Chapter 3
DECISIONS UNDER RISK: PROBABILITY

3-1. Maximizing Expected Values

When Michael Smith applied to Peoples Insurance Company for a life insurance policy in the amount of $25,000 at a premium of $100 for one year, he was required to fill out a lengthy health and occupation questionnaire and submit to an examination by a physician. According to the results, Smith fell into the category of persons with a 5% death rate for the coming year. This enabled Peoples to assign a probability of .05 to Smith's dying within the year, and thus the decision whether or not to insure him became one *under risk*.

I will represent this decision problem by means of decision table 3-1. Notice that I have put both outcome values and probabilities in the same square.

3-1	Smith Dies	Smith Lives
Insure Smith	−$25,000 .05	$100 .95
Do Not	$0 .05	$0 .95

Let us leave the Peoples example for a moment and turn to Sally Harding's decision. She needs a car and has a choice of buying an old heap for $400 or a four-year-old car for $3,000. Harding plans to leave the country at the end of the year, so either car need serve her for only the year. That is why she is considering the old heap. On the other hand, if either car breaks down before the end of the year, she intends to rent one at an estimated cost of $200. Harding would pay cash for the older car and junk it at the end of the year. If, on the other hand, she buys the newer car, she will finance it and sell it when she leaves. She believes that her net cost for the newer car would come to $500. Being an experienced mechanic, Harding estimates that the chances of the old heap surviving the year are 5 in 10 and that the newer car's chances are 9 in 10.

Harding's decision is again one under risk and can be represented by table 3-2.

3-2

	No Need to Rent	Must Rent
Buy New	− \$500 .90	− \$700 .10
Buy Old	− \$400 .50	− \$600 .50

You might be tempted to solve Harding's problem by appealing to the dominance principle, for buying the old heap costs less no matter what happens. But remember, the chances of Harding having to rent a car are a function of which car she buys. Buying the older car increases her chances of having to rent by a factor of 5. This is exactly the sort of case where the dominance principle should not be applied.

In Harding's decision problem, the probabilities of the states vary with the acts. Writing different probability numbers in different rows reflects this. In the insurance problem the acts do not affect the probabilities of the states; hence, the probabilities of the states conditional on the acts collapse to absolute probabilities. That is why the same probability numbers are written in each row.

Peoples Insurance Company came to a decision quite easily. They decided they could not afford to insure Smith—at least not at a \$100 premium. They reasoned that if they were to insure a hundred people like Smith, five would die. That would cost the company \$125,000 in death benefits and would be counterbalanced by only \$10,000. So the net loss would be \$115,000. Peoples concluded that, whether they insured a hundred, a thousand, or a million people similar to Smith, their losses would average close to \$1,150 per person. On the other hand, they would neither gain nor lose by not insuring Smith and others like him.

Peoples proceeded as insurance companies are wont to do. They calculated the *expected monetary value* (EMV) of each option and chose the one with the highest EMV. You can calculate the EMV for an act from a decision table quite easily: Multiply the monetary value in each square by the probability number in that square and sum across the row. Do this for Sally Harding's decision. You will see that the EMV of buying the newer car is − \$520 and the EMV of buying the older one is − \$500.

Should Sally Harding buy the older car? On the average, that would save her the most money. But what about the pleasures of riding in a newer car? And what of the worry that the older one will break down? These relevant factors have been neglected. But perhaps these could be assigned a monetary value and the decision could be made on the basis of a revised set of EMVs. There remains another problem. Sally Harding will buy a car only once this year, and to the best of our knowledge, no one else has been confronted with a very similar decision. So how can we speak of what happens "on the average" here? This is a one-shot deal so there is nothing to average out. And even if there were, Sally

Harding should be concerned with what happens to her, not to the average person like her.

We can deal with these problems by replacing monetary values with utility values and showing that maximizing expected utility does not suffer from the drawbacks of maximizing EMVs. Utility is thus one of the chief ingredients of the theory of decisions under risk. But before we can attend to it we must achieve a better understanding of the other major ingredient, namely, probability.

3-2. Probability Theory

Probability judgments are now a commonplace of our daily life. Every time we turn on the radio or pick up a newspaper we are apt to encounter a weather, economic, or political forecast framed in probabilistic terms. Probability is also essential to theoretical science since it is the backbone of certain areas of physics, genetics, and the social sciences. Despite its pervasiveness in our culture, there is still much debate among philosophers, statisticians, and scientists concerning what probability statements mean and how they may be justified or applied. Some believe, for instance, that it is not proper to assign probabilities to single events, and hence that probability judgments inform us only of proportions. They claim that, strictly speaking, we should not say that the chances are 1 in 10 that Jackson will die within the year; rather, we should say that in a large group of people like Jackson, 10% will die. Others believe that single events or statements can be assigned probabilities by simply refining our best hunches. Some believe that probability is a measure of the *strength of beliefs* or lack thereof; others think it is a property of reality.

There is, however, a common focal point for the study of probability—*the probability calculus*. This is a mathematical theory that enables us to calculate additional probabilities from those we already know. For instance, if I know that the probability that it will rain tomorrow is 50% and that there is a 10% probability that the price of gold will drop tomorrow, the calculus allows me to conclude that the probability that both will occur tomorrow is 5%. The beauty of the probability calculus is that it can be used with almost all the interpretations of probability that have been proposed. Furthermore, much of the superstructure of the theory of decision under risk depends on the probability calculus. Let us then turn to this calculus before we plunge into the more philosophical questions about probability.

