the available options. When the argument is tightened sufficiently, it turns out, it does not vindicate just one rule of inductive inference; instead, it equally justifies an infinite class of rules. Serious efforts—up to this time—to find a satisfactory basis for selecting a unique rule have been unsuccessful, (the technical details are discussed in Salmon 1967, Chapter 6).

Where do things stand now—250 years after the publication of Hume's *Treatise of Human Nature*—with respect to the problem we have inherited from him? Although many ingenious attempts have been made to solve or dissolve it there is still no consensus. It still stands as an item of "unfinished business" for philosophy of science (see Salmon 1978a). The problem may, perhaps, best be summarized by a passage from Hume himself:

> Let the course of things be allowed hitherto ever so regular, that alone, without some new argument or inference, proves not that for the future it will continue so. In vain do you pretend to have learned the nature of bodies from your past experience. Their secret nature, and consequently all their effects and influence, may change without any change in their sensible qualities. This happens sometimes, and with regard to some objects: Why may it not happen always, and with regard to all objects? What logic, what process or argument secures you against this supposition? My practice, you say, refutes my doubts. But you mistake the purport of my question. As an agent, I am quite satisfied in the point; but as a philosopher . . . I want to learn the foundation of this inference. (1748, Section 4)

As Hume makes abundantly clear, however, life—and science—go on in spite of these troubling philosophical doubts.

Part III: Probability

2.7 THE MATHEMATICAL THEORY OF PROBABILITY

Our discussion up to this point has been carried on without the aid of a powerful tool—the calculus of probability. The time has come to invoke it. The defects of the qualitative approaches to confirmation discussed in Sections 2.3 and 2.4 suggest that an adequate account of the confirmation of scientific statements must resort to quantitative or probabilistic methods. In support of this suggestion, recall that we have already come across the concept of probability in the discussion of the qualitative approaches. In our discussion of the H-D method, for instance, we encountered the concept of probability in at least two ways. First, noting that a positive result of an H-D test does not conclusively establish a hypothesis, we remarked that it might render the hypothesis a little more probable than it was before the test. Second, in dealing with the problem of statistical hypotheses, we saw that only probabilistic observational predictions can be derived from such test hypotheses. In order to pursue our investigation of the issues that have been raised we must take a closer look at the concept or concepts of probability.

The modern theory of probability had its origins in the seventeenth century. Legend has it that a famous gentleman, the Chevalier de Méré, posed some questions about games of chance to the philosopher-mathematician Blaise Pascal. Pascal communicated the problems to the mathematician Pierre de Fermat, and that was how it all began. Be that as it may, the serious study of mathematical probability theory began around 1660, and Pascal and Fermat, along with Christian Huygens, played crucial roles in that development, (for an historical account see Hacking 1975 and Stigler 1986).

In order to introduce the theory of probability, we take probability to be a relationship between events of two different types—for example, between tossing a standard die and getting a six, or drawing from a standard bridge deck and getting a king. We designate probabilities by means of the following notation:

$Pr(B|A)$ is the probability of a result of the type B given an event of the type A.

If A is a toss of a standard die and B is getting a three, then "$Pr(B|A)$" stands for the probability of getting a three if you toss a standard die. As the theory of probability is seen today, all of the elementary rules of probability can be derived from a few simple axioms. The meanings of these axioms and rules can be made intuitively obvious by citing examples from games of chance that use such devices as cards and dice. After some elementary features of the mathematical calculus of probability have been introduced in this section, we look in the following section at a variety of interpretations of probability that have been proposed.

AXIOMS (BASIC RULES)

Axiom (rule) 1: Every probability is a unique real number between zero and one inclusive; that is,

$$O \leq Pr(B|A) \leq 1.$$

Axiom (rule) 2: If A logically entails B, then $Pr(B|A) = 1$.

Definition: Events of types B and C are *mutually exclusive* if it is impossible for both B and C to happen on any given occasion. Thus, for example, on any draw from a standard deck, drawing a heart and drawing a spade are mutually exclusive, for no card is both a heart and a spade.

Axiom (rule) 3: If B and C are mutually exclusive, then

$$Pr(B \vee \mathrm{C|A}) = Pr(B|A) + Pr(C|A).$$

This axiom is also known as the *special addition rule*.

Example: The probability of drawing a heart or a spade equals the probability of drawing a heart plus the probability of drawing a spade.

Axiom (rule) 4: The probability of a *joint occurrence*—that is, of a conjunction of B and C—is equal to the probability of the first multiplied by the probability of the second given that the first has occurred:

$$Pr(B.C|A) = Pr(B|A) \times Pr(C|A.B).$$

This axiom is also known as the *general multiplication rule*.

