpler example, let us say the cue ball strikes the eight ball and causes the eight ball to move. Here again we have two conceptually distinct events: the cue ball striking the eight ball and the eight ball moving. As they can be held distinct in this manner, they are not necessarily or analytically connected. To connect them we must appeal to experience and something known as inductive inference.

The deeper problem, according to Hume, has to do with our concept of causality itself. This concept includes three primary elements. The first is temporality. That is, when we assert that A causes B, it is presumed that A temporally precedes B. You never have an effect before a cause. The second element is spatial contiguity. The cause and the effect must come into physical contact. We say the cue ball causes the eight ball to move because it strikes the eight ball. We do not attribute the movement of the eight ball to the five ball, which is resting still at the other end of the table. Even when we seem to attribute causality at a distance, we don't really (setting aside quantum mechanics for the moment). To say that pressing the power button on the remote control causes the television (TV) to turn on leaves unsaid but implied that pressing the power button causes an infrared beam, which, when operating on a sensor on the TV, causes the TV to turn on. When we say the remote power button causes the TV to turn on, there is implied a causal chain. But beyond these two elements, there is a third, which is more conceptually troublesome. Saying that A causes B means more than that A precedes B (temporality) and that A and B occur in the same place (spatial contiguity). Also part of our concept of causality is that there is some power in A to bring about B. Hume referred to this element as the *necessary connection*. The trouble with this element is that while the first two (temporality and spatial contiguity) are empirical, this necessary connection is not. A necessary connection implies that any causal claim is analytical, but as we have already seen, it is not. This necessary connection is something that we cannot see or perceive in any manner. "Causality" is, itself, beyond perception. Any causal inference we make can only be based on the former two elements: repeated observations of A occurring before B, and A and B occurring at the same place. Thus, we inductively generalize from particular observations of these two elements to the claim that the two events are connected by this seemingly metaphysical concept of causality. Yet, Hume doubted that such an inference is rationally supported:

> The bread, which I formerly eat, nourished me; that is, a body of such sensible qualities was … endued with such secret powers: but does it follow, that other bread must also nourish me at another time, and that like sensible qualities must always be attended with like secret powers? (Hume, 1748/1974, p. 329)

As with the sun rising tomorrow there is no logical reason to accept that bread will continue to nourish me, or that the cue ball striking the eight ball will in the future cause movement in the eight ball, that HIV will continue to result in AIDS. Hume ultimately could find nothing more than custom or habit to attribute our inductive inferences to. That is, from viewing two events together again and again, we by habit or custom come to expect them to be conjoined in the future: a new day conjoined with a rising sun, HIV conjoined with AIDS, the striking of the eight ball conjoined with movement of that ball, the eating of bread conjoined to nourishment. But this expectation is merely psychological, not logical.

Now Hume did not completely reject causality or induction in general. As a human being he had to accept and work with those concepts on a daily basis, as we all do. His main problem was that as a philosopher, especially one interested in understanding and justifying scientific knowledge, he could find no logical justification for induction. And therein lay the problem: as a practical human being or as a scientist he has to employ and accept the practice of induction. But as a philosopher he can find no reasoning to back it up, leaving the status of inductively "inferred" claims as *knowledge* fundamentally uncertain.

HYPOTHETICO-DEDUCTIVISM

Empiricists in the 19th century (e.g., John Stuart Mill) and in the 20th century (logical positivists/empiricists) worked to find a better inductive methodology. In much of the logical positivist/empiricist work you can find many hopeful statements about the improvements on induction to come. A reliance on induction follows logically from the central logical positivist doctrine of verifiability. Many philosophers proposed strategies to solve the problem of induction. Although none was successful in addressing the fundamental problem, a number of these strategies were valuable for highlighting problems attendant to the problem of induction and possibly making the problem of induction seem less intractable.