There are a number of alternative presentations of the probability calculus. I will use a formulation in which probabilities are assigned to simple and compound statement forms, since this will enable many philosophical readers to draw on their training in symbolic logic.

The basic formulas of the calculus will take the form

$$P(S) = a.$$

Here S represents a statement form, such as "p or q," "p & q," or "p & (q or r)," and a represents a numeral. The entire expression "$P(S) = a$" should be read as: "the probability of S is a." Some examples are

$P(p) = 1/2$, $P(p \text{ or } q) = 2/3$, $P(\text{not } q \text{ or } r) = .9$.

In applications the letters will be replaced with sentences or sentence abbreviations. Thus we can have:

P(the coin comes up heads) = 1/2
P(heart or diamond) = 1/2
P(ace and spade) = 1/52.

These statements are all *absolute* probability statements.

We will also need *conditional* probability statements, which we will write as

$P(S/W) = a$

and read as "the probability of S given $W = a$." Because it is essential to appreciate the difference between conditional and absolute probability statements, let us reflect on an example. The probability of drawing a heart at random from a fair deck is 1/4, but if all the black cards were removed and only red cards remained, the probability in question would be the probability of drawing a heart *given* that the card to be drawn is red. That would be 1/2. We express the first probability judgment as an absolute probability statement

P(heart) = 1/4,

but the second should be formulated as the conditional probability

P(heart/red) = 1/2.

Now one might think that the conditional probability of a heart given a red card is just the absolute probability of the conditional "If the card is red, then it is a heart." But that will not work, given our current methods for treating conditionals in logic and mathematics. For the conditional in question is equivalent to "either the card is not red or it is a heart," and

P(not red or heart) = 3/4

since thirty-nine out of fifty-two cards are either not red or hearts.

Another important point: $P(S/W)$ and $P(W/S)$ are generally distinct. Thus P(heart/red) = 1/2 but P(red/heart) = 1, since every heart is red.

Conditional and absolute probability are related, however, through a number of laws of probability. This is one of the most important of those laws: $P(p \;\&\; q) = P(p) \times P(q/p)$. It says that the probability of a conjunction is the probability of the first component times the probability of the second given the first. The idea here is that it is less probable that two things will be true together than that either one will be true separately. Since probabilities are less than or equal to 1, multiplying them will produce a number smaller than or equal to either. But in general we cannot simply multiply the probabilities of the two conjuncts, since their truth and falsity might be linked. There is, for instance, no chance for a coin to come up both heads and tails, yet the simple product of the probabilities that it does is 1/4. On the other hand, multiplying by the conditional probability of the second component given the first avoids this difficulty:

P[heads and tails (on the same toss)] $= 0 = P(\text{heads}) \times P(\text{tails/heads})$
$= 1/2 \times 0$.

Example. What is the probability of drawing two aces in a row from an ordinary deck of fifty-two cards when the first card is not put back into the deck? This is the probability of drawing an ace on the first draw and then again on the second draw. So we must calculate as follows:

$P(\text{ace on draw 1 and ace on draw 2}) = P(\text{ace on draw 1}) \times$
$P(\text{ace on draw 2/ace on draw 1}) = 4/52 \times 3/51 = 3/663$.

The first probability is simply the ratio of the aces to the total number of cards in the deck. The second is figured by noting that if an ace is drawn on the first draw and not replaced, three aces out of a total of fifty-one cards remain.

This example illustrates another important concept of probability theory. Because the first card is not put back into the deck, the outcome of the first draw affects the outcome of the second draw: There are fifty-one cards to draw after the first draw, and one less ace if the first card drawn was an ace. On the other hand, if the first card drawn is replaced and the deck reshuffled, the outcome of the first draw has no effect on the outcome of the second draw. In this case, we say that the outcomes are *independent*. When we replace and reshuffle the cards

$P(\text{ace on draw 2}) = P(\text{ace on draw 2/ace on draw 1})$.

This leads to the following definition.

Definition 1. p is *independent* of q if and only if $P(p) = P(p/q)$.

Another important concept we will need is that of *mutual exclusiveness*. If a single card is drawn, then drawing an ace and simultaneously drawing a king are mutually exclusive. But drawing an ace and drawing a spade are not; one can draw the ace of spades. Generalizing we have:

Definition 2. p and q are mutually exclusive if and only if it is impossible for both to be true.

If p and q are *mutually exclusive*, if one is true the other must be false. So if p and q are mutually exclusive, $P(p/q)$ and $P(q/p)$ are 0. Consequently, if p and q are mutually exclusive, p and q will not be independent of each other—unless each already has a probability of 0.

With these definitions in hand we can now lay down the basic laws or *axioms* of the probability calculus. All the other laws of the calculus can be derived from these using purely logical and mathematical reasoning.

A (logically or mathematically) impossible statement has no probability of being true; so 0 is a natural lower bound for probabilities. On the other hand, the probability of a certainty being true is 100%; so 1 is a natural upper bound for probabilities. Statements that are neither impossible nor certain have probabilities between 0 and 1. These considerations motivate the first two axioms of the calculus.

AXIOM 1. a. $0 \leq P(p) \leq 1$.
b. $0 \leq P(p/q) \leq 1$.

