Notice that, although the *likelihood* of a defective product is twice as great for the old machine (0.02) as for the new (0.01), the *posterior probability* that a defective frisbee was produced by the new machine (2/3) is twice as great as the probability that it was produced by the old one (1/3).

In Section 2.9 we return to the problem of assigning probabilities to hypotheses, which is the main subject of this chapter.

2.8 THE MEANING OF PROBABILITY

In the preceding section we discussed the notion of probability in a formal manner. That is, we introduced a symbol, ''*Pr*(|),'' to stand for probability, and we laid down some formal rules governing the use of that symbol. We illustrated the rules with concrete examples, to give an intuitive feel for them, but we never tried to say what the word ''probability'' or the symbol ''*Pr*'' means. That is the task of this section.

As we discuss various suggested meanings of this term, it is important to recall that we laid down certain basic rules (axioms). If a proposed definition of ''probability'' satisfies the basic rules—and, consequently, the derived rules, since they are deduced from the basic rules—we say that the suggested definition provides an *admissible interpretation* of the probability concept. If a proposed interpretation violates those rules, we consider it a serious drawback.

1. The classical interpretation. One famous attempt to define the concept of probability was given by the philosopher-scientist Pierre Simon de Laplace ([1814] 1951). It is known as the *classical interpretation.* According to this definition, *the probability of an outcome is the ratio of favorable cases to the number of equally possible cases.* Consider a simple example. A standard die (singular of ''dice'') has six faces numbered 1–6. When it is tossed in the standard way there are six *possible* outcomes. If we want to know the probability of getting a 6, the answer is 1/6, for only one possible outcome is *favorable.* The probability of getting an even number is 3/6, for three of the possible outcomes (2, 4, 6) are favorable.

Laplace was fully aware of a fundamental problem with this definition. The definition refers not just to possible outcomes, but to *equally possible* outcomes. Consider another example. Suppose two standard coins are flipped simultaneously. What is the probability of getting two heads? Someone might say it is 1/3, for there are three possible outcomes, two heads, one head and one tail, or two tails. We see immediately that this answer is incorrect, for these possible outcomes are not equally possible. That is because one head and one tail can occur in two different ways—head on coin #1 and tail on coin #2, or tail on coin #1 and head on coin #2. Hence, we should say that there are four *equally possible* cases, so the probability of two heads is 1/4.

In order to clarify his definition Laplace needed to say what is meant by ''equally possible,'' and he endeavored to do so by offering the famous *principle of indifference.* According to this principle, two outcomes are equally possible—we might as well say ''equally probable''—*if we have no reason to prefer one to the other.*

> Compare the coin example with the following from modern physics. Suppose you have two helium-4 atoms in a box. Each one has a fifty-fifty chance of being in the left-hand side of the box at any given time. What is the probability of both atoms being in the left-hand side at a particular time? The answer is 1/3. Since the two atoms are in principle indistinguishable—unlike the coins, which are obviously distinguishable—we cannot regard atom #1 in the left-hand side and atom #2 in the right-hand side as a case distinct from atom #1 in the right-hand side and atom #2 in the left-hand side. Indeed, it does not even make sense to talk about atom #1 and atom #2 since we have no way, even in principle, of telling which is which.

Suppose, for example, that we examine a coin very carefully and find that it is perfectly symmetrical. Any reason one might give to suppose it will come up heads can be matched by an equally good reason to suppose it will land tails up. We say that the two sides are equally possible, and we conclude that the probability of heads is 1/2. If, however, we toss the coin a large number of times and find that it lands heads up in about 3/4 of all tosses and tails up in about 1/4 of all tosses, we *do have* good reason to prefer one outcome to the other, so we would *not* declare them equally possible. The basic idea behind the principle of indifference is this: when we have no *reason* to consider one outcome more probable than another, we should not *arbitrarily* choose one outcome to favor over another. This seems like a sound principle of probabilistic reasoning.

There is, however, a profound difficulty connected with the principle of indifference; its use can lead to outright inconsistency. The problem is that it can be applied in different ways to the same situation, yielding incompatible values for a particular probability. Again, consider an example, namely, the case of Joe, the sloppy bartender. When a customer orders a 3:1 martini (3 parts of gin to 1 part of dry vermouth), Joe may mix anything from a 2:1 to a 4:1 martini, and there is no further information to tell us where in that range the mix may lie. According to the principle of indifference, then, we may say that there is a fifty-fifty chance that the mix will be between 2:1 and 3:1, and an equal chance that it will be between 3:1 and 4:1. Fair enough. But there is another way to look at the same situation. A 2:1 martini contains 1/3 vermouth, and a 4:1 martini contains 1/5 vermouth. Since we have no further information about the proportion of vermouth we can apply the principle of indifference once more. Since 1/3 = 20/60 and 1/5 = 12/60, we can say that there is a fifty-fifty chance that the proportion of vermouth is between 20/60 and 16/60 and an equal chance that it is between 16/60 and 12/60. So far, so good?

Unfortunately, no. We have just contradicted ourselves. A 3:1 martini contains 25 percent vermouth, which is equal to 15/60, *not* 16/60. The principle of indifference has told us *both* that there is a fifty-fifty chance that the proportion of vermouth is between 20/60 and 16/60, *and also* that there is a fifty-fifty chance that it is between

20/60 and 15/60. The situation is shown graphically in Figure 2.2. As the graph shows, the same result occurs for those who prefer their martinis drier; the numbers are, however, not as easy to handle.