One simple strategy is the broadening of the role of the hypothesis. If we look back to the inductive method at the beginning of modern science, in the work of Francis Bacon (1561–1626), we see a very simple approach in his *Novum Organum*. He outlined his empirical investigation into the concept and phenomenon of heat. His method included merely listing everything that had the properties of heat (fire, the sun, etc.) and those things that lacked heat (ice, cold earth, etc.), then abstracting from all the hot things what they had in common, and which those things on the list of non-hot things lacked, to determine what caused heat. There is a simple practical problem to this method that reflects the more general problem of induction. A list of hot things and a list of non-hot things would each be at least indefinite if not infinite in length.[2] The scientific method of hypothetico-deductivism, as employed by the logical positivists/empiricists, like Carl Hempel (1966/2000), was intended to address this problem. The basic structure of the method is to suggest a likely *hypothesis* to explain an event or predict the recurrence of an event, then logically deduce what would and would not follow logically (*deductively*) if that hypothesis were true—thus, "hypothetico-deductivism." Experiments then are structured to see if the logically deduced consequences follow. The introduction of a hypothesis provides some boundaries and guidelines, thereby limiting the observations that would be relevant to include in an investigation, for, in the words of Hempel, "a collection of *all* the facts would have to await the end of the world" (1966/2000, p. 45). But what exactly does "relevant" mean? What makes a fact or an observation relevant to an investigation? First, it means relevant to the hypothesis in question. A little more deeply it means that "either its occurrence or its nonoccurrence can be inferred from" the hypothesis in question. Hempel employs the example of physician Ignaz Semmelweis, who, while working in the 1840s at Vienna General Hospital, investigated the cause of childbed fever, which took the lives of many new mothers in Vienna. What was particularly interesting and concerning was that there was much more childbed fever in the First Maternity Division than in the Second Maternity Division. One hypothesis he investigated was that the birthing position might be the cause. Women in the Second Division more commonly birthed laterally, and those in the First Division more commonly birthed on their backs. It seemed that those delivering on their sides had less incidence of childbed fever. Thus, the hypothesis was that birthing in a supine position causes childbed fever. The question of relevance then hinges on what facts would be logically implied by this hypothesis. Most obviously a fact such as a rise or decrease in childbed fever on changing the birthing position would be logically relevant. That is, if the hypothesis were true, we would expect a change in the First Division to lateral birthing to decrease the incidence of childbed fever. If the hypothesis were not true, we would expect no change in changing the birthing position in the First Division. In fact, he found that changing the birthing position had no effect on the incidence of childbed fever and concluded then that the hypothesis was false. The logical structure of such an investigation would be as follows (where H is some hypothesis and E is a logically relevant event or fact):

If H is true, then E would follow (is true).
E does not follow (is not true).
Therefore, H is not true.

This is a logically valid argument. The form is known as *modus tollens*. Here is a simple example to illustrate the validity of the form in general:

If Bill is in Philadelphia, then Bill is in Pennsylvania.
Bill is not in Pennsylvania.
Therefore, Bill is not in Philadelphia.

Here it should be clear that the conclusion follows logically from the premises. If one is not in Pennsylvania, then one cannot be in Philadelphia—given that we are speaking of the specific Philadelphia that is in Pennsylvania, as the first premise asserts. The previous argument has the same *modus tollens* form and thus is deductively valid also. So it seems that childbed fever is not caused by birthing position. Another hypothesis Semmelweis investigated was that "cadaveric matter" was the cause. Most of the women in the Second Division were attended by midwives, while most of the women in the First Division were attended by physicians and medical students who would often come directly from performing autopsies. To test this hypothesis he had physicians and medical students wash their hands before moving from autopsies to birthing. This may seem like a ridiculously obvious solution, but bear in mind that in the 1840s the germ theory of disease was still a contentious, not widely accepted theory. And, as we might expect, on instituting a hand washing policy the occurrence of childbed fever decreased. Semmelweis thus concluded that childbed fever was caused by an infection delivered from cadavers by physicians and medical students to new mothers. The formal structure of the reasoning is a bit different in this case:

If H is true, then E would follow (is true).
E does follow (is true).
Therefore, H is true.

Logicians have a name for this argument structure also: Affirming the Consequent. For in a conditional (if–then) statement like the first premise, the smaller statement that appears between the "if" and the "then" is called the antecedent, and the smaller statement that follows the "then" is called the consequent. The second premise then affirms the truth of the consequent. The conclusion is the affirmation of the first premise's antecedent. The problem here though is that Affirming the Consequent is not a valid argument form. It is an invalid form—also called a formal fallacy. Compare this simpler example:

If Bill is in Philadelphia, then Bill is in Pennsylvania.
Bill is in Pennsylvania
Therefore, Bill is in Philadelphia.