AXIOM 2. If p is certain, then $P(p) = 1$.

What is the probability of drawing a face card or an ace? Four cards out of fifty-two are aces and twelve are face cards. So the probability of drawing an ace or a face card is 16/52 or 5/13. This example illustrates the next axiom.

AXIOM 3. If p and q are mutually exclusive, then
$P(p \text{ or } q) = P(p) + P(q)$.

We can immediately apply these axioms to derive this law.

THEOREM 1. $P(p) + P(\text{not } p) = 1$.

PROOF. "p" and "not p" are mutually exclusive, and their disjunction is certain. Thus by axioms 2 and 3 we have:

$$1 = P(p \text{ or not } p) = P(p) + P(\text{not } p).$$

Theorem 1 yields a rule for calculating negated statements. If we already know $P(p)$, we can find $P(\text{not } p)$ by subtracting $P(p)$ from 1.

Example. What is the probability of taking more than one roll of a fair die to get a 6? Since taking more than one roll to get a 6 is the same as not getting a 6 on the first roll, $P(\text{not } 6)$ is the probability we want. But it must equal 5/6, since $P(6) = 1/6$.

We can also generalize the proof of theorem 1 to establish that the probabilities of any set of mutually exclusive and exhaustive alternatives sum to 1. (Alternatives are exhaustive if and only if it is certain that at least one of them is true.) This is important for decision theory, since it tells us that the probabilities in each row of a properly specified decision table must total 1.

Two equivalent statements must both be true together or both be false together. Thus there is no chance for the one to be true when the other is not. This leads to the following theorem.

THEOREM 2. If p and q are equivalent, then $P(p) = P(q)$.

PROOF. Suppose that p and q are equivalent. Then one is true just in case the other is. But then (a) "either not p or q" is certain and (b) q and not p are mutually exclusive. From (a) and axiom 2, we conclude that

$$P(\text{not } p \text{ or } q) = 1.$$

From (b) and axiom 3, we get

$$P(\text{not } p \text{ or } q) = P(\text{not } p) + P(q).$$

Putting this together with an application of theorem 1 yields

$$1 = 1 - P(p) + P(q)$$

from which $P(p) = P(q)$ follows immediately by algebra.

This theorem lets us draw on mathematics and logic to show that certain statements or statement forms are equiprobable. (In applying this and axioms 2 and 3 to examples outside mathematics and logic, we may also use those equivalences, certainties, and impossibilities that are part of the background assumptions of the application. For example, we may ordinarily assume that it is certain that a coin will land either heads or tails, though one tossed onto a sand pile need not.) The next theorem appeals to logical equivalences.

THEOREM 3. $P(p \text{ or } q) = P(p) + P(q) - P(p \,\&\, q)$.

PROOF. "p or q" is logically equivalent to "either p & q or p & not q or else not p & q"; so their probabilities must be equal. But the first two disjuncts are mutually exclusive with the third. Thus by substituting in axiom 3, we obtain

(1) $P(\text{either } p \,\&\, q \text{ or } p \,\&\, \text{not } q \text{ or else not } p \,\&\, q) = P(\text{either } p \,\&\, q \text{ or } p \,\&\, \text{not } q) + P(\text{not } p \,\&\, q)$.

But "either p & q or p & not q" is equivalent to "p"; so if we apply theorem 2 and substitute in equation (1) we get:

(2) $P(\text{either } p \,\&\, q \text{ or } p \,\&\, \text{not } q \text{ or else not } p \,\&\, q) = P(p) + P(\text{not } p \,\&\, q)$.

But, remember, the left side of equation (1) equals $P(p \text{ or } q)$; so we may get

(3) $P(p \text{ or } q) = P(p) + P(\text{not } p \,\&\, q)$.

Adding $P(p \,\&\, q)$ to both sides of equation (3) we get

(4) $P(p \text{ or } q) + P(p \,\&\, q) = P(p) + P(\text{not } p \,\&\, q) + P(p \,\&\, q)$.

Noting that "not p & q or else p & q" is equivalent to "q" and that its disjuncts are mutually exclusive, we may apply axioms 3 and 4 to equation (4) to obtain:

(5) $P(p \text{ or } q) + P(p \,\&\, q) = P(p) + P(q)$.

The theorem then follows by subtracting $P(p \,\&\, q)$ from both sides of this equation.

Theorem 3 and axiom 3 can both be used to calculate the probabilities of disjunctions from the probabilities of their components. Axiom 3 is simpler to use, but it does not always apply, whereas there is no restriction on theorem 3. (Note that theorem 3 has axiom 3 as a special case; for when p and q are mutually exclusive $P(p \,\&\, q) = 0$.) The more complicated appearance of theorem 3 is to prevent double counting when calculating probabilities of disjunctions whose components do not exclude each other. For example, the probability of a heart or a king is *not* $1/4 + 1/13$ because the king of hearts would be counted twice; rather the probability is $1/4 + 1/13 - 1/52$ where the double count has been subtracted. Notice that this follows the model of theorem 3.

Example. What is the probability of getting exactly two heads on three tosses of a fair coin? The two heads might occur in any one of three mutually exclusive ways: *HHT*, *HTH*, and *THH*. Each of these has a probability of 1/8 since each is one of the eight possibilities. Thus the answer is 3/8.

