CHAPTER 1

Evidence

Scientists and philosophers of science often emphasize that science is a fallible enterprise. The evidence that scientists have for their theories does not render those theories certain. This point about *evidence* is often represented by citing a fact about *logic*: The evidence we have at hand does not deductively entail that our theories must be true. In a *deductively valid argument*, the conclusion must be true if the premises are. Consider the following old saw:

All human beings are mortal.
Socrates is a human being.
―――――――――――――――
Socrates is mortal.

If the premises are true, you cannot go wrong in believing the conclusion. The standard point about science's fallibility is that the relationship of evidence to theory is *not* like this. The correctness of this point is most obvious when the theories in question are far more *general* than the evidence we can bring to bear on them. For example, theories in physics such as the general theory of relativity and quantum mechanics make claims about what is true at *all* places and *all* times in the entire universe. Our observations, however, are limited to a very small portion of that immense totality. What happens here and now (and in the vicinity thereof) does not deductively entail what happens in distant places and at times remote from our own.

If the evidence that science assembles does not provide certainty about which theories are true, what, then, does the evidence tell us? It seems entirely natural to say that science uses the evidence at hand to say which theories are *probably* true. This statement leaves room for science to be fallible and for the scientific picture of the world to change when new evidence rolls in. As sensible as this position sounds, it is deeply controversial. The controversy I have in mind is not between science and nonscience; I do not mean that scientists view themselves as assessing how probable theories are while postmodernists and religious zealots debunk science and seek to undermine its authority. No, the controversy I have in mind is alive *within* science. For the past seventy years, there has been a dispute in the foundations of statistics between Bayesians and frequentists. They disagree about many issues, but perhaps their most basic disagreement concerns whether science is in a position to judge which theories are probably true. Bayesians think that the answer is *yes* while frequentists emphatically disagree. This controversy is not confined to a question that statisticians and philosophers of science address; scientists use the methods that statisticians make available, and so scientists in all fields must choose which model of scientific reasoning they will adopt.

The debate between Bayesians and frequentists has come to resemble the trench warfare of World War I. Both sides have dug in well; they have their standard arguments, which they lob like grenades across the no-man's-land that divides the two armies. The arguments have become familiar and so have the responses. Neither side views the situation as a stalemate, since each regards its own arguments as compelling. And yet the warfare continues. Fortunately, the debate has not brought science to a standstill, since scientists frequently find themselves in the convenient situation of not having to care which of the two approaches they should use. Often, when a Bayesian and a frequentist consider a biological theory in the light of a body of evidence, they both give the theory high marks. This allows biologists to walk away happy; they've got their answer to the biological question of interest and don't need to worry whether Bayesianism or frequentism is the better statistical philosophy. Biologists care about making discoveries about *organisms*; the *nature of reasoning* is not their subject, and they are usually content to leave such "philosophical" disputes for statisticians and philosophers to ponder. Scientists are *consumers* of statistical methods, and their attitude towards methodology often resembles the attitude that most of us have towards consumer products like cars and computers. We read *Consumer Reports* and other magazines to get expert advice on what to buy, but we rarely delve deeply into what makes cars and computers tick. Empirical scientists often use statisticians, and the "canned" statistical packages they provide, in the same way that consumers use *Consumer Reports*. This is why the trench warfare just described is not something in which most biologists feel themselves to be engulfed. They live, or try to live, in neutral Switzerland; the Battle of the Marne (they hope) involves others, far from home.

This book is about the concept of evidence as it applies in evolutionary biology; the present chapter concerns general issues about evidence that will be relevant in subsequent chapters. I do not aim here to provide anything like a complete treatment of the debate between Bayesianism and frequentism, nor is my aim to end the trench warfare that has persisted for so long. Rather, I hope to help the reader to understand what the shooting has been about. I intend to start at the beginning, to not use jargon, and to make the main points clear by way of simple examples. There are depths that I will not attempt to plumb. Even so, my treatment will not be neutral; in fact, it is apt to irritate both of the entrenched armies. I will argue that Bayesianism makes excellent sense for many scientific inferences. However, I do agree with frequentists that applying Bayesian methods in other contexts is highly problematic. But, unlike many frequentists, I do not want to throw out the Bayesian baby with the bathwater. I also will argue that some standard frequentist ideas are flawed but that others are more promising. With respect to frequentism as well, I feel the need to pick and choose. My approach will be "eclectic"; no single unified account of all scientific inference will be defended here, much as I would like there to be a grand unified theory.