We must recall, at this point, our first axiom, which states, in part, that the probability of a given outcome under specified conditions is a *unique* real number. As we have just seen, the classical interpretation of probability does not furnish unique results; we have just found two different probabilities for the same outcome. Thus, it turns out, the classical interpretation is *not* an admissible interpretation of probability.

You might be tempted to think the case of the sloppy bartender is an isolated and inconsequential fictitious example. Nothing could be farther from the truth. This example illustrates a broad range of cases in which the principle of indifference leads to contradiction. The source of the difficulty lies in the fact that we have two quantities—the ratio of gin to vermouth and the proportion of vermouth—that are interdefinable; if you know one you can calculate the other. However, as Figure 2.2 clearly shows, the definitional relation is not linear; the graph is not a straight line. We can state generally: Whenever there is a nonlinear definitional relationship between two quantities, the principle of indifference can lead to a similar contradiction. To convince yourself of this point, work out the details of another example. Suppose there is a square piece of metal inside of a closed box. You cannot see it. But you are told that its area is somewhere between 1 square inch and 4 square inches, but nothing else is known about the area. First apply the principle of indifference to the area of the square, and then apply it to the length of the side which is, of course, directly

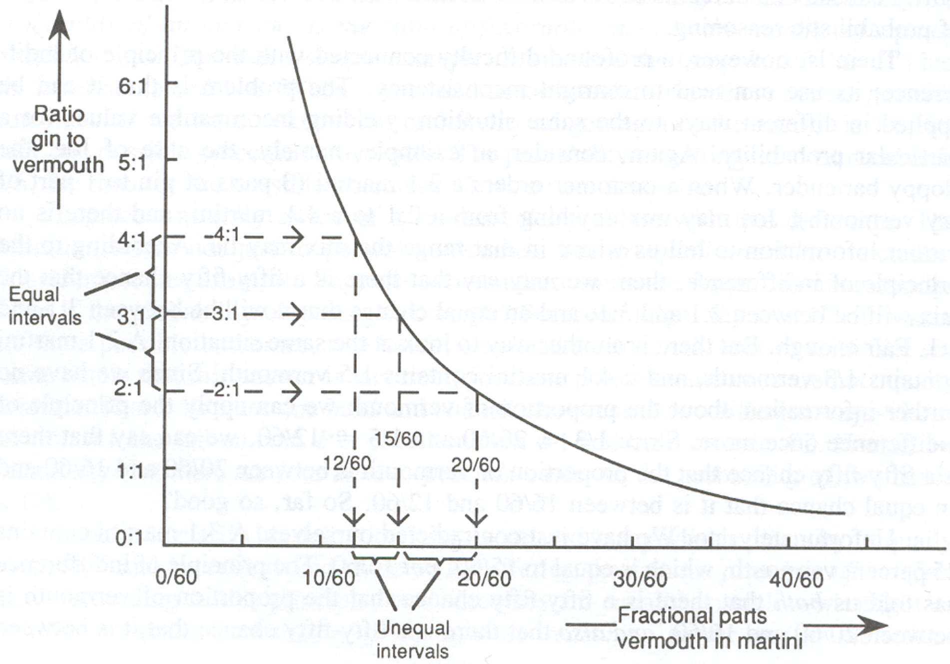

Figure 2.2

ascertainable from the area. (For another example, involving a car on a racetrack, see Salmon 1967, 66–67.)

Although the classical interpretation fails to provide a satisfactory basic definition of the probability concept, that does not mean that the idea of the ratio of favorable to equiprobable possible outcomes is useless. The trouble lies with the principle of indifference, and its aim of transforming ignorance of probabilities into values of probabilities. However, in situations in which we have positive knowledge that we *are* dealing with alternatives that have equal probabilities, the strategy of counting equiprobable favorable cases and forming the ratio of favorable to equiprobable possible cases is often handy for facilitating computations.

2. The frequency interpretation. The frequency interpretation has a venerable history, going all the way back to Aristotle (4th century B.C.), who said that the probable is that which happens often. It was first elaborated with precision and in detail by the English logician John Venn (1866, [1888] 1962). The basic idea is easily illustrated. Consider an ordinary coin that is being flipped in the standard way. As it is flipped repeatedly a sequence of outcomes is generated:

H T H T T T H H T T H T T T T H T H T T T H H H H . . .[10]

We can associate with this sequence of results a sequence of *relative frequencies*—that is, the proportion of tosses that have resulted in heads up to a given point in the sequence—as follows:

1/1, 1/2, 2/3, 2/4, 2/5, 2/6, 3/7, 4/8, 4/9, 4/10, 5/11, 5/12, 5/13/ 5/14, 5/15, 6/16, 6/17, 7/18, 7/19, 7/20, 7/21, 8/22, 9/23, 10/24, 11/25, . . .

The denominator in each fraction represents the number of tosses made up to that point; the numerator represents the number of heads up to that point. We could, of course, continue flipping the coin, recording the results, and tabulating the associated relative frequencies. We are reasonably convinced that this coin is fair and that it was flipped in an unbiased manner. Thus, we believe that the probability of heads is 1/2. If that belief is correct, then, as the number of tosses increases, the relative frequencies will become and remain close to 1/2. The situation is shown graphically in Figure 2.3. There is no particular number of tosses at which the fraction of heads is and remains precisely 1/2; indeed, in an odd number of tosses the ratio cannot possibly equal 1/2. Moreover, if, at some point in the sequence, the relative frequency does equal precisely 1/2, it will necessarily differ from that value on the next flip. Instead of saying that the relative frequency must equal 1/2 in any particular number of throws, we say that it approaches 1/2 *in the long run*.