It does not take much reflection to realize here that the conclusion does not follow with certainty from the premises. Given that Bill is in Pennsylvania and that Philadelphia is in Pennsylvania, it *might* be the case that Bill is in Philadelphia. But "might" is not good enough for deductive validity. There are hundreds of other cities and towns he might be in other than Philadelphia. A little reflection will demonstrate also how the aforementioned argument also does not guarantee its conclusion of the truth of the hypothesis. Thus, logically, Semmelweis has not established the truth of his hypothesis. Even replicating this test with the same results again and again will not logically verify the truth of the hypothesis. However, following the basic understanding of inductive reasoning—which as a matter of course sets aside Hume's actual conclusions and focuses on the hope that a better inductive mechanism will be found—"extensive testing with entirely favorable results … provides … more or less strong support for" the hypothesis (Hempel, 1966/2000, p. 48). This more or less strong support, something short of "proof," is commonly referred to by logical positivists and empiricists as "confirmation," leaving then scientific claims to be merely inductively confirmed through replicated tests, rather than deductively proven.

PROBABILITY THEORY

CLASSICAL PROBABILITY

Yet, how many positive results does a hypothesis need in order to be confirmed? No clear answer seems forthcoming. In order to provide such an answer, some philosophers of science turned, as did some early modern scientists like Isaac Newton, to mathematics. Particularly, they turned to probability theory. Most simply, probability theory refers to the study of the likelihood that certain events will occur. This science/mathematic began in the 17th century, with the work of Blaise Pascal (1623–1662), Christiaan Huygens (1629–1695), Jacob Bernoulli (1654–1705), and Pierre-Simon Laplace (1749–1827), as "classical probability." In classical probability, any event which we are sure will occur is given a value of 1, and any event which we are sure will not occur is given the value of 0. An example of the first case would be flipping a fair coin and having it land on either heads or tails. An example of the second case would be flipping a fair coin and having it land on both heads and tails. In between these extremes, the probability of the occurrence of an event is expressed as a ratio of number of possibilities of the occurrence in question over the total possible outcomes. Therefore, in flipping a fair coin, the probability that it will land heads is $\frac{1}{2}$, as there is one heads side and two total possible outcomes. In rolling a fair die, there are six total possible outcomes. Any one of them on any one roll has a probability of $\frac{1}{6}$. The probability that an even number will be rolled is $\frac{3}{6}$, which can be reduced to $\frac{1}{2}$. In drawing a card from a randomly shuffled deck of 52 cards, the probability of drawing a spade is $\frac{13}{52}$ or $\frac{1}{4}$.

Classical probability is also called an a priori theory of probability, because everything can be known without any empirical foreknowledge. All that is needed are the initial conditions. In regard to coin flipping, what is needed to know is that the coin is "fair" and that the coin has two distinct sides. In regard to die rolling, all that is needed to know is that the die is "fair" and has six sides labeled 1 to 6. In regard to card drawing, all that is needed to know is that the deck is randomly shuffled, that there are 52 distinct cards, and that there are four suits, each comprising 13 of the 52 cards. No experiments or trial runs are necessary in any of these cases. Probability can be calculated purely from these initial conditions.

Taken further, classical probability theory can calculate the probability of more complex events. Say we want to determine the probability of two independent events, two events whose occurrences have no influence on the occurrence of the other. For example, we might want to know what the probability of flipping a fair coin twice and landing heads twice. In this case, we would multiply the independent probability of each event: $P(a) \times P(b)$, where "a" is the first toss and "b" is the second toss. Each toss of a fair coin has a $\frac{1}{2}$ probability of landing heads. Hence, to apply the formula: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Thus, the probability of two tosses landing heads is $\frac{1}{4}$. Alternatively, we might wish to calculate the probability of two events that do influence each other—in other words, are not independent. We might want to calculate the probability of drawing two kings from a randomly shuffled deck of cards with two draws. Here the formula is a little different: $P(a) \times P(b \text{ if } a)$, meaning that we take the probability of drawing a king on the first draw and multiply that by the probability of drawing a king on the second draw. However, the probability of the second draw has been affected by the first draw because there is now one fewer king and one fewer card in total. So, on the first draw the probability of drawing a king is $\frac{4}{52}$ or $\frac{1}{13}$. On the second draw, the probability is $\frac{3}{51}$. To apply our formula: $\frac{1}{13} \times \frac{3}{51} = \frac{3}{663}$ or $\frac{1}{221}$. We can also calculate the probability of getting two mutually exclusive alternative events. Mutually exclusive means that the two events cannot both happen; the occurrence of one excludes the occurrence of the other. For example, we might want to know the chances of getting either two heads or two tails with two tosses of a fair coin. In this case, we would add the probabilities of the two events: *P*(