One further comment before we begin: I have contrasted Bayesianism and frequentism and will return to this dichotomy in what follows. However, there are different varieties of Bayesianism, and the same is true of frequentism. In addition, there is a third alternative, likelihoodism (though frequentists often see Bayesianism and likelihoodism as two sides of the same deplorable coin). We will separate these inferential philosophies more carefully in what follows. But for now we begin with a stark contrast: Bayesians attempt to assess how probable different scientific theories are, or, more modestly, they try to say which theories are more probable and which are less. Frequentists hold that this is not what the game of science is about. But what do frequentists regard as an attainable goal? Hold that question in mind; we will return to it.

1.1 Royall's three questions

The statistician Richard Royall begins his excellent book on the concept of evidence (Royall 1997: 4) by distinguishing three questions:

(1) What does the present evidence say?
(2) What should you believe?
(3) What should you do?

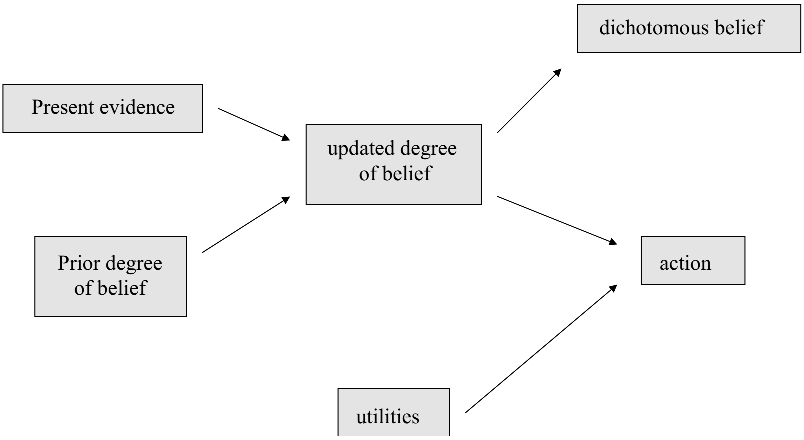

Figure 1.1 Present evidence and its downstream consequences.

If you are rational, you form your beliefs by consulting the evidence you have just gained, and when you decide what to do (which actions to perform), you should take account of what you believe. But answering question (2) requires more than an answer to (1), and answering question (3) requires more than an answer to (2). The extra elements needed are depicted in Figure 1.1.

Suppose you are a physician and you are talking to the patient in your office about the result of his tuberculosis test. The report from the lab says "positive." This is your present evidence. Should you conclude that the patient has tuberculosis? You want to take the lab report into account, but you have other information besides. For example, you previously had conducted a physical exam. Before you looked at the test report, you had some opinion about whether your patient has tuberculosis. The lab report may modify how certain you are about this. You update your degree of belief by integrating the new evidence with your prior information. This may lead you say to him "your probability of tuberculosis is 0.999."