Example. What is the probability of getting a heart or an ace on at least one of two draws from a deck of cards, where the card drawn is replaced and the deck reshuffled after the first draw? The probability of getting a heart or an ace on the first draw is $13/52 + 4/52 - 1/52$ or $4/13$. But the probabilities are the same for the second draw. Furthermore, the question allows for the possibility that a heart or an ace is drawn both times; so the probability in question is $8/13$.

We still lack methods for calculating the probability of conjunctions. This is remedied by the next axiom, which we discussed in connection with conditional probability.

AXIOM 4. $P(p \;\&\; q) = P(p) \times P(q/p)$.

If $P(p) \neq 0$, we can divide both sides of axiom 4 by it, obtaining the formula of the next theorem.

THEOREM 4. If $P(p) \neq 0$, then
$P(q/p) = P(p \;\&\; q)/P(p)$.

(In many presentations theorem 4 is taken as the definition of conditional probability.)

According to definition 1, if q is independent of p, $P(q) = P(q/p)$. Thus, by axiom 4, we have:

THEOREM 5. If q is independent of p, then
$P(p \;\&\; q) = P(p) \times P(q)$.

Theorem 5 and axiom 4 let us calculate the probabilities of conjunctions in terms of their components. This is illustrated in the next example.

Example. What is the probability of getting twenty heads on twenty tosses of a fair coin? This is the probability of getting heads on the first toss and on the second toss and . . . and on the twentieth toss. But each toss is independent of the others, so the probability of getting twenty heads is the probability of getting one head multiplied by itself nineteen more times, i.e., $(1/2)^{20}$.

The next theorem shows that independence is almost always mutual.

THEOREM 6. p is independent of q if and only if q is independent of p, provided that $P(p)$ and $P(q)$ are both nonzero.

PROOF. Suppose that both $P(p)$ and $P(q)$ are not zero. Suppose that p is independent of q. Then $P(p) = P(p/q)$. Then, by theorem 4,

$$\begin{aligned} P(q/p) &= P(q \;\&\; p)/P(p) = [P(q) \times P(p/q)]/P(p) \\ &= [P(q) \times P(p)]/P(p) = P(q). \end{aligned}$$

That means that q is independent of p. By interchanging "q" and "p" in this proof we can show that p is independent of q if q is independent of p.

The next theorem relates independence and mutual exclusiveness.

THEOREM 7. If p and q are mutually exclusive and both $P(p)$ and $P(q)$ are nonzero, then p and q are not independent.

PROOF. If p and q are mutually exclusive, then the negation of their conjunction is certain; so by theorem 1 and axiom 2, we have $P(p \,\&\, q) = 0$. On the other hand, if either p or q is independent of the other, then

$$P(p \,\&\, q) = 0 = P(p) \times P(q),$$

which implies that one of $P(p)$ or $P(q)$ is 0. That would contradict the hypothesis of the theorem.

Notice that the converse of theorem 7 does not hold: Independent statements need not be mutually exclusive. Getting heads on the second toss of a coin is independent of getting heads on the first toss, but they do not exclude each other.

The next pair of theorems, known as the inverse probability law and Bayes's Theorem, respectively, have been of fundamental importance in decision theory, statistics, and the philosophy of science.

THEOREM 8 (the inverse probability law). If $P(q) \neq 0$, then
$$P(p/q) = [P(p) \times P(q/p)]/P(q).$$

PROOF. Since "p & q" and "q & p" are equivalent, theorem 2 and axiom 4 yield

$$P(q) \times P(p/q) = P(p) \times P(q/p),$$

and the theorem follows by dividing by $P(q)$.

THEOREM 9 (Bayes's theorem). If $P(q) \neq 0$, then

$$P(p/q) = \frac{P(p) \times P(q/p)}{[P(p) \times P(q/p)] + [P(\text{not } p) \times P(q/\text{not } p)]}$$

PROOF. By theorem 8 we have

$$(1)\ P(p/q) = \frac{P(p) \times P(q/p)}{P(q)}\ .$$

But "q" is equivalent to "either p & q or not p & q." The probability of this disjunction is equal, by axioms 3 and 4, to the denominator of the theorem. Thus the theorem follows from equation (1) and theorem 2.

To give these theorems some meaning, let us consider the situation of a physician who has just observed a spot on an X ray of some patient's lung. Let us assume that the physician knows the probability of observing such spots *given* that the patient has tuberculosis and that she also knows the incidence of TB and the incidence of lung spots. Letting "S" stand for "the patient has a lung spot," the known probabilities are $P(S)$, $P(TB)$, and $P(S/TB)$. By applying the inverse probability law our physician can calculate the probability that the patient has TB *given* that he has a lung spot. This is

$$P(TB/S) = [P(TB) \times P(S/TB)]/P(S).$$

Having observed the lung spot, it would be legitimate for our physician to take $P(TB/S)$ as the probability that the patient has TB and to base her decisions on it.

The probability assigned to TB before the spot was observed—$P(TB)$—is called the *prior* probability of TB. The conditional probability, $P(TB/S)$, the physician uses after observing the spot is called the *posterior* probability of TB. (The physician also uses a prior probability for lung spots, but does not obtain a posterior probability for them in this example.) Posterior probabilities can also be used as new priors in further applications of the inverse probability law. For example, if the physician now tests her patient with a TB skin test and obtains a positive reading, she can apply the formula again using the posterior probability for TB as her new prior.