Example: If you make two draws *without replacement* from a standard deck, what is the probability of getting two aces? The probability of getting an ace on the first draw is 4/52; the probability of getting an ace on the second draw *if you have already drawn an ace on the first draw* is 3/51, because there are only 51 cards left in the deck and only 3 of them are aces. Thus, the probability of getting two aces is

$$4/52 \times 3/51 = 12/2652 = 1/221.$$

Some Derived Rules

From the four axioms (basic rules) just stated, several other rules are easy to derive that are extremely useful in calculating probabilities. First, we need a definition:

Definition: The events B and C are *independent* if and only if

$$Pr(C|A.B) = Pr(C|A).$$

When the events B and C are *independent* of one another, the multiplication rule (axiom 4) takes on a very simple form:

Rule 5: If B and C are independent, given A, then

$$Pr(B.C|A) = Pr(B|A) \times Pr(C|A).$$

This rule is known as the *special multiplication rule.* (Proofs, sketches of proofs, and other technical items will be placed in boxes. They can be omitted on first reading.)

Proof of Rule 5: Substitute $Pr(C|A)$ for $Pr(B.C|A)$ in Axiom 4.

Example: What is the probability of getting double 6 (''boxcars'') when a standard pair of dice is thrown? Since the outcomes on the two dice are independent, and the probability of 6 on each die is 1/6, the probability of double 6 is

$$1/6 \times 1/6 = 1/36$$

Example: What is the probability of drawing two spades on two consecutive draws when the drawing is done *with replacement*? The probability of getting a spade on the first draw is 13/52 = 1/4. After the first card is drawn, whether it is a spade or not, it is put back in the deck and the deck is reshuffled. Then the second card is drawn. Because of the replacement, the outcome of the second draw is independent of the outcome of the first draw. Therefore, the probability of getting a spade on the second draw is just the same as it was on the first draw. Thus, the probability of getting two spades on two consecutive draws is

$$1/4 \times 1/4 = 1/16$$

NOTE CAREFULLY. If the drawing is done *without replacement,* the special multiplication rule cannot be used because the outcomes are *not independent.* In that case Rule 4 must be used.

Rule 6: $Pr(\sim B|A) = 1 - Pr(B|A)$.

This simple rule is known as the *negation rule.* It is very useful.

Example: Suppose you would like to know the probability of getting at least one 6 if you toss a standard die three times.[8] That means you want to know the probability of getting a 6 on the first toss *or* on the second toss *or* on the third toss, where this is an *inclusive or.* Thus, the outcomes are *not* mutually exclusive, so you cannot use the special addition rule (Axiom 3). We can approach this problem via the negation. To fail to get at least one 6 in three tosses means to get non-6 on the first toss *and* non-6 on the second toss *and* non-6 on the third toss. Since the probability of 6 is 1/6, the negation rule tells us that the probability of non-6 is 5/6. Because the outcomes on the three tosses are *independent,* we can use Rule 5 to obtain the probability of non-6 on all three tosses as

$$5/6 \times 5/6 \times 5/6 = 125/216.$$

The probability of getting at least one 6, which is the negation of not getting any 6, is therefore

$$1 - 125/216 = 91/216.$$

NOTE CAREFULLY: The probability of getting at least one 6 in three tosses is *not* 1/2. It is equal to 91/216, which is approximately 0.42.

Proof of Rule 6: Obviously every A is either a B or not a B. Therefore, by Axiom 2,

$$Pr(B \vee \sim B|A) = 1.$$

Since B and $\sim B$ are mutually exclusive, Axiom 3 yields

$$Pr(B|A) + Pr(\sim B|A) = 1.$$

Rule 6 results from subtracting $Pr(B|A)$ from both sides.

Rule 7: $Pr(B \vee C|A) = Pr(B|A) + Pr(C|A) - Pr(B.C|A)$.

This is the *general addition rule.* Unlike Rule 3, this rule applies to outcomes B and C even if they are not mutually exclusive.