If your patient is a philosopher who enjoys perverse conversation, he may reply, "but tell me, doctor, do I have tuberculosis, or not?" He doesn't want to know how *probable* it is that he has tuberculosis; he wants to know *whether* he has the disease – *yes or no*. This raises the question of whether a proposition's having a probability of 0.999 suffices for one to believe it, where belief is conceptualized as a dichotomous category: Either you believe the proposition or you do not. It may seem that a high degree of belief suffices for believing a proposition (even if it does not suffice for being certain that the proposition is true), but there are complications. Consider Kyburg's (1970) lottery paradox. Suppose 1,000 lottery tickets are sold and the lottery is fair. *Fair* means that one ticket will win and each has the same chance of winning. If high probability suffices for belief, you are entitled to believe that ticket no. 1 will not win, since the probability of ticket 1's not winning is $\frac{999}{1000}$. The same is true of ticket no. 2; you should believe that it won't win. And so on, for each of the 1,000 tickets. But if you put these 1,000 beliefs (each of the form *ticket i will not win*) together with the rest of what you believe, your beliefs have become contradictory: You believe that some ticket will win (since you believe the lottery is fair), and you have just accepted the proposition that no ticket will win. Kyburg's solution to this puzzle is to say that acceptance does not obey a rule of conjunction; you can accept *A* and accept *B* without having to accept the conjunction *A&B*.[1] This may be the best one can do for the concept of dichotomous belief, but it raises the question of whether we really need such a concept. After all, our everyday thought is littered with dichotomies that, upon reflection, seem to be crudely grafted to an underlying continuum. For example, we speak of people being *bald*, but we know that there is no threshold number of hairs that marks the boundary.[2] We are happy to abandon these crude categories when we need to, but we return to them when they are convenient and harmless.

If it makes sense to talk about rational acceptance and rational rejection, those concepts must bear the following relation to the concept of evidence:

> If learning that *E* is true justifies you in *rejecting* (i.e., disbelieving) the proposition *P*, and you were not justified in rejecting *P* before you gained this information, then *E* must be evidence *against P*.
>
> If learning that *E* is true justifies you in *accepting* (i.e., believing) the proposition *P*, and you were not justified in accepting *P* before you gained this information, then *E* must be evidence *for P*.

A theory of rational acceptance and rejection must provide more than this modest principle, which may seem like a mere crumb, hardly worth

[1] See Kaplan (1996) for a theory of rational acceptance that, unlike Kyburg's, obeys the conjunction principle.

[2] I say we "know" this, but Williamson (1994) and Sorenson (2001) have argued that in each use of a vague term, there is a cutoff, even if speakers are not aware of what it is. Their position is counterintuitive, but it cannot be dismissed without attending to their arguments (which we won't do here).

mentioning at all. But, in fact, it *is* worth stating, since later in this chapter it will do some important philosophical work.[3]

Even if this modest principle linking evidence and rational acceptance seems obvious, there is an old philosophical reason for pausing to ponder it. In the seventeenth century, Blaise Pascal sketched an argument that came to be called *Pascal's wager*. Earlier proofs of the existence of God had tried to demonstrate that there is evidence that God exists; Pascal endeavored to show that one ought to believe in God even if all the evidence one has is evidence *against*. The rough idea is this: If there is a God, you'll go to Heaven if you're a believer and go to Hell if you're not; on the other hand, if there is no God, it won't much affect your well-being whether or not you believe. Pascal wrote when probability theory was just starting to take its modern mathematical form, and his argument is a nice illustration of ideas that came to be assembled in *decision theory*. Though there is room to dispute the details of this argument (on which see Mougin and Sober 1994), the wager is of interest here because it appears to challenge the "modest" principle just enunciated. The wager purports to provide a reason for accepting the proposition that God exists even though it does not cite any evidence that there is a God. It is easy to think of nontheological arguments that pose the same challenge. Suppose I promise to give you $1,000,000 if you can get yourself to believe that the President is now juggling candy bars. If I am trustworthy, I have given you a reason to believe the proposition though I have not provided any evidence that it is true.

Commentators on Pascal's wager often distinguish two types of rational acceptance. The *act of accepting* a proposition can make good prudential sense, but that does not mean that *the proposition accepted* is well supported by evidence. When acceptance is driven by the costs and benefits that attach to the act of believing, I'll call this "prudential acceptance." When it is driven by the bearing of evidence on the proposition believed, I'll use the term "evidential acceptance." The modest principle linking evidence and "acceptance" really pertains to *evidential* acceptance. The principle, modified in this way, is true; in fact, it may even be true *by definition*. However, this does not settle whether it is ever permissible to