It is likely that a physician will know or can easily find out the probability of observing lung spots *given* that a patient has TB, and it is likely that she can find the probability of a patient's having TB, since there are much medical data concerning these matters. But it is less likely that a physician will have access to data concerning the probability of observing lung spots per se. Then the inverse probability law will not apply, but Bayes's theorem might. For suppose the physician knows that the patient has TB or lung cancer but not both. Then lung cancer (LC) can play the role of not TB in Bayes's Theorem and we obtain:

$$P(TB/S) = \frac{[P(TB) \times P(S/TB)]}{[P(TB) \times P(S/TB)] + [P(LC) \times P(S/LC)]}$$

Example. Suppose that on any given day the probability of rain (R) is .25, that of clouds (C) is .4, and that of clouds given rain is 1. You observe a cloudy sky. Now what are the chances of rain? Using the inverse probability law we obtain:

$$P(R/C) = [P(R) \times P(C/R)]/P(C) = [1/4 \times 1]/(4/10) = 5/8.$$

PROBLEMS

1. Assume a card is drawn at random from an ordinary fifty-two card deck.
   a. What is the probability of drawing an ace?
   b. The ace of hearts?
   c. The ace of hearts or the king of hearts?
   d. An ace or a heart?
2. A card is drawn and not replaced, then another card is drawn.
   a. What is the probability of the ace of hearts on the first draw and the king of hearts on the second?
   b. What is the probability of two aces?
   c. What is the probability of no ace on either draw?
   d. What is the probability of at least one ace and at least one heart for the two draws?
3. $P(p) = 1/2$, $P(q) = 1/2$, $P(p \ \& \ q) = 1/4$. Are p and q mutually exclusive? What is $P(p \text{ or } q)$?
4. A die is loaded so that the probability of rolling a 2 is twice that of a 1, that of a 3 three times that of a 1, that of a 4 four times that of a 1, etc. What is the probability of rolling an odd number?

5. Prove that if "p" implies "q" and $P(p) \neq 0$, then $P(q/p) = 1$. (Hint: "p" implies "q" just in case "p" is equivalent to "p & q.")
6. Prove that $P(p \;\&\; q) \leq P(p)$.
7. Suppose that $P(p) = 1/4$, $P(q/p) = 1$, $P(q/\text{not } p) = 1/5$. Find $P(p/q)$.
8. There is a room filled with urns of two types. Type I urns contain six blue balls and four red balls; urns of type II contain nine red balls and one blue ball. There are 800 type I urns and 200 type II urns in the room. They are distributed randomly and look alike. An urn is selected from the room and a ball drawn from it.
   a. What is the (prior) probability of the urn's being type I?
   b. What is the probability that the ball drawn is red?
   c. What is the probability that the ball drawn is blue?
   d. If a blue ball is drawn, what is the (posterior) probability that the urn is of type I?
   e. What is it if a red ball is drawn?
9. Suppose you could be certain that the urn in the last example is of type II. Explain why seeing a blue ball drawn from the urn would not produce a lower posterior probability for the urn being of type II.

3-2a. Bayes's Theorem without Priors

Suppose you are traveling in a somewhat magical land where some of the coins are biased to land tails 75% of the time. You find a coin and you and a friend try to determine whether the coin is biased. There is a certain test using magnets that will tell whether the coin is biased, but you do not have any magnets. So you flip the coin ten times. Each time the coin lands tails up. Can you conclude that the coin is more likely to be biased than not?

It would seem natural to try to apply Bayes's theorem (or the inverse probability law) here. But what is the prior probability that the coin is biased? Of course, if you knew that, say, 70% of the coins in this land are biased and that your coin was "randomly" selected, it would be reasonable for you to use .7 as your prior. However, as far as the story goes, you know no such thing.

Some statisticians and decision theorists claim that in a situation such as this you should take your best hunch as the prior probability and use it to apply Bayes's theorem. These people are known as *Bayesians*. This is not only because they desire to use Bayes's theorem (when other statisticians believe it should not be applied), but also because they have constructed an argument in support of their position that is itself based on Bayes's theorem.

In brief, their argument is this. Suppose you (or a group of individuals) come to a situation in which you use your best hunch (or hunches) to estimate the prior probability that some statement p is true. Next suppose you are exposed to a large amount of data bearing on the truth of p and you use Bayes's theorem or the inverse probability law to generate posterior probabilities for the truth of p, taking the posterior probabilities so yielded as their new priors, and repeat this process each time you receive new data. Then as you are exposed to more and more data your probability estimates will come closer and closer to the "objective" or statistically based probability – if there is one. Furthermore, if several individuals are involved, their several (and possibly quite different) personal probability estimates will converge to each other. The claim is that, in effect, large amounts of data bearing on the truth of p can "wash out" poor initial probability estimates.