a) + *P*(*b*). For the probability of two heads we return to our first formula: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. The probability of two tails is the same. So then we apply our formula: $\frac{1}{4}+\frac{1}{4}=\frac{2}{4}$ or $\frac{1}{2}$. So the probability of getting either two heads or two tails is $\frac{1}{2}$. For the probability of alternative events that are not mutually exclusive, such as getting at least one head on two coin tosses, we amend this formula slightly. Let us call the coin landing heads at least once "*a*" and the coin landing other than heads on both tosses "not-*a*." As we know we will end up with a result of either "*a*" or "not-*a*"; the probability as a whole is 1, certainty. So the probability of the coin landing heads at least once will be 1 minus the probability of "not-*a*." The probability of "not-*a*" will be the product of the probability of landing other than heads on each toss: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Hence, we apply our formula: $1-\frac{1}{4}=\frac{3}{4}$. Thus, the probability of getting heads on at least one of two tosses is $\frac{3}{4}$. A little reflection or reapplication of the formula will demonstrate that this is also the probability of getting at least one tail on two coin flips.[3]

The ubiquitous examples of coin flips, die throws, and card draws are no accident. It is believed that the study of classical probability began with a dispute between Blaise Pascal and Pierre de Fermat (1601–1665) over a game of chance (Copi & Cohen, 1994). This grounding in games of chance points to a limitation of classical probability, which led to the development of more sophisticated theories of probability more applicable to scientific methodology. Classical probability contains a presumed principle known as equipossibility or the principle of indifference. Note the constant use of the word *fair* earlier to qualify die or coin. A fair coin is one that is not biased, not weighted or altered in some way to make landing heads or tails more than 50% likely. A fair die similarly is one that is not biased (or "loaded" in the vernacular) to land on any number more than the others. In other words, all possibilities are equally likely. The problem is that when we get beyond artificial games of chance to real-world questions of probability, equipossibility does not seem to hold. When it comes to games of chance, "the possible outcomes can be classified neatly into *n* mutually exclusive, completely exhaustive cases that fulfill the conditions of equipossibility" (Carnap, 1966, p. 25). But when it comes to more complex natural and human events (events studied by meteorology, physics, the social sciences, etc.), possible outcomes cannot be exhausted in an a priori manner. Of course even games of chance are arguably not so simply determined, because the assumption of fair may not always be warranted and the number of variables may exceed those we can account for. Bear in mind that the purpose of probability theory, consistent with inductive logic and empirical science, is to utilize knowledge we do have to infer about knowledge that is beyond our direct perception. When flipping a fair coin we know that it will land either heads or tails. But we do not know which. However, if we could calculate not only the formal probabilities covered by classical probability but the infinitude of other variables, we could predict with certainty the outcome of each flip. Pierre-Simon Laplace proposed a thought experiment that has come to be known as Laplace's demon. Imagine a being of immense knowledge and understanding who knows both the position of every particle of matter in the universe and every law of motion controlling that matter. Given some initial state of matter, this being would be able to predict the future with complete accuracy. It would be like us looking at the children's game Mousetrap. By following the Rube Goldbergian path that turning the crank will send the ball bearing, we can predict that the trap will fall on the mouse. To Laplace's demon the whole of the universe would be as simple to take in as the game Mousetrap is to us. Laplace's demon would then be able to predict the flip of any coin by knowing not just the formal probabilities of a fair coin with two sides but the specific peculiarities of the hand flipping the coin, air currents that may affect the flip, and even minor peculiarities of the coin itself that may make it slightly less than fair and any number of other seemingly incidental variables. One result of all this is that even with classical probability, the outcomes are "merely" theoretical, not absolutely conclusive. Thus, even though the chance of a coin landing heads is 1:2, it is possible to have 10 heads flips in a row. It is possible to have 100 as well. Probability theory does not tell us that if you flip a coin 1,000 times that 500 will be heads and 500 will be tails. It does not even tell us that 1,000 heads in a row is not possible. The most it can tell us is that the ideal case will result, at least eventually, in a 50:50 situation. It is worth noting that if a pair or one of a pair of gamblers suspected that a coin or die was not fair, they would test it empirically. Although in testing the coin or die, they would test it against the a priori expectations—perhaps not seeking perfect compliance, but not a very great deviation.