Although we know that no coin can ever be flipped an infinite number of times, it is useful, as a mathematical idealization, to think in terms of a *potentially infinite* sequence of tosses. That is, we imagine that, no matter how many throws have been

[10] These are the results of 25 flips made in an actual trial by the authors.

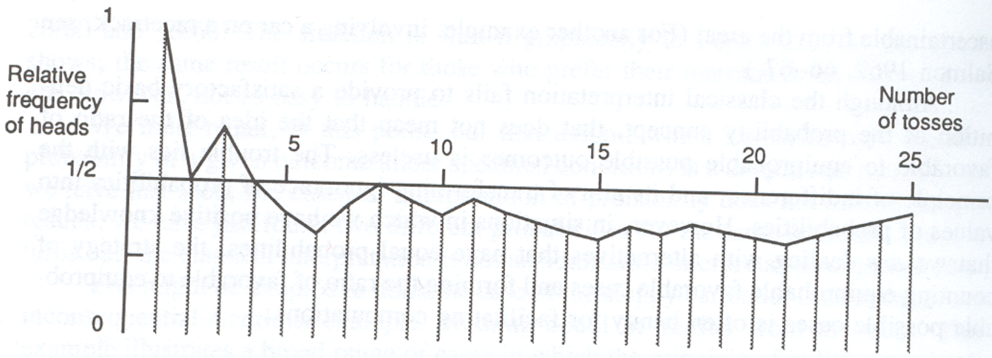

Figure 2.3

made, it is still possible to make more; that is, there is no particular finite number N at which point the sequence of tosses is considered complete. Then we can say that the *limit of the sequence* of relative frequencies equals the probability; this is the meaning of the statement that the *probability* of a particular sort of occurrence is, by definition, its long run relative frequency.

What is the meaning of the phrase "limit of the relative frequency"? Let $f_1, f_2, f_3, \ldots$ be the successive terms of the sequence of relative frequencies. In the example above, $f_1 = 1, f_2 = 1/2, f_3 = 2/3$, and so on. Suppose that p is the limit of the relative frequency. This means that the values of f_n become and remain arbitrarily close to p as n becomes larger and larger. More precisely, let δ be any small number greater than 0. Then, there exists some finite integer N such that, for any $n > N$, f_n does not differ from p by more than δ.

Many objections have been lodged against the frequency interpretation of probability. One of the least significant is that mentioned above, namely, the finitude of all actual sequences of events, at least within the scope of human experience. The reason this does not carry much weight is the fact that science is full of similar sorts of idealizations. In applying geometry to the physical world we deal with ideal straight lines and perfect circles. In using the infinitesimal calculus we assume that certain quantities—such as electric charge—can vary continuously, when we know that they are actually discrete. Such practices carry no danger provided we are clearly aware of the idealizations we are using. Dealing with infinite sequences is technically easier than dealing with finite sequences having huge numbers of members.

A much more serious problem arises when we ask how we are supposed to ascertain the values of these limiting frequencies. It seems that we observe some limited portion of such a sequence and then extrapolate on the basis of what has been observed. We may not want to judge the probability of heads for a certain coin on the basis of 25 flips, but we might well be willing to do so on the basis of several hundred. Nevertheless, there are several logical problems with this procedure. First, no matter how many flips we have observed, it is always *possible* for a long run of heads to occur that would raise the relative frequency of heads well above 1/2. Similarly, a long run of future tails could reduce the relative frequency far below 1/2.

Another way to see the same point is this. Suppose that, for each n, m/n is the fraction of heads to tosses as of the nth toss. Suppose also that f_n does have the limiting value p. Let a and b be any two fixed positive integers where $a \leq b$. If we add the constant a to every value of m and the constant b to every value of n, the resulting sequence $(m + a)/(n + b)$ will converge to the very same value p. That means that you could attach any sequence of b tosses, a of which are heads, to the beginning of your sequence, without changing the limiting value of the relative frequency. Moreover, you can chop off any finite number b of members, a of which are heads, from the beginning of your sequence without changing the limiting frequency p. As m and n get very large, the addition or subtraction of fixed numbers a and b has less and less effect on the value of the fraction. This seems to mean that *the observed relative frequency in any finite sample is irrelevant to the limiting frequency*. How, then, are we supposed to find out what these limiting frequencies—probabilities—are?

It would seem that things could not get much worse for the frequency interpretation of probability, but they do. For any sequence, such as our sequence of coin tosses, there is no guarantee that any limit of the relative frequency even exists. It is logically possible that long runs of heads followed by longer runs of tails followed by still longer runs of heads, and so on, might make the relative frequency of heads fluctuate between widely separated extremes throughout the infinite remainder of the sequence. If no limit exists *there is no such thing as the probability of a head when this coin is tossed*.

In spite of these difficulties, the frequency concept of probability seems to be used widely in the sciences. In Chapter 1, for instance, we mentioned the spontaneous decay of C^{14} atoms, commenting that the half-life is 5730 years. That is the rate at which atoms of this type have decayed in the past; we confidently predict that they will continue to do so. The relative frequency of disintegration of C^{14} atoms within 5730 years is 1/2. This type of example is of considerable interest to archaeologists, physicists, and geophysicists. In the biological sciences it has been noted, for example, that there is a very stable excess of human male births over human female births, and that is expected to continue. Social scientists note, however, that human females, on average, live longer than human males. This frequency is also extrapolated.

It is easy to prove that the frequency interpretation satisfies the axioms of probability laid down in the preceding section. This interpretation is, therefore, admissible. Its main difficulty lies in the area of *ascertainability*. How are we to establish values of probabilities of this sort? This question again raises Hume's problem of justification of induction.