I have put this argument in an overbold and imprecise form. Unfortunately, a careful formulation would require more mathematics than is appropriate for this book. Perhaps we can illustrate the phenomenon with which the argument is concerned by returning to our example. Recall that you have flipped the coin ten times and each time it has come up tails. These data favor the biased coin, since it is biased to produce tails on three tosses out of four. Now suppose that before you decided to toss the coin you assigned a probability of .01 to its being biased (B) and a probability of .99 to its being not biased (not B). Each toss of the coin is independent of the others, so the probabilities of ten tails in a row conditional on either coin are:

$$P(10 \text{ tails}/B) = (3/4)^{10}$$
$$P(10 \text{ tails/not } B) = (1/2)^{10}.$$

Now instead of calculating $P(B/10 \text{ tails})$ and $P(\text{not } B/10 \text{ tails})$, let us calculate the ratio of the latter to the former. Using the inverse probability law we obtain:

$$\frac{P(B/10 \text{ tails})}{P(\text{not } B/10 \text{ tails})} = \frac{[P(B) \times P(10 \text{ tails}/B)]/P(10 \text{ tails})}{[P(\text{not } B) \times P(10 \text{ tails/not } B)]/P(10 \text{ tails})}$$
$$= \frac{P(B) \times P(10 \text{ tails}/B)}{P(\text{not } B) \times P(10 \text{ tails/not } B)}$$
$$= (.01 \times (3/4)^{10})/(.99 \times (1/2)^{10})$$
$$= (1/99)(3/2)^{10}.$$

This is approximately 57.7, which means that the probability you assigned to the coin's being biased has gone from ninety-nine times smaller than that of its being unbiased to almost fifty-eight times larger. If you had flipped it a hundred times and had gotten a hundred tails, the probability of its being biased would be over 3,000 times larger than its being not biased.

Of course, we have looked at one of the simplest and most favorable cases. More complicated mathematics is required to analyze the cases in which some proportion of the tosses are heads while the balance are tails. It is not hard to show, for example, that if 1/4 of the tosses turned out to be heads, that you would assign a higher probability to the coin's being biased, and that the probability would increase as the number of tosses did.

PROBLEMS

1. Calculate the posterior probability that the coin is biased given that you flip the coin ten times and observe eight tails followed by two heads. [$P(B) - .01$.]
2. Calculate the posterior probability that the coin is biased given that you flip

it ten times and observe any combination of eight tails and two heads. [$P(B) = .01$.]

3. Suppose you assigned a probability of 0 to the coin's being biased. Show that, no matter how many tails in a row you observed, neither Bayes's theorem nor the inverse probability law would lead to a nonzero posterior probability for the coin's being biased.

3-2b. Bayes's Theorem and the Value of Additional Information

Another important application of Bayes's theorem and the inverse probability law in decision theory is their use to determine the value of additional information. It is a commonplace that having more facts on which to base a decision can make a radical difference to our choices. But how can we determine how much those facts are worth? The basic idea for a decision theoretic answer is this. In decisions under risk the choices we make are a function of the values we assign to outcomes and the probabilities we assign to states. As we obtain more information we often revise our probability assignments and, consequently, the choices made on that basis. Additional information may save us from serious mistakes, or it may leave our decisions unchanged. Often, by using Bayes's theorem or the inverse probability law, we can calculate how our probabilities and decisions would change if we had an additional piece of information and how much our expectations would be raised. The latter may be used to determine upper bounds on the value of that information. There are many applications of this technique ranging from evaluating diagnostic tests in medicine to designing surveys for business and politics to the evaluation of scientific research programs. The following example is a simple illustration of the method used.

Clark is deciding whether to invest $50,000 in the Daltex Oil Company. The company is a small one owned by some acquaintances of his, and Clark has heard a rumor that Daltex will sell shares of stock publicly within the year. If that happens he will double his money; otherwise he will earn only the unattractive return of 5% for the year and would be better off taking his other choice—buying a 10% savings certificate. He believes there is about an even chance that Daltex will go public. This has led him to decision table 3-3. Clark makes his

3-3	Daltex Goes Public	Does Not
Invest in Daltex	$100,000 .5	$52,500 .5
Buy a Savings Certificate	$55,000 .5	$55,000 .5

decisions on the basis of expected monetary values; so he has tentatively decided to invest in Daltex since that has an EMV of $76,250. Suppose he could pay some completely reliable person to tell him now whether Daltex will go public.

As things now stand, he is looking at an EMV of \$76,250. But if he learned that Daltex was certain to go public, his EMV would increase to \$100,000, whereas if he learned that it was certain not to go public, he would buy the savings certificate and decrease his EMV to \$55,000. As of now, he believes that he has an even chance of learning either piece of information. Thus his expectations prior to paying the completely reliable person for the truth about Daltex are \$100,000(.5) + \$55,000(.5) or \$77,500. This is an increase of \$1,250 over his current expectation of \$76,250. Learning the truth about Daltex in order to revise his decision is thus not worth more to him than \$1,250.

So far we have not used Bayes's theorem or the inverse probability law, so let us change the example. Now let us suppose that Clark knows Daltex is preparing a confidential annual report and he also knows that if they are going public there is a chance of .9 that they will say so in the report and only a .1 chance that they will deny it. On the other hand, if they are not going public there is a chance of .5 that they will say that they are not and .5 chance that they will lie and say they are.