Yet experience has taught us that classical probability is good enough to provide us reliable knowledge in these limited situations with limited possible outcomes and limited "significant" variables. This is not the case in many real-world cases where the lack of equipossibility is a real problem. One development in probability theory to deal with this problem is probability based on relative frequency. Classical probability is based on absolute frequency because we know, a priori, all the possible outcomes. With relative frequency we cannot presume such knowledge. What makes this technique relative is that it is relative to the available evidence, not based on the absolute knowledge of only six sides to a die or two sides to a coin. As noted previously, this absolute knowledge may not be as absolute as commonly presumed; however, as also noted earlier, it seems good enough to make classical probability a practical method for games of chance and other limited cases. This also means that probability of relative frequency is not a priori because the information needed cannot be attained merely from fully known initial conditions but instead requires empirical investigation. The probability of relative frequency is the type of probability associated with statistics. If we wish to understand how many Hispanic expectant women receive prenatal care, we would take a sample (e.g., 1,000 for the sake of simplicity) of women who fit into the class of interest and determine how many of them received prenatal care. Let us say that number is 683, which would be 68.3% of our sample. For any pregnant Hispanic woman we could then attribute a 68.3% chance of receiving prenatal care. This is the type of probability deemed more appropriate for scientific investigation. Whether applied to physics, meteorology, or the social sciences, it takes into account the complexity of the world and provides not simply a vague sense of a more or less strongly confirmed hypothesis but a more quantifiable "degree of confirmation" (Carnap, 1966, p. 22). In addition, the aforementioned classical probability formulae, with appropriate consideration for a lack of the principle of indifference, functions in statistical (relative frequency) probability as well.

SUBJECTIVE PROBABILITY AND BAYES' THEOREM

Another theory of probability that many scientists and philosophers of science find more appropriate for scientific investigation is the subjectivist theory of probability. It is called subjectivist because instead of the objective information of the initial conditions utilized in classical probability (the number of sides of a coin, the number of sides of a die, or the number of cards in a deck) or the objective empirical information utilized in relative frequency probability, subjective probability is based on the beliefs of individuals. Like classical probability, there is a gaming application to this technique. When betting on a football game, there are not the stable initial conditions found in flipping coins or throwing dice. The probability that one team will win is determined by what an individual is willing to risk in a bet. A football fan who is willing to take a 3-to-1 bet on her team believes that the team has a 75% chance of winning. This technique has the odd and seemingly counterintuitive consequence of assigning different, "valid" probabilities to any single event. The aforementioned objective theories worked to avoid this type of situation. But this variability is acceptable in part because it is recognized that probabilities do not inhere in events themselves. They are, rather, attributes of belief. This justification is acceptable, because both classical probability and relative frequency probability accept this principle, based on the recognition that probability is a function not of events themselves but of our limited knowledge of the world. There is an implicit appeal here to what social scientists call rational choice theory (see Chapter 14). In addition, for much scientific investigation and hypothesis testing, there are clearly not the a priori conditions assumed in classical probability but also not the empirical observations necessary for the theory of relative frequency. New hypotheses may have no history to draw from in order to assign a probability. The same may be true of new evidence. Absent these objective determinants, probability can subjectively be assigned to the initial conditions. Yet, this theoretical approach is not wholly subjective, as these initial values will be plugged into the mathematical formulae of probability.

Philosophers of science who advocate subjective probability theory are often called Bayesians because of their application of a formula named for and designed by mathematician Reverend Thomas Bayes (1702–1761), known as Bayes' Theorem. Bayesians assert that scientists should, or in fact implicitly *do*, utilize Bayes' Theorem in testing and confirming hypotheses. Bayes' Theorem can be used to calculate the degree that evidence increases the probable truth of a hypothesis. Let us say a midwife hypothesizes that a particular birthing position reduces the pain and discomfort of childbirth. Call this hypothesis *H*. As a matter of subjective belief, there is a probability that this hypothesis is true. If it were true, we would expect application of this technique to result in a decrease in pain and discomfort in childbirth. Call the occurrence of this outcome *E*. There is similarly a probability that *E* will occur. And there is a probability that *E* will occur assuming that *H* is true: $p(E/H)$. Once we (subjectively) assign probabilities to these basic elements, we can begin to plug them into Bayes' Theorem, which formally is structured as follows:

$$p(H/E) = \frac{p(E/H) \times p(H)}{p(E)}.$$

What we want to calculate, $p(H/E)$, is the probability that the hypothesis is true given this new evidence. So let us assign a probability of 0.7 to the midwife's hypothesis, a probability of 0.75 to the proposed evidence, and a probability of 0.9 to the occurrence of the evidence assuming the truth of the hypothesis:

$$\frac{0.9 \times 0.7}{0.75} = \frac{0.63}{0.75} = 0.84.$$

Thus, the occurrence of a decrease in pain on implementation of the proposed birthing position will increase the probability that the hypothesis is true from 70% to 84%. To some readers, this may all seem too arbitrary to be of any value, even given the justifications for the method provided in the previous paragraph. That is an understandable reaction. Yet Bayesians maintain that not only may the initial probability values be merely the subjective (though rational) beliefs of the scientists involved, but not even that minimal level of stability is necessary. Those initial values may be entirely arbitrary. This is because, according to Bayesians, as Bayes' Theorem is employed again and again and "more and more data come in … the successive values of *p(h/e)* will converge on the correct value" (Rosenberg, 2005, p. 135). This is a remarkable claim but one seemingly borne out by experience. It does not matter, then, what values one begins with. What matter are the logically derived results. And those appear to achieve the kind of universality and objectivity that we traditionally associate with knowledge.

Despite the seeming success of Bayes' Theorem, it remains a highly controversial method. One odd problem is called the problem of old evidence. We would expect that a hypothesis that explains phenomena of the past or ongoing present to be provided more support by this fact. For example, our midwife's hypothesis might seem to be supported by the fact that women of a particular culture that traditionally employs the hypothesized birthing position are known to be unusually stoic during childbirth, whereas women from a culture that traditionally employs an alternative birthing position are known to be much more expressive about pain during childbirth. Let us for the moment set aside some obvious problems of interpretation that may taint this example. Any information that is known to be true already (old evidence) will be given a probability value of 1. It has already happened or is continuing to happen, so there is no question of its probable truth: $p(E) = 1$. Also, the truth or falsity of a hypothesis will not affect the probability of what we already know to be true. So, similarly, $p(E/H) = 1$. Now, let's plug these new numbers into Bayes' Theorem:

$$\frac{1 \times 7}{1} = \frac{7}{1} = 7.$$

So, contrary to expectations, old evidence, which is absolutely certain in truth, does not raise the probability that the hypothesis is true. Rather, it leaves that probability unchanged, according to Bayes' Theorem. One Bayesian response to this problem is to say that this unexpected, seemingly counterintuitive result is in fact consistent with and an expression of a basic scientific value: Hypotheses and theories should not be designed merely to fit preconceptions. Old evidence may instill in us certain preconceptions of how the world works and we may, consciously or unconsciously, shape our hypotheses according to these preconceptions. But we should describe and explain the world from an objective point of view, not from preconceived notions. However, this answer may fail to properly distinguish from hypotheses improperly influenced by preconceived notions and those that simply (because they are right) are consistent with, and even descriptive or explanatory of, old evidence. One other Bayesian response might be to attempt to give old evidence not a value of 1 but a value that it would be given at a time before it occurred. The problem with this response is that it takes the most controversial aspect of Bayesianism to resolve this relatively simpler difficulty. Beyond the subjectivism of individually assigning probability values to events and hypotheses at question now, scientists now have to imagine themselves at another time before this old evidence is known to be true. And, as just mentioned, subjectivism in general is the most serious and fundamental problem critics of Bayesianism note. This fundamental use of subjectivity in this theoretical approach seems squarely at odds with the objective standards of knowledge traditional in scientific inquiry.

THE HYPOTHETICO-DEDUCTIVE METHOD

Probability theory seems quite a strong method, and indeed it has proven so through much of modern science. It has proven quite fruitful and quite accurate in its abilities to predict and explain much of the natural and human world. However, all this empirical evidence aside, it still leaves untouched the basic problem outlined by Hume. No matter how much mathematics is added, the conclusions of inductive inferences are still not conclusive and no good logical reasoning has been found to develop belief in these conclusions. That is, we seem to be in the same place as Hume was more than 200 years ago. Induction, especially with probability added, seems empirically justified, but rationally, logically, there is no clear justification. For this reason, some philosophers of science still resist reliance on induction as central to scientific investigation. Most important among these thinkers is the illustrious Karl Popper.