A further problem remains. Probabilities of the frequency variety are used in two ways. On the one hand, they appear in statistical laws, such as the law of radioactive decay of unstable species of nuclei. On the other hand, they are often applied in making predictions of single events, or finite classes of events. Pollsters, for example, predict outcomes of single elections on the basis of interviews with samples of voters. If, however, probability is defined as a limiting frequency in a potentially infinite sequence of events, it does not seem to make any sense to talk about probabilities of single occurrences. The *problem of the single case* raises a problem about the *applicability* of the frequency interpretation of probability.

Before we leave the frequency interpretation a word of caution is in order. The frequency interpretation and the classical interpretations are completely different from one another, and they should not be confused. When the classical interpretation refers to possible outcomes and favorable outcomes it is referring to types or classes of events—for example, the class of all cases in which heads comes up is *one* possible outcome; the class of cases in which tails comes up is *one* other possible outcome. In this example there are only two possible outcomes. These *classes—not their members*—are what you count for purposes of the classical interpretation. In the frequency interpretation, it is the *members* of these classes that are counted. If the coin is tossed a large number of times there are many heads and many tails. In the frequency interpretation, the numbers of items of which ratios are formed keep changing as the number of individual events increases. In the classical interpretation, the probability does not depend in any way on how many heads or tails actually occur.

3. The propensity interpretation. The propensity interpretation is a relatively recent innovation in the theory of probability. Although suggested earlier, particularly by Charles Saunders Peirce, it was first clearly articulated by Popper (1957b, 1960). It was introduced specifically to deal with the problem of the single case.

The sort of situation Popper originally envisaged was a potentially infinite sequence of tosses of a loaded die that was biased in such a way that side 6 had a probability of 1/4. The limiting frequency of 6 in this sequence is, of course, 1/4. Suppose, however, that three of the tosses were *not* made with the biased die, but rather with a fair die. Whatever the outcomes of these three throws, they would have no effect on the limiting frequency. Nevertheless, Popper maintained, we surely want to say that the probability of 6 on those three tosses was 1/6—*not* 1/4. Popper argued that the appropriate way to deal with such cases is to associate the probability with the *chance setup* that produces the outcome, rather than to define it in terms of the sequence of outcomes themselves. Thus, he claims, each time the fair die is thrown, the mechanism—consisting of the die and the thrower—has a causal tendency or *propensity* of 1/6 to produce the outcome 6. Similarly, each time the loaded die is tossed, the mechanism has a propensity of 1/4 to produce the outcome 6.

Although this idea of propensity—probabilistic causal tendency—is important and valuable, it does not provide an admissible interpretation of the probability calculus. This can easily be seen in terms of the case of the frisbee factory introduced in the preceding section. That example, we recall, consisted of two machines, each of which had a certain propensity or tendency to produce defective frisbees. For the new machine the propensity was 0.01; for the old machine it was 0.02. Using the rule of total probability we calculated the propensity of the factory to produce faulty frisbees; it was 0.012. So far, so good.

The problem arises in connection with Bayes's rule. Having picked a defective frisbee at random from the day's production, we asked for the probability that it was produced by the new machine; the answer was 2/3. This is a perfectly legitimate probability, but it cannot be construed as a propensity. It makes no sense to say that this frisbee has a propensity of 2/3 to have been produced by the new machine. Either it was produced by the new machine or by the old. It does not have a tendency of 1/3 to have been produced by the old machine and a tendency of 2/3 to have been produced by the new one. The basic point is that causes precede their effects and causes produce their effects, even if the causal relationship has probabilistic aspects. We can speak meaningfully of the causal tendency of a machine to produce a faulty product. Effects do not produce their causes. It does *not* make sense to talk about the causal tendency of the effect to have been produced by one cause or another.

Bayes's rule enables us to compute what are sometimes called *inverse probabilities.* Whereas the rule of total probability enables us to calculate the *forward* probability of an effect, given suitable information about antecedent causal factors, Bayes's rule allows us to compute the inverse probability that a given effect was produced by a particular cause. These inverse probabilities are an integral part of the mathematical calculus of probability, but no propensities correspond to them. For this reason the propensity interpretation is not an admissible interpretation of the probability calculus.

4. The subjective interpretation. Both the frequency interpretation and the propensity interpretation are regarded by their proponents as types of *physical probabilities.* They are objective features of the real world. But probability seems to many philosophers and mathematicians to have a subjective side as well. This aspect has something to do with the degree of conviction with which an individual believes in one proposition or another. For instance, Mary Smith is sure that it will be cold in Montana next winter—that is, in some place in that state the temperature will fall below 50 degrees Fahrenheit between 21 December and 21 March. Her subjective probability for this event is extremely close to 1. Also, she disbelieves completely that Antarctica will be hot any time during its summer—that is, she is sure that the temperature will not rise above 100 degrees Fahrenheit between 21 December and 21 March. Her subjective probability for real heat in Antarctica in summer is very close to 0. She neither believes in rain in Pittsburgh tomorrow, nor disbelieves in rain in Pittsburgh tomorrow; her conviction for either one of these alternatives is just as strong as for the other. Her subjective probability for rain tomorrow in Pittsburgh is just about 1/2. As she runs through the various propositions in which she might believe or disbelieve she finds a range of degrees of conviction spanning the whole scale from 0 to 1. Other people will, of course, have different degrees of conviction in these same propositions.