Clark knows someone in Daltex who will show him a copy of the report—for a price. So he decides to use Bayes's theorem to calculate the probabilities that Daltex will (will not) go public given that they say (deny) that they will in the report. Where P stands for their going public, Y stands for their saying they will, and D stands for their denying it, he obtains

$$P(P/Y) = \frac{P(P) \times P(Y/P)}{P(P) \times P(Y/P) + P(\text{not } P) \times P(Y/\text{not } P)}$$

$$= \frac{.5 \times .9}{.5 \times .9 + .5 \times .5} = .64+.$$

Similarly, $P(\text{not } P/Y) = .35+$; $P(P/N) = .16+$; $P(\text{not } P/N) = .83+$. Clark then considers two revised decision tables—one based on the probabilities he would use after reading an affirmation of going public, the other based on those he would use after reading a denial. He finds that on either version the EMV of investing in Daltex would still be higher than that of buying the savings certificate. So he decides not to offer the bribe.

By changing the example appropriately we can arrange for Clark to discover that reading a denial in the report would change his original decision. He would then proceed as in the first version of the example to calculate the gains he can expect under the two scenarios and average these to find the maximum price he should pay to see the report.

As the example illustrates, applications of this method can become quite complicated. When nonmonetary values are involved, as happens in medicine or science, it may also be necessary to enlist some advanced ideas in utility theory. So I will not discuss the topic of the value of additional information further in this book.

PROBLEMS

1. Suppose $P(Y/P) = .01$, $P(N/P) = .99$, $P(Y/\text{ not } P) = .5$, and $P(N/\text{ not } P) = .5$. How much should Clark be willing to pay to see the report?
2. A closet contains 800 type I urns and 200 type two urns. Urns of both types appear identical but type I urns contain six blue balls and four red ones; type II urns contain one blue ball and nine red ones. An urn is drawn at random from the closet and you must bet on the type of the urn. If you bet on type I and it is one, you win \$20, otherwise you lose \$10. If you bet on type II and it is one, you win \$80, otherwise you lose \$10. Assume that you maximize expected monetary values.
   a. Set up a decision table for the choice between the two bets and calculate the EMVs of the two bets. Which one would you choose?
   b. Prior to making your choice, what is the maximum amount you should pay to learn the type of the urn?
   c. Assume that a blue ball has been drawn from the urn. Appropriately revise your table and calculate the new EMVs. Which bet would you choose now?
   d. Assume that the ball drawn is red and and then follow the rest of the instructions for c.
   e. Prior to seeing the ball and making your choice, what is the maximum amount that you should pay to see it?

3-2c. Statistical Decision Theory and Decisions under Ignorance

Several of the rules for making decisions under ignorance were originally developed by statisticians for handling statistical problems in which some of the data necessary for applying Bayes's theorem are unavailable. I will illustrate their thinking with a typical problem in applied statistics—the predicament faced by a drug company that needs to test a new batch of pills prior to marketing them. Let us suppose that we own such a company and our laboratory staff has just manufactured a new batch of pills. Their usual practice is to test new lots by feeding some of the pills to a hundred rats and observing how many die. From their previous experience with this and similar drugs, they have developed good estimates of the percentage of rats that will die if the batch is defective. Unfortunately, due to an irregularity in making this particular batch, our staff has no idea of the probability of its being defective. Suppose the test is run and five rats die. Since the rats in our laboratory die from time to time from various other causes, the staff cannot be certain that the pills are defective. Can they use the test results to calculate the probability of the batch's being defective? They could use Bayes's theorem if they could assign a probability to the batch's being defective *prior* to running the test. But that is exactly the probability they cannot assign.

Bayesian statisticians would urge our staff to try to use their best hunches as the prior probability that the batch is defective. And they could even offer some methods for refining and checking these hunches.

Many statisticians dissent from this recommendation and urge that the problem be treated as one under ignorance. But it is not as simple as a two-act/two-state choice of marketing or not marketing the pills against the possibilities of the batch's being fine or being defective for that would not respond to the test results. They have proposed that prior to running the test we should choose between various strategies for responding to the test results. For instance, we might adopt the strategy of marketing the pills no matter how many rats die or marketing them if no more than two die or not marketing them at all. Then after the test is run we market or withhold the pills according to the prescriptions of the strategy we have selected. This allows the test results to influence our actions without depending on guesses as to the prior probability that the batch is defective.

But how do we choose a strategy? We presumably know the value of marketing defective (or good) pills and the value of withholding them. Also we can use the probabilities that various percentages of the rats die given that the batch is defective (or fine) to calculate the various expected values of our strategies under the assumption that the batch is defective, and we can make similar calculations under the assumption that it is fine. Using this we can form decision table 3-4. Finally, we can apply one of the rules for making decisions under ignorance to this table.

3-4	Batch Defective	Batch Fine
Strategy 1	x	y
Strategy 2	z	w
	$\vdots$	$\vdots$
Strategy n	u	v

Now this certainly does not settle the philosophical and methodological issues raised by the absence of prior probabilities. For we must once again face the question of the proper rule to use for making our decision under ignorance. We will not pursue that question further here, although we will examine more closely the Bayesian case for the use of subjective priors.

PROBLEMS

1. Suppose you formulate strategies as depending on whether fewer than ten rats die or whether ten or more die. This yields four strategies. Two are: (1) market if fewer than ten die; market if ten or more die; (2) market if fewer than ten die; withhold if ten or more die. List the remaining two.
2. Suppose that the probability that ten or more rats die given that the batch is defective is .5 and that it is .01 given that the batch is fine. Construct two tables—one assuming that the batch is defective, the other that it is fine—that

will tabulate your probabilities of marketing and withholding the pills under each strategy. (For example, in the row for strategy 2 of the first table there is a probability of .5 that you will market and one of .5 that you will withhold.) These tables give your *action probabilities*.