Example: What is the probability of getting a spade or a face card in a draw from a standard deck? These two alternatives are not mutually exclusive, for there are three

[8] This example is closely related to one of the problems posed by the Chevalier de Méré. How many tosses of a pair of dice, he asked, are required to have at least a fifty-fifty chance of getting at least one double 6? It seems that a common opinion among gamblers at the time was that 24 tosses would be sufficient. The Chevalier doubted that answer, and it turned out that he was right. One needs 25 tosses to have at least a fifty-fifty chance.

cards—king, queen, and jack of spades—that are both face cards and spades. Since there are 12 face cards and 13 spades, the probability of a spade or a face card is

$$12/52 + 13/52 - 3/52 = 22/52$$

It is easy to see why this rule has the form that it does. If B and C are not mutually exclusive, then some outcomes may be both B and C. Any such items will be counted twice—once when we count the Bs and again when we count the Cs. (In the foregoing example, the king of spades is counted once as a face card and again as a spade. The same goes for the queen and jack of spades.) Thus, we must subtract the number of items that are both B and C, in order that they be counted only once.

> How to prove Rule 7. First, we note that the class of things that are B or C in the *inclusive* sense consists of those things that are $B.C$ or $\sim B.C$ or $B.\sim C$, where these latter three classes are *mutually exclusive*. Thus, Rule 3 can be applied, giving
>
> $$Pr(B \vee C|A) = Pr(B.C|A) + Pr(\sim B.C|A) + Pr(B.\sim C|A).$$
>
> Rule 4 is applied to each of the three terms on the right-hand side, and then Rule 6 is used to get rid of the negations inside of the parentheses. A bit of simple algebra yields Rule 7.

Rule 8: $Pr(C|A) = Pr(B|A) \times Pr(C|A.B) + Pr(\sim B|A) \times Pr(C|A.\sim B)$.

This is the *rule of total probability*. It can be illustrated as follows:

Example: Imagine a factory that produces frisbees. The factory contains just two machines, a new machine **B** that produces 800 frisbees each day, and an old machine **~B** that produces 200 frisbees per day. Among the frisbees produced by the new machine, 1% are defective; among the frisbees produced by the old machine, 2% are defective. Let A stand for the frisbees produced in a given day at that factory. Let B stand for the frisbees produced by the new machine; $\sim B$ then stands for those produced by the old machine. Let C stand for defective frisbees. Then,

$Pr(B|A)$ = the probability that a frisbee is produced by machine **B** = 0.8

$Pr(\sim B|A)$ = the probability that a frisbee is produced by machine **~B** = 0.2

$Pr(C|A.B)$ = the probability that a frisbee produced by machine **B** is defective = 0.01

$Pr(C|A.\sim B)$ = the probability that a frisbee produced by machine **~B** is defective = 0.02

Therefore, the probability that a frisbee is defective =

$$0.8 \times 0.01 + 0.2 \times 0.02 = 0.012$$

As can be seen from this artificial example, the rule of total probability can be used to calculate the probability of an outcome that can occur in either of two ways, either by the occurrence of some intermediate event B or by the nonoccurrence of B. The situation can be shown in a diagram:

$$\begin{tikzcd} & B \arrow[dr] & \\ A \arrow[ur] \arrow[dr] & & C \\ & \sim B \arrow[ur] & \end{tikzcd}$$

Proof of Rule 8: Since every C is either a B or not a B, the class C is identical to the class $B.C \vee \sim B.C$; moreover, since nothing is both a B and not a B, the classes $B.C$ and $\sim B.C$ are mutually exclusive. Hence,

$$\begin{aligned} Pr(C|A) &= Pr(C.[B \vee \sim B]\,|A) \\ &= Pr([B.C \vee \sim B.C]|A) \\ &= Pr(B.C|A) + Pr(\sim B.C|A) \qquad \text{by Rule 3} \\ &= Pr(B|A) \times Pr(C|A.B) + Pr(\sim B|A) \times Pr(C|A.\sim B) \\ &\qquad \text{by Rule 4 applied twice} \end{aligned}$$

We now come to the rule of probability that has special application to the problem of confirmation of hypotheses.

Rule 9: $$Pr(B|A.C) = \frac{Pr(B|A) \times Pr(C|A.B)}{Pr(C|A)}$$

$$= \frac{Pr(B|A) \times Pr(C|A.B)}{Pr(B|A) \times Pr(C|A.B) + Pr(\sim B|A) \times Pr(C|A.\sim B)}$$

provided that $Pr(C|A) \neq 0$. The fact that these two forms are equivalent follows immediately from the rule of total probability (Rule 8), which shows that the denominators of the right-hand sides are equal to one another.

Rule 9 is known as *Bayes's rule*; it has extremely important applications. For purposes of illustration, however, let us go back to the trivial example of the frisbee factory that was used to illustrate the rule of total probability.