[3] It is interesting that the concept of evidence relates pairs of propositions to each other, while the concepts of acceptance and rejection relate propositions to persons. Smoke is evidence for fire, regardless of whether any agent takes this fact to heart. However, rational acceptance (or rejection) means that a person is justified in accepting (or rejecting) some proposition. The present disciplinary divide between philosophers of science and epistemologists coincides to a considerable degree with this distinction between questions concerning how propositions are related to each other and questions concerning how propositions are related to persons.

indulge in prudential acceptance. William James (1897) defends the right to believe when the evidence is silent in his essay "The Will to Believe." W. K. Clifford (1999) replies, in "The Ethics of Belief," that it is always wrong "to believe upon insufficient evidence." I will not try to adjudicate between these two positions. Suffice it to say that the modest principle stated earlier is binding on those who commit to having evidence control what they believe.

It may seem a long jump from Pascal's seventeenth-century theology to the hard edges of twentieth-century statistics, but Pascal's concept of prudential acceptance lives on in frequentism. The following remark by Neyman and Pearson (1933: 291) has often been quoted:

> No test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood of [an] hypothesis [. . .] But we may look at the purpose of tests from another viewpoint. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behavior with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong.

Neyman and Pearson think of acceptance and rejection as *behaviors*, which should be regulated by prudential considerations, not by "evidence," which, for them, is a will o' the wisp. The prudential considerations they have in mind do not involve going to Heaven or Hell, but rather pertain to having true beliefs or false ones. There is no such thing as allowing "evidence" to regulate what we believe. Rather, we must embrace a policy and stick to it. If we do so, we can be certain (or, at least, it is overwhelmingly probable) that the percentage of false beliefs we accumulate over the long run will be held below some predesignated minimum. Not that present-day frequentists are all so dismissive of the concept of evidence (§1.4). But frequentists, early and late, have often embraced the idea of *prudential* belief.

Let us return to Figure 1.1. Suppose you, the physician, are 99.9 percent certain that your patient has tuberculosis, this degree of belief being based on the present tuberculosis test result and on other information you had from before. The thing to notice next is that your degree of belief does not, by itself, dictate what you should *say* or *do*. Should you tell your patient what you think? Should you remain silent? Should you lie? Should you hand him the pink pills you have in your desk? A rational decision about what to do requires more than the evidence you have and more than the degree of belief you have; a choice of action requires the input of values (which economists call *utilities*).

1.2 THE ABCs OF BAYESIANISM

Bayesianism is an answer to Royall's question (2): What should you believe? Bayesianism refines this question, substituting the concept of degree of belief for the dichotomous concept of believing or not believing a proposition. In our running example, Bayesianism addresses the question of how certain you should be that your patient has tuberculosis, given that his tuberculosis test came back positive.

Bayes' theorem

Bayesianism is based on Bayes' theorem, but the two are different. Bayes' theorem is a result in mathematics.[4] It is called a theorem because it is derivable from the axioms of probability theory (in fact, from a standard definition of conditional probability). As a piece of mathematics, the theorem is not controversial. Bayesianism, on the other hand, is a philosophical theory – it is an epistemology. It proposes that the mathematics of probability theory can be put to work in a certain way to explicate various concepts connected with issues about evidence, inference, and rationality.

Here is the rough idea of how Bayesianism uses Bayes' theorem: Before you make an observation, you assign a probability to the hypothesis H; this probability may be high, medium, or low (all probabilities by definition must be between 0 and 1, inclusive). After you make the observation, thereby learning that some observation statement O is true, you update the probability you assigned to H to take account of what you just learned. The probability that H has before the observation is called its *prior probability*; it is represented by $Pr(H)$. The word "prior" just means *before*; it doesn't mean that you know its value a priori (i.e., without any empirical input at all). The probability that H has in the light of the evidence O is called H's *posterior probability*; it is represented by the conditional probability $Pr(H \mid O)$; read this as "the probability of H, given O." Bayes' theorem shows how the prior and the posterior probability are related.