It is easy to see immediately that subjective degrees of commitment do not provide an admissible interpretation of the probability calculus. Take a simple example. Many people believe that the probability of getting a 6 with a fair die is 1/6, and that the outcomes of successive tosses are independent of one another. They also believe that we have a fifty-fifty chance of getting 6 at least once in three throws. As we saw in the previous section, however, that probability is significantly below 1/2. Therefore, the preceding set of degrees of conviction violate the mathematical calculus of probability. Of course, not everyone makes that particular mistake, but extensive empirical research has shown that most of us do make various kinds of mistakes in dealing with probabilities. In general, a given individual's degrees of conviction fail to satisfy the mathematical calculus.

5. Personal probabilities. What if there were a person whose degrees of conviction did not violate the probability calculus? That person's subjective probabilities would constitute an admissible interpretation. Whether there actually is any such person, we can think of such an organization of our degrees of conviction as an ideal.

Compare this situation with deductive logic. One of its main functions is to help us avoid certain types of logical errors. Anyone who believes, for example, that all humans are mortal and Socrates is human, but that Socrates is immortal, is guilty of self-contradiction. Whoever wants to believe only what is true must try to avoid contradictions, for contradictions cannot possibly be true. In this example, among the three statements, "All humans are mortal," "Socrates is human," and "Socrates is immortal," at least one must be false. Logic does not tell us which statement is false, but it does tell us to make some change in our set of beliefs if we do not want to believe falsehoods. A person who avoids logical contradictions—inconsistencies—has a consistent set of beliefs.

A set of degrees of conviction that violate the calculus of probability is said to be *incoherent*. Anyone who holds a degree of conviction of 1/6 that a fair die, when tossed, will come up 6, and who also considers successive tosses independent (whose degree of conviction in 6 on the next toss is not affected by the outcome of previous tosses), and who is convinced to the degree 1/2 that 6 will come up at least once in three tosses, is being incoherent. So also is anyone who assigns two different values to the probability that a martini mixed by Joe, the sloppy bartender, is between 3:1 and 4:1.

A serious penalty results from being incoherent. A person who has an incoherent set of degrees of conviction is vulnerable to a *Dutch book*. A Dutch book is a set of bets such that, no matter what the outcome of the event on which the bets are made, the subject loses. Conside a very simple example. The negation rule of the probability calculus tells us that $Pr(B|A)$ and $Pr(\sim B|A)$ must add up to 1. Suppose someone has a degree of conviction of 2/3 that the next toss of a particular coin will result in heads, and also a degree of conviction of 2/3 that it will result in tails. This person should be willing to bet at odds of 2 to 1 that the coin will come up heads, and also at odds of 2 to 1 that it will come up tails. These bets constitute a Dutch book because, if the coin comes up heads the subject wins \$1 but loses \$2, and if it comes up tails the subject loses \$2 and wins \$1. Since these are the only possible outcomes, the subject loses \$1 no matter what happens.

It has been proved in general that a person is subject to a Dutch book if and only if that person holds an incoherent set of degrees of conviction. Thus, we can look at the probability calculus as a kind of system of logic—the logic of degrees of conviction. Conforming to the rules of the probability calculus enables us to avoid certain kinds of blunders in probabilistic reasoning, namely, the type of error that makes one subject to a Dutch book. In light of these considerations, *personal probabilities* have been defined as *coherent sets of degrees of conviction*. It follows immediately that personal probabilities constitute an admissible interpretation of the probability calculus, for they have been defined in just that way.

One of the major motivations of those who accept the personalist interpretation of probability lies in the use of Bayes's rule; indeed, those who adhere to personal probabilities are often called "Bayesians." To see why, let us take another look at Bayes's rule (Rule 9):

$$Pr(H|K.E) = \frac{Pr(H|K) \times Pr(E|K.H)}{Pr(H|K) \times Pr(E|K.H) + Pr(\sim H|K) \times Pr(E|K.\sim H)}$$

provided that $Pr(E|K) \neq 0$.

Consider the following simple example. Suppose that someone in the next room is flipping a penny, and that we receive a reliable report of the outcome after each toss. We cannot inspect the penny, but for some reason we suspect that it is a two-headed coin. To keep the arithmetic simple, let us assume that the coin is either two-headed or fair. Let K stand for our background knowledge and opinion, H for the hypothesis that the coin is two-headed, and E for the results of the flips. For any given individual $Pr(H|K)$—the *prior probability*—represents that person's antecedent degree of conviction that the coin is two-headed before any of the outcomes have been reported. Probability $Pr(E|K.H)$—one of the *likelihoods*—is the probability of the outcome reported to us, given that the coin being flipped is two-headed. If an outcome of tails is reported, that probability obviously equals zero, and the hypothesis H is refuted.

If one or more heads are reported, that probability clearly equals 1. The probability $Pr(\sim H|K)$ is the prior probability that the coin is not two-headed—that is, that it is fair. On pain of incoherence, this probability must equal $1 - Pr(H|K)$. The probability $Pr(E|\sim H.K)$ is also a likelihood; it is the probability of the outcomes reported to us given that the coin is not two-headed. The probability $Pr(H|K.E)$—the *posterior probability*—is the probability that the coin is two-headed given both our background knowledge and knowledge of the results of the tosses. That probability represents an assessment of the hypothesis in the light of the observational evidence (reported reliably to us).