Example: Suppose we have chosen a frisbee at random from the day's production and it turns out to be defective. We did not see which machine produced it. What is the probability—$Pr(B|A\ .\ C)$—that it was produced by the new machine? Bayes's rule gives the answer:

$$\frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.2 \times 0.02} = \frac{0.008}{0.012} = 2/3$$

The really important fact about Bayes's rule is that it tells us a great deal about the confirmation of hypotheses. The frisbee example illustrates this point. We have a frisbee produced at this factory (A) that turns out, on inspection, to be defective (C), and we wonder whether it was produced (caused) by the new machine (B). In other

Proof of Bayes's rule: Bayes's rule has two forms as given above; we show how to prove both. We begin by writing Rule 4 twice; in the second case we interchange B and C.

$$Pr(B.C|A) = Pr(B|A) \times Pr(C|A.B)$$

$$Pr(C.B|A) = Pr(C|A) \times Pr(B|A./C)$$

Since the class $B.C$ is obviously identical to the class $C.B$ we can equate the right-hand sides of the two equations:

$$Pr(C|A) \times Pr(B|A.C) = Pr(B|A\ ) \times Pr(C|A.B)$$

Assuming that $P(C|A) \neq 0$, we divide both sides by that quantity:

$$Pr(B|A.C) = \frac{Pr(B|A) \times Pr(C|A.B)}{Pr(C|A)}$$

This is the first form. Using Rule 8, the rule of total probability, we replace the denominator, yielding the second form:

$$Pr(B|A.C) = \frac{Pr(B|A) \times Pr(C|A.B)}{Pr(B|A) \times Pr(C|A.B) + Pr(\sim B|A) \times Pr(C|A.\sim B)}$$

words, we are evaluating the hypothesis that the new machine produced this defective frisbee. As we have just seen, the probability is 2/3.

Inasmuch as we are changing our viewpoint from talking about types of objects and events A, B, C, . . . to talking about hypotheses, let us make a small change in notation to help in the transition. Instead of using "A" to stand for the day's production of frisbees, we shall use "K" to stand for our background *knowledge* about the situation in that factory. Instead of using "B" to stand for the frisbees produced by the new machine **B**, we shall use "H" to stand for the *hypothesis* that a given frisbee was produced by machine **B**. And instead of using "C" to stand for defective frisbees, we shall use "E" to stand for the *evidence* that the given frisbee is defective. Now Bayes's rule reads as follows:

$$\text{Rule 9: } Pr(H|K.E) = \frac{Pr(H|K) \times Pr(E\ |K.H)}{Pr(E|K)}$$

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

Changing the letters in the formula (always replacing the same old letter for the same new letter) obviously makes no difference to the significance of the rule. If the axioms are rewritten making the same changes in variables, Rule 9 would follow from them in exactly the same way. And inasmuch as we are still talking about *probabilities*—albeit the probabilities of hypotheses instead of the probabilities of events—we still need the same rules.

We can now think of the probability expressions that occur in Bayes's rule in the following terms:

> $Pr(H|K)$ is the *prior probability* of hypothesis H just on the basis of our background knowledge K without taking into account the specific new evidence E. (In our example, it is the probability that a given frisbee was produced by machine **B**.)
> $Pr(\sim H|K)$ is the *prior probability* that our hypothesis H is false. (In our example, it is the probability that a given frisbee was produced by machine **~B**.) Notice that H and $\sim H$ must exhaust all of the possibilities.

By the negation rule (Rule 6), these two prior probabilities must add up to 1; hence, if one of them is known the other can immediately be calculated.

> $Pr(E|K.H)$ is the probability that evidence E would obtain given the truth of hypothesis H in addition to our background knowledge K. (In our example, it is the probability that a particular frisbee is defective, given that it was produced by machine **B**.) This probability is known as a *likelihood*.
> $Pr(E|K.\sim H)$ is the probability that evidence E would obtain if our hypothesis H is false. (In our example, it is the probability that a particular frisbee is defective if it was not produced by machine **B**.) This probability is also a *likelihood*.

The two likelihoods—in sharp contrast to the prior probabilities—are independent of one another. Given only the value of one of them, it is *impossible* to calculate the value of the other.

> $Pr(E|K)$ is the probability that our evidence E would obtain, regardless of whether hypothesis H is true or false. (In our example, it is the probability that a given frisbee is defective, regardless of which machine produced it.) This probability is often called the *expectedness* of the evidence.[9]
> $Pr(H|K . E)$ is the probability of our hypothesis, judged in terms of our background knowledge K and the specific evidence E. It is known as the *posterior probability*. This is the probability we are trying to ascertain. (In our example, it is the probability that the frisbee was produced by the new machine. Since the posterior probability of H is different from the prior probability of H, the fact that the frisbee is defective is evidence relevant to that hypothesis.)