Popper proposed a variation on the hypothetico-deductivism of the logical positivists presented previously, a version that remains truer to the deductive part of the method. Popper also proposes that a scientist must construct a hypothesis as a likely explanation for a phenomenon. That hypothesis must be tested with deducible results in mind. For the positivists, if those results are obtained, the theory is confirmed—not proven, but confirmed. Taking the problem of induction more to heart, Popper is not ready to accept even this weakened standard of knowledge justification. Therefore, according to Popper, positive results of an experiment do not “confirm” a hypothesis. However, following the employment of the *modus tollens* argument, Popper asserted that a hypothesis can only be refuted—never confirmed. In making a universal claim, as scientific claims typically do, merely one counterinstance can falsify such a bold claim. This then is an appeal to Popper’s falsifiability thesis. It is Popper’s position that scientists should not pretend to confirm hypotheses. The only real goal is to try to falsify or refute them. This strategy maintains a critical attitude toward scientific claims which would in theory lead to the strongest claims, as they are continually tested. Only the most likely to be true claims would be left, but no claim could ever be proclaimed true. To go along with this critical attitude, scientists are supposed to maintain, only the riskiest, boldest hypotheses should be proposed and tested. Testing and failing to refute a weak hypothesis does not prove much. But to test and fail to refute a risky, bold hypothesis makes a stronger statement. And to test and refute a risky, bold hypothesis expresses the scientist’s proper attitude and orientation as an objective knowledge seeker, not someone simply seeking to prove her hypothesis (and ultimately herself) right. Scientific investigation is not about any individual but about the method itself and about the cumulative effect of testing. The most that positive results can provide is what Popper called “corroboration.” For Popper, corroboration is a distinct concept from confirmation. Confirmation is meant to supply affirmative reasons to believe a claim, although, as noted earlier, confirmation is epistemically weaker than “proof,” which provides irrefutable reasons for belief. Corroboration merely refers to a theory’s track record of standing up to testing, its record of not being refuted in the face of good faith attempts to do so. An appropriate scientific attitude, according to Popper, will never put full faith and belief in any scientific claim, no matter how well corroborated. Scientific claims can only be accepted provisionally or tentatively, mainly for the purpose of testing further claims, building on corroborated knowledge.

Popper’s view of methodology may seem more logically coherent, yet there are both theoretical and real-world problems with it. Falsifiability falls to the same problem that W. V. O. Quine noted regarding verifiability. Recall from Chapter 6 that Quine’s (1951/2000) landmark paper, “Two Dogmas of Empiricism,” placed the logical positivist thesis of verifiability into question by noting that no statement can be tested in isolation. Every experiment tests not only the claim in question but a host of attendant claims or presumptions: the accuracy of measurement instruments, theoretical presuppositions, and so forth. This turns out to be a problem for Popper’s falsifiability thesis as well—and by extension a problem for his form of hypothetico-deductive methodology. For, if a result appears to falsify a hypothesis, the hypothesis itself may in fact be true and the real problem may lie with one of attendant claims being implicitly tested. This leaves us in a seemingly untenable logical position regarding scientific claims, succinctly described by Imre Lakatos: “Scientific theories are not only equally unprovable, and equally improbable, but they are also equally undisprovable” (1970, p. 103).

Due to this problem, Popper replaced this simple form of falsificationism (sometimes called naive or dogmatic) with a more sophisticated version. According to sophisticated falsificationism, merely producing a refuting instance to a hypothesis or theory is not enough to refute it. In addition, there must be another theory to replace the falsified theory. And this new theory must explain all that the falsified theory explains and more (excess content); furthermore, this new content must be corroborated. These new criteria make for a more stringent methodology and provide more stability to scientific knowledge. Under naive falsificationism, any scientific claim, hypothesis, law, or theory could be recklessly disposed of by some possibly sloppy scientific work or some untruthful attendant claims. With sophisticated falsificationism, claims are not so easily disposed of. The extra criteria noted previously must be met, bringing about a more stable collection of knowledge claims of which science is comprised. However, these new criteria, while providing more stability, do not actually address the fundamental problem. The Quinean issue involving the inability to test a single claim in isolation still holds. In addition, there are claims that real science does not function as Popper described. First, scientists often do not reject a hypothesis following a single falsifying result, even with the more stable version of sophisticated falsificationism (Bauer, 1994; Kuhn, 1970; Rosenberg, 2005). More typically, with a cherished belief (theory, hypothesis) at stake, especially one on which other important beliefs are connected (which following Quine's "web of belief" concept would be most), a scientist may be more likely to make ad hoc adjustments in order to "save the theory." These would be adjustments not necessarily implied by reason or empirical evidence but adjustments (in auxiliary hypotheses) that would keep the theory (for now) from being falsified. Similarly, a scientist may blame her equipment or other variables instead of the truth of the theory before accepting a conclusion of falsification. Second, it does not seem true from the history of science and actions of real scientists that science gives as little value to confirming instances as Popper does (Bauer, 1994; Rosenberg, 2005). Indeed, the piling on of confirming instances seems in real science to have far more epistemic value than Popper's concept of "corroboration" suggests.