Now for the derivation of the theorem. Forget for just a moment that H means hypothesis and O means observation. Just regard them as any two

[4] A special case of the theorem was derived by Thomas Bayes and was published posthumously in the *Proceedings of the Royal Society* for 1764. Bayes' derivation was laborious and not fully general, very unlike the now-standard streamlined derivation I'll describe here.

propositions. Kolmogorov's (1950) definition of conditional probability is this:

$$Pr(H \mid O) = \frac{Pr(H \ \& \ O)}{Pr(O)}.$$

The definition is intuitive. For example, what is the probability that a card drawn at random from a standard deck is a heart, given that it is red? According to the Kolmogorov definition, this conditional probability has the same value as the ratio $Pr(\text{heart} \ \& \ \text{red})/Pr(\text{red})$. The denominator has a value of $\frac{1}{2}$. The proposition in the numerator, *heart & red*, is equivalent to *heart*, so the value for the numerator is $\frac{1}{4}$. Hence, the conditional probability has a value of $\frac{1}{2}$. By switching *H*s and *O*s with each other in the Kolmogorov definition, you can see that it also is true that

$$Pr(O \mid H) = \frac{Pr(O \ \& \ H)}{Pr(H)}.$$

This means that the probability of the conjunction *H&O* can be expressed in two different ways:

$$Pr(H \ \& \ O) = Pr(H \mid O) \ Pr(O) = Pr(O \mid H)Pr(H).$$

From the second equality in the previous line, we obtain

$$\text{Bayes' theorem: } Pr(H \mid O) = \frac{Pr(O \mid H)Pr(H)}{Pr(O)}.$$

Here is some more terminology. I've already mentioned the *posterior probability* and the *prior probability* that appear in Bayes' theorem, but two other quantities are also mentioned. $Pr(O)$ is the *unconditional probability of the observations*. And R. A. Fisher dubbed $Pr(O \mid H)$ the *likelihood of H*. Because Fisher's terminology has become standard in statistics, I will use it here. However, this terminology is confusing, since in ordinary English, "likely" and "probably" are synonymous. So, beware! You need to remember that "likelihood" is a technical term. The likelihood of *H*, $Pr(O \mid H)$, and the posterior probability of *H*, $Pr(H \mid O)$, are different quantities and they can have different values. The likelihood of H is the probability that H confers on O, not the probability that O confers on H. Suppose you hear a noise coming from the attic of your house. You consider the hypothesis that there are gremlins up there bowling. The likelihood of this hypothesis is very high, since if there are gremlins bowling in the attic, there probably will be noise. But surely you don't think that the noise makes it very probable that there are gremlins up there bowling. In this example, $Pr(O \mid H)$ is high and $Pr(H \mid O)$ is low. The gremlin hypothesis has a high likelihood (in the technical sense) but a low probability.

Let me add two more details that underscore the distinction between H's probability and its likelihood.

$$Pr(H) + Pr(notH) = 1$$

and

$$Pr(H \mid O) + Pr(notH \mid O) = 1$$

as well. The probability of a proposition and the probability of its negation sum to one; this is true for prior and also for posterior probabilities. But likelihoods need not sum to one; $Pr(O \mid H) + Pr(O \mid notH)$ can be less than 1, or more. Suppose you observe that Sue is a millionaire and wonder whether she won her wealth in last week's lottery. Your observation is very improbable under the hypothesis that she bought a ticket in the lottery and also under the hypothesis that she did not. To summarize this point: If you know the probability of H, you thereby know the probability of *notH*; but knowing the likelihood of H leaves the likelihood of *notH* completely open.

Another difference between likelihoods and probabilities concerns the difference between logically stronger and logically weaker hypotheses. Consider the following two hypotheses about the next card you'll be dealt from a standard deck:

H_1 = It's a heart.

H_2 = It's the Ace of Hearts.

The hypothesis H_2 is *logically stronger* than H_1; this means that H_2 entails H_1, but not conversely. Suppose the dealer is careless and you catch a glimpse of the card before it is dealt; you observe O = the card is red. Notice that H_1 has the higher posterior probability; $Pr(H_1 \mid O) = \frac{1}{2}$ while