Suppose that John's prior personal probability that the coin is two-headed is 1/100. The result of the first toss is reported, and it is a head. Using Bayes's rule, he computes the posterior probability as follows:

$$\frac{1/100 \times 1}{1/100 \times 1 + 99/100 \times 1/2} = 2/101 \approx 0.02.$$

After two heads the result would be

$$\frac{1/100 \times 1}{1/100 \times 1 + 99/100 \times 1/4} = 4/103 \approx 0.04.$$

After ten heads the result would be

$$\frac{1/100 \times 1}{1/100 \times 1 + 99/100 \times 1/1024} = 1024/1123 \approx 0.91.$$

Suppose Wes's personal prior probability, before any outcomes are known, is much higher than John's; Wes has a prior conviction of 1/2 that the coin is two-headed. After the first head, he makes the following calculation:

$$\frac{1/2 \times 1}{1/2 \times 1 + 1/2 \times 1/2} = 2/3 \approx 0.67.$$

After the second head, he has

$$\frac{1/2 \times 1}{1/2 \times 1 + 1/2 \times 1/4} = 4/5 = 0.80.$$

After ten heads, he has

$$\frac{1/2 \times 1}{1/2 \times 1 + 1/2 \times 1/1024} = 1024/1025 > 0.99.$$

These calculations show two things. First, they show how Bayes's rule can be used to ascertain the probability of a hypothesis *if we have values for the prior probabilities*. If we employ *personal probabilities* the prior probabilities become available. They are simply a person's degrees of conviction in the hypothesis prior to receipt of the observational evidence. In this kind of example the likelihoods can be calculated from assumptions we share concerning the behavior of fair and two-headed coins.

Second, these calculations illustrate a phenomenon known as *washing out of the priors* or *swamping of the priors*. Notice that we did two sets of calculations—one for John and one for Wes. We started with widely divergent degrees of conviction in the hypothesis; Wes's was 1/2 and John's was 1/100. As the evidence accumulated our degrees of conviction became closer and closer. After ten heads, Wes's degree of conviction is approximately 0.99 and John's is approximately 0.91. As more heads occur our agreement becomes even stronger. This illustrates a general feature of Bayes's rule. Suppose there are two people with differing prior probabilities—as far apart as you like provided neither has an extreme value of 0 or 1. Then, if they agree on the likelihoods and if they share the same observational evidence, their posterior probabilities will get closer and closer together as the evidence accumulates. The influence of the prior probabilities on the posterior probabilities decreases as more evidence becomes available. This phenomenon of washing out of the priors should help to ease the worry we might have about appealing to admittedly subjective degrees of conviction in our evaluations of scientific hypotheses.

Still, profound problems are associated with the personalistic interpretation of probability. The only restriction imposed by this interpretation on the values of probabilities is that they be coherent—that they satisfy the rules of mathematical probability. This is a very weak constraint. If we look at the rules of probability we note that, with a couple of trivial exceptions, the mathematical calculus of probability does not by itself furnish us with any values of probabilities. The exceptions are that a logically necessary proposition must have probability 1 and a contradiction must have probability 0. In all other cases, the rules of probability enable us to calculate some probability values from others. You plug in some probability values, turn the crank, and others come out. This means that there need be little contact between our personal probabilities and what goes on in the external world. For example, it is possible for a person to have a degree of conviction of 9/10 that the next toss of a coin will result in heads even though the coin has been tossed hundreds of times and has come up tails on the vast majority of these tosses. By suitably adjusting one's other probabilities one can have such personal probabilities as these without becoming incoherent. If our probabilities are to represent *reasonable* degrees of conviction some stronger restrictions surely appear to be needed.

6. The logical interpretation. One of the most ambitious twentieth-century attempts to deal with the problems of probability and confirmation was the construction of a theory of *logical probability* by Rudolf Carnap. Carnap was not the first to make efforts in that direction, but his was the most systematic and precise. In fact, Carnap maintained that there are two important and legitimate concepts of probability—relative frequencies and logical probabilities—but his main work was directed toward the latter. He referred to logical probability as *degree of confirmation*. Many philosophers refer to logical probability as *inductive probability*. The three terms are essentially interchangeable.

Carnap's program was straightforward in intent. He believed that it is possible to develop a formal inductive logic along much the same lines as formal deductive logic. In fact, he constructed a basic logical language in which both deductive and inductive relations would reside. In deductive logic, if a statement E entails another statement H, E supports H completely—if E is true H must also be true. In inductive logic, if E is evidence for a hypothesis H, E provides some sort of *partial support* for H; indeed, this type of partial support is often referred to as *partial entailment*.

The easiest way to understand what Carnap did is to work out the details of a simple and highly artificial example. Let us construct a language which deals with a universe containing only three entities, and each of these entities has or lacks one property. We let a, b, and c denote the three individuals, and we use F to designate the property. To make the example concrete, we can think of the individuals as three balls and the property as red. The notation Fa says that the first ball is red; $\sim Fa$ says that it is not red. We need a few other basic logical symbols. We use x and y as variables for individuals, and (x), which is known as the *universal quantifier*, means "for every x." The notation $(\exists x)$, which is known as the *existential quantifier*, means "there exists at least one x such that." A dot "." is used for the conjunction *and*; a wedge "v" for the disjunction *or*. That is about all of the logical equipment we will need.

The model universe we are discussing is a very simple place, and we can describe it completely; indeed, we can describe every *logically possible* state of this universe. Any such complete description of a possible state is a *state description*; there are eight:

1. $Fa.Fb.Fc$	5. $Fa.\sim Fb.\sim Fc$
2. $Fa.Fb.\sim Fc$	6. $\sim Fa.Fb.\sim Fc$
3. $Fa.\sim Fb.Fc$	7. $\sim Fa.\sim Fb.Fc$
4. $\sim Fa.Fb.Fc$	8. $\sim Fa.\sim Fb.\sim Fc$

Any consistent statement that we can form in this miniature language can be expressed by means of these state descriptions. For example, $(x)Fx$, which says that every ball is red, is equivalent to state description 1. The statement $(\exists x)Fx$, which says that at least one ball is red, is equivalent to the disjunction of state descriptions 1–7; that is, it says that either state description 1 or 2 or 3 or 4 or 5 or 6 or 7 is true. Fa is equivalent to the disjunction of state descriptions 1, 2, 3, and 5. $Fa.Fb$ is equivalent to the disjunction of state descriptions 1 and 2. If we agree to admit—just for the sake of convenience—that there can be disjunctions with only one term, we can say that every consistent statement is equivalent to some disjunction of state descriptions. The state descriptions in any such disjunction constitute the *range* of that statement. A contradictory statement is equivalent to the denial of all eight of the state descriptions. Its range is empty.