SUMMARY

We seem to be in a very similar position to where we started out with following Hume. That is, for Hume, induction seemed to work but he was unable to find a clear, philosophical justification for it. As for us, science seems to work. We see apparent successes throughout time, throughout the world. We have eliminated or nearly eliminated some devastating diseases, invented flying machines, and even travelled to the moon. Probability theory, even though it does not solve the fundamental question of the problem of induction, has proved invaluable to scientific progress. So, just as Hume had to accept the practicality of induction even though he could not justify it philosophically, we, it seems, must accept the practicality of a pluralistic scientific method, even though no complete philosophical justification seems forthcoming.

QUESTIONS FOR REFLECTION

1. What do you remember being taught in grade school or high school about the scientific method?
2. What, simply stated, is the problem of induction?
3. You may have heard the moral dictum "Never generalize." Yet we constantly generalize, perhaps even necessarily generalize, often to our great benefit. So, why this apparent discrepancy between this moral dictum and our epistemic and even mundane needs?
4. Will the sun rise tomorrow? On what basis do you believe this?
5. Does HIV cause AIDS? Does smoking cause cancer? How do we answer these questions? How do we affirm causality in such cases?
6. What difficulties are there in establishing that HIV causes AIDS or that smoking causes cancer? In what way are these in fact different questions?
7. Construct your own, original argument in *modus tollens* form.
8. Construct your own, original invalid argument in Affirming the Consequent form.
9. If science is supposed to produce objective knowledge, why do many scientists affirm a subjectivist approach to probability?
10. Why, according to Popper, does "riskiness" mean better science?

NOTES

1. Indeed, the fact that many will read this and accept this example as relevant and meaningful at indefinite points in the future reveals my implicit acceptance of this thesis, not only for tomorrow but for many tomorrows to come.
2. Of course, this particular investigation into heat may suffer more fundamentally from some rather pre-modern misconceptions about the nature of the world as we understand it today. But this error would not necessarily be an error with the methodology itself.
3. Many of these examples are borrowed from Copi and Cohen (1994, pp. 574–588).

REFERENCES

Bauer, H. H. (1994). *Scientific literacy and the myth of the scientific method*. Urbana and Chicago, IL: University of Illinois Press.

Carnap, R. (1966). Induction and statistical probability. In M. Gardner (Ed.), *Philosophical foundations of physics: An introduction to the philosophy of science* (pp. 19–28). New York, NY, and London, UK: Basic Books.

Copi, I. M., & Cohen, C. (1994). *Introduction to logic*. New York, NY: MacMillan.

Hempel, C. G. (2000). The role of induction in scientific inquiry. In T. Schick (Ed.), *Readings in the philosophy of science: From positivism to postmodernism* (pp. 41–49). Mountain View, CA: Mayfield Publishing. (Original work published 1966)

Hume, D. (1974). An enquiry concerning human understanding. In J. Locke, G. Berkeley, & D. Hume (Authors), *The empiricists* (pp. 307–430). Garden City, NY: Anchor Books. (Original work published 1748)

Kuhn, T. S. (1970). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 1–23). Cambridge, UK: Cambridge University Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, UK: Cambridge University Press.

Quine, W. V. O. (2000). Two dogmas of empiricism. In J. McErlean (Ed.), *Philosophies of science: From foundations to contemporary issues* (pp. 115–128). Belmont, CA: Wadsworth. (Original work published 1951)

Rosenberg, A. (2005). *Philosophy of science: A contemporary introduction*. London, UK, and New York, NY: Routledge.

Steps of the Scientific Method. (2009). Retrieved from http://www.sciencebuddies.org/mentoring/project_scientific_method.shtml