[9] Expectedness is the opposite of *surprisingness*. If the expectedness of the evidence is small the evidence is surprising. Since the expectedness occurs in the denominator of the fraction, the smaller the expectedness, the greater the value of the fraction. Surprising evidence confirms hypotheses more than evidence that is to be expected regardless of the hypothesis.

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

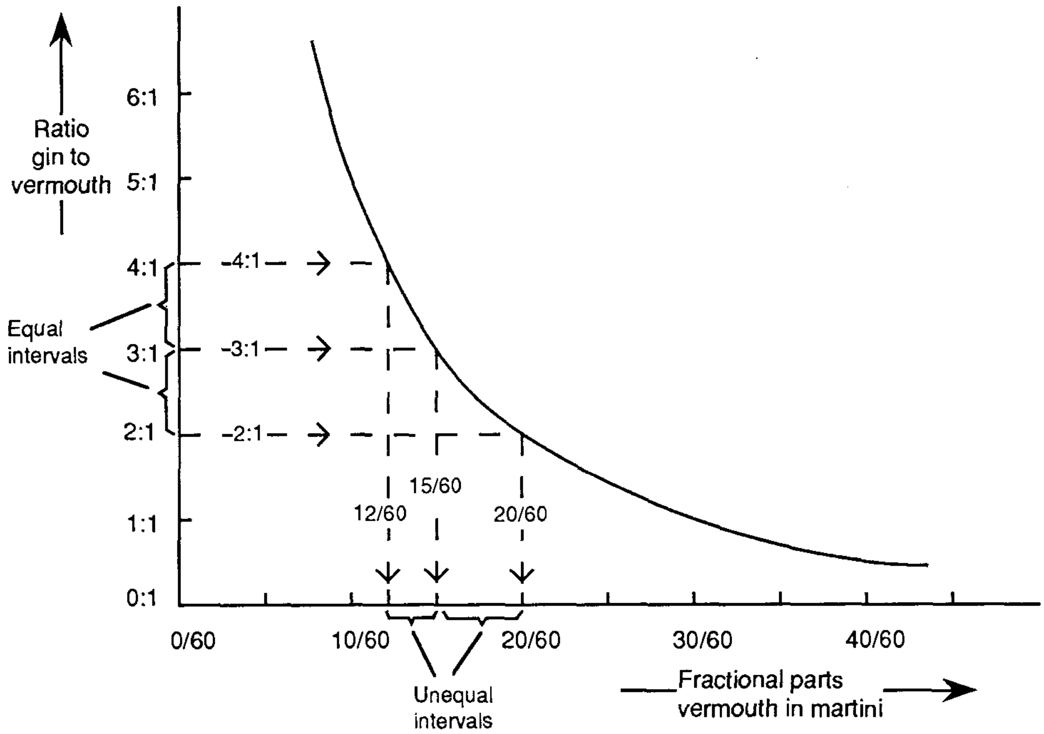

Figure 2.2

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.

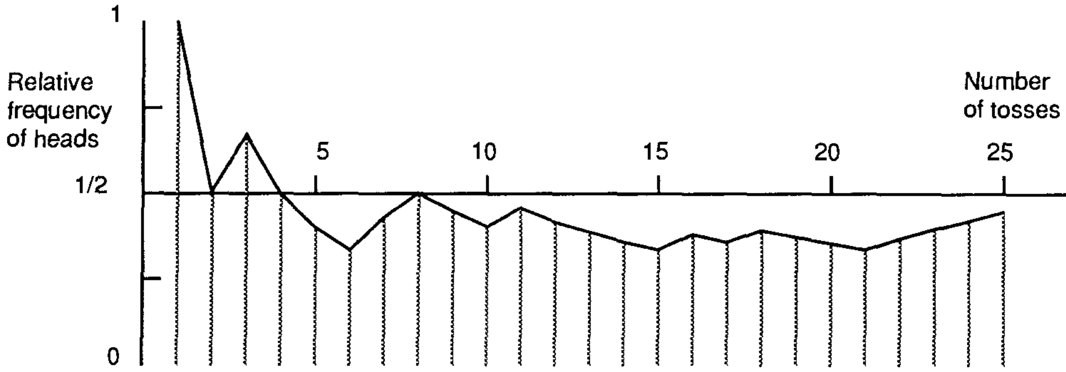

Figure 2.3

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 *n*th 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 proba-