In the following discussion, H is any statement that is being taken as a hypothesis and E any statement that is being taken as evidence. In this discussion any consistent statement that can be formulated in our language can serve as a statement of evidence, and any statement—consistent or inconsistent—can serve as a hypothesis. Now, consider the hypothesis $(\exists x)Fx$ and evidence Fc. Clearly this evidence deductively entails this hypothesis; if the third ball is red at least one *must* be red. If we look at the ranges of this evidence and this hypothesis, we see that the range of Fc (state descriptions 1, 3, 4, 7) is entirely included in the range of $(\exists x)Fx$ (state descriptions 1–7). This situation always holds. If one statement entails another, the range of the first is included within the range of the second. This means that every possible state of the universe in which the first is true is a possible state of the universe in which the second is true. If two statements have identical ranges, they are logically equivalent, and each one entails the other. If two statements are logically incompatible with one another, their ranges do not overlap at all—that is, there is no possible state of the universe in which they can both be true. We see, then, that deductive relationships can be represented as relationships among the ranges of the statements involved.

Let us now turn to inductive relationships. Consider the hypothesis $(x)Fx$ and the evidence Fa. This evidence obviously does not entail the hypothesis, but it seems reasonable to suppose that it provides some degree of inductive support or confirmation. The range of the evidence (1, 2, 3, 5) is not completely included in the range of the hypothesis (1), but it does overlap that range—the two ranges have state description 1 in common. What we need is a way of expressing the idea of confirmation in terms of the overlapping of ranges. When we take any statement E as evidence, we are accepting it as true; in so doing we are ruling out all possible states of the universe that are incompatible with the evidence E. Having ruled out all of those, we want to know to what degree the possible states in which the evidence holds true are possible states in which the hypothesis also holds true. This can be expressed in the form of a ratio, range $(E.H)$/range (E), and this is the basic idea behind the concept of *degree of confirmation*.

Consider the range of $(x)Fx$; this hypothesis holds in one state description out of eight. If, however, we learn that Fa is true, we rule out four of the state descriptions, leaving only four as possibilities. Now the hypothesis holds in one out of four. If we now discover that Fb is also true, our combined evidence $Fa.Fb$ holds in only two state descriptions, and our hypothesis holds in one of the two. It looks reasonable to say that our hypothesis had a probability of 1/8 on the basis of no evidence, a probability of 1/4 on the basis of the first bit of evidence, and a probability of 1/2 on the two pieces of evidence. (This suggestion was offered by Wittgenstein 1922). But appearances are deceiving in this case.

If we were to adopt this suggestion as it stands, Carnap realized, we would rule out altogether the possibility of learning from experience; we would have no basis at all for predicting future occurrences. Consider, instead of $(x)Fx$, the hypothesis Fc. By itself, this hypothesis holds in four (1, 3, 4, 7) out of eight state descriptions. Suppose we find as evidence that Fa. The range of this evidence is four state descriptions (1, 2, 3, 5), and the hypothesis holds in two of them. But $4/8 = 2/4$, so the evidence has done nothing to support the hypothesis. Moreover, if we learn that Fb is true our new evidence is $Fa.Fb$, which holds in two state descriptions (1, 2), and our hypothesis holds in only one of them, giving us a ratio of 1/2. Hence, according to this way of defining confirmation, what we observe in the past and present has no bearing on what will occur in the future. This is an unacceptable consequence. When we examined the hypothesis $(x)Fx$ in the preceding paragraph we appeared to be achieving genuine confirmation, but that was not happening at all. The hypothesis $(x)Fx$ simply states that a, b, and c all have property F. When we find out by observing the first ball that it is red, we have simply reduced the predictive content of h. At first it predicted the color of three balls; after we examine the first ball it predicts the color of only two balls. After we observe the second ball, the hypothesis predicts the color of only one ball. If we were to examine the third ball and find it to be red, our hypothesis would have no predictive content at all. Instead of confirming our hypothesis we were actually simply reducing its predictive import.

In order to get around the foregoing difficulty, Carnap proposed a different way of evaluating the ranges of statements. The method adopted by Wittgenstein amounts to assigning equal weights to all of the state descriptions. Carnap suggested assigning unequal weights on the following basis. Let us take another look at our list of state descriptions in Table 2.2:

TABLE 2.2

State Description	Weight	Structure Description	Weight
1. $Fa.Fb.Fc$	1/4	All F	1/4
2. $Fa.Fb.{\sim}Fc$	1/12		
3. $Fa.{\sim}Fb.Fc$	1/12	2 F, 1 ${\sim}F$	1/4
4. ${\sim}Fa.Fb.Fc$	1/12		
5. $Fa.{\sim}Fb.{\sim}Fc$	1/12		
6. ${\sim}Fa.Fb.{\sim}Fc$	1/12	1 F, 2 ${\sim}F$	1/4
7. ${\sim}Fa.{\sim}Fb.Fc$	1/12		
8. ${\sim}Fa.{\sim}Fb.{\sim}Fc$	1/4	No F	1/4

Carnap noticed that state descriptions 2, 3, and 4 make similar statements about our miniature universe; they say that two entities have property F and one lacks it. Taken together, they describe a certain structure. They differ from one another in identifying the ball that is not red, but Carnap suggests that that is a secondary consideration. Similarly, state descriptions 5, 6, and 7, taken together describe a certain structure, namely, a universe in which one individual has property F and two lack it. Again, they differ in identifying the object that has this property. In contrast, state description 1, all by itself, describes a particular structure, namely, all threeentities have property F. Similarly, state description 8 describes the structure in which no object has that property.

Having identified the structure descriptions, Carnap proceeds to assign equal weights to them (each gets 1/4); he then assigns equal weights to the state descriptions within each structure description. The resulting system of weights is shown above. These weights are then used as a measure of the ranges of statements;[11] this system of measures is called m^*. A confirmation function c^* is defined as follows:[12]

$$c^*(H|E) = m^*(H.E)/m^*(E).$$

To see how it works, let us reconsider the hypothesis Fc in the light of different bits of evidence. First, the range of Fc consists of state description 1, which has weight 1/4, and 3, 4, and 7, each of which has weight 1/12. The sum of all of them is 1/2; that is, the probability of our hypothesis before we have any evidence. Now, we find that Fa; its measure is 1/2. The range of $Fa.Fc$ is state descriptions 1 and 3, whose weights are, respectively, 1/4 and 1/12, for a total of 1/3. We can now calculate the degree of confirmation of our hypothesis on this evidence:

$$c^*(H|E) = m^*(E.H)/m^*(E) = 1/3 \div 1/2 = 2/3.$$

Carrying out the same sort of calculation for evidence $Fa.Fb$ we find that our hypothesis has degree of confirmation 3/4. If, however, our first bit of evidence had been $\sim Fa$, the degree of confirmation of our hypothesis would have been 1/3. If our second bit of evidence had been $\sim Fb$, that would have reduced its degree of confirmation to 1/4. The confirmation function c^* seems to do the right sorts of things. When the evidence is what we normally consider to be positive, the degree of confirmation goes up. When the evidence is what we usually take to be negative, the degree of confirmation goes down. Clearly, c^* allows for learning from experience.

A serious philosophical problem arises, however. Once we start playing the game of assigning weights to state descriptions, we face a huge plethora of possibilities. In setting up the machinery of state descriptions and weights, Carnap demands only that the weights for all of the state descriptions add up to 1, and that each state description have a weight greater than 0. These conditions are sufficient to guarantee an admissible interpretation of the probability calculus. Carnap recognized the obvious fact that infinitely many confirmation functions satisfying this basic requirement are possible. The question is how to make an appropriate choice. It can easily be shown that choosing a confirmation function is precisely the same as assigning prior probabilities to all of the hypotheses that can be stated in the given language.

Consider the following possibility for a measure function:

[11] The measure of the range of any statement H can be identified with the prior probability of that statement in the absence of any background knowledge K. It is an *a priori* prior probability.

[12] Wittgenstein's measure function assigns the weight ⅛ to each state description; the confirmation function based upon it is designated $c\dagger$.

TABLE 2.3

State Description	Weight	Structure Description	Weight
1. *Fa.Fb.Fc*	1/20	All *F*	1/20
2. *Fa.Fb.~Fc*	3/20		
3. *Fa.~Fb.Fc*	3/20	2 *F*, 1 ~*F*	9/20
4. *~Fa.Fb.Fc*	3/20		
5. *Fa.~Fb.~Fc*	3/20		
6. *~Fa.Fb.~Fc*	3/20	1 *F*, 2 ~*F*	9/20
7. *~Fa.~Fb.Fc*	3/20		
8. *~Fa.~Fb.~Fc*	1/20	No *F*	1/20

(The idea of a confirmation function of this type was given in Burks 1953; the philosophical issues are further discussed in Burks 1977, Chapter 3.) This method of weighting, which may be designated $m^{\diamond}$, yields a confirmation function $C^{\diamond}$, which is a sort of counterinductive method. Whereas m^* places higher weights on the first and last state descriptions, which are state descriptions for universes with a great deal of uniformity (either every object has the property, or none has it), $m^{\diamond}$ places lower weights on descriptions of uniform universes. Like c^*, $c^{\diamond}$ allows for "learning from experience," but it is a funny kind of anti-inductive "learning." Before we reject $m^{\diamond}$ out of hand, however, we should ask ourselves if we have any a priori guarantee that our universe is uniform. Can we select a suitable confirmation function without being totally arbitrary about it? This is the basic problem with the logical interpretation of probability.

Part IV: Confirmation and Probability

2.9 THE BAYESIAN ANALYSIS OF CONFIRMATION

We now turn to the task of illustrating how the probabilistic apparatus developed above can be used to illuminate various issues concerning the confirmation of scientific statements. Bayes's theorem (Rule 9) will appear again and again in these illustrations, justifying the appellation of Bayesian confirmation theory.

Various ways are available to connect the probabilistic concept of confirmation back to the qualitative concept, but perhaps the most widely followed route utilizes an incremental notion of confirmation: E confirms H relative to the background knowledge K just in case the addition of E to K raises the probability of H, that is, $Pr(H|E.K) > Pr(H|K)$.[13] Hempel's study of instance confirmation in terms of a

[13] Sometimes, when we say that a hypothesis has been confirmed, we mean that it has been rendered highly probable by the evidence. This is a *high probability* or *absolute* concept of confirmation, and it should be carefully distinguished from the *incremental* concept now under discussion (see Carnap 1962, Salmon 1973, and Salmon 1975). Salmon (1973) is the most elementary discussion.