3. Now assume that your value for marketing a defective batch is $-1{,}000$, for marketing a fine one is 100, for withholding a defective one is 0, and for withholding a fine one is -100. Use these values and your action probabilities to calculate the expected values for each strategy under the assumption that the batch is defective. Do the same under the assumption that it is fine. (For example, the first value for the second strategy is -500 or $[.5][-1000]+[.5][0]$.) This should enable you to complete table 3-5. If

3-5	Batch Defective	Batch Fine
Strategy 1		
Strategy 2	-500	
Strategy 3		
Strategy 4		

you use the maximin rule, which strategy do you choose? Do you think this is the only reasonable choice to make in this situation?

3-3. Interpretations of Probability

There are many contending views concerning what probability statements mean, when and how they may be applied, and how their truth may be ascertained. I will not pretend to resolve the complicated and heated debate that surrounds the various views on probability. Here I will only review several of the major views and discuss some of the objections that have been raised against them.

I will classify the interpretations of probability as *objective* or *subjective*. The *objective* interpretations see probability as measuring something independent of human judgments, and constant from person to person. The *subjective* views regard probability as the measure of an individual's belief or confidence in a statement and permit it to vary from person to person. Objective views are further classified as *logical* or *empirical*, according to whether they count probability as a property defined in terms of logical or mathematical structures or as an empirically defined property. To illustrate these distinctions, suppose I assert the statement:

> The probability that 2/3 of the next 100 tosses of this coin will land head up is .75.

Subjective views will construe this as meaning that I am reasonably confident that 2/3 of the tosses will result in heads. Logical views will see it as reflecting a logical or mathematical analysis of the various possible tosses and (perhaps) my evidence concerning them. Finally, the empirical views will see my claim as about the behavior of the coin and as testable in terms of it.

Until quite recently probability theorists maintained that a satisfactory interpretation of probability must satisfy the probability calculus. The views I will discuss do, and I will show this by verifying that each interpretation satisfies the axioms of the calculus. Since the theorems of the calculus follow logically from the axioms, any interpretation that makes the latter true must verify the former as well.

3-3a. The Classical View

The classical interpretation of probability, also known as the Laplacean view after one of its founders, is the oldest and simplest view of probability. It is an objective and logical view, which is best applied to games of chance and other clearly specified situations that can be divided into a number of equally likely cases. We have used it implicitly in illustrating the probability calculus, since most of our examples have concerned random card drawings, tosses of fair coins, and rolls of unloaded dice, where one can reasonably assume that each card has an equal chance of being drawn and that each face of the coin or die has an equal chance of landing up.

To state the view in its general form, let us think of each statement as having a finite set of possibilities associated with it. For example, the statement "The coin will land heads in at least one of the next two tosses" is associated with the four possible outcomes of tossing the coin twice (*HH, HT, TH, TT*). Some of the possibilities associated with a statement verify it, others falsify it. Thus the possibility of getting two heads (*HH*) verifies the statement about the coin, while the possibility of getting two tails (*TT*) falsifies it. Given a statement p and the possibilities associated with it, let us call those that verify it the *p-cases*. Then the classical view may be put as the claim that the probability of p is the ratio of the number of p-cases to the total number of cases or possibilities:

$$P(p) = \#(p\text{-cases})/\#(\text{total possibilities}).$$

We must interpret conditional probability too, since it figures in the axioms of the calculus. Ordinarily, $P(q/p)$ is the number of p-cases that are also q-cases. However, when there are no p-cases, it is zero. This leads to:

$$\begin{aligned} P(q/p) &= \#(p \ \&\ q\text{-cases})/\#(p\text{-cases}) \text{ if } \#(p\text{-cases}) > 0, \\ &= 0 \text{ if } \#(p\text{-cases}) = 0. \end{aligned}$$

To see that this works, consider the probability that the card you have drawn is an ace given that it is a heart. This is just the ratio of the number of aces of hearts to the number of hearts, that is 1/13.

Before discussing the philosophical objections to the classical view let us verify that it does satisfy the probability calculus. That comes to showing that each axiom of the calculus becomes true when interpreted by construing "$P(p)$" and "$P(q/p)$" as previously defined. This is easy to see in the case of axiom 1. For the number of possibilities associated with a statement is never negative and the number of p-cases (p & q-cases) never exceeds the total number of cases (the number of p-cases). Thus $P(p)$ [$P(q/p)$] must be a number between 0 and 1 inclusively.

Axiom 2 is easily verified too. A certainty is true no matter what; thus the cases in which it is true must be identical with all the cases associated with it and the ratio of the number of the one to that of the other must be 1.

Turning now to axiom 3, we must remember that in the probability calculus "p or q" is construed as meaning that either p is true, q is true, or *both* p and q are true. The probability of "p or q" is then the ratio of the number of (p or q)-cases to the total number. If p and q are mutually exclusive (as the condition on axiom 3 states), the (p or q)-cases are simply the cases in which either p or q (but not the other) is true. Thus we have:

$$\begin{aligned} P(p \text{ or } q) &= \#[(p \text{ or } q)\text{-cases}]/\#(\text{total cases}) \\ &= \#(p\text{-cases})/\#(\text{total cases}) + \#(q\text{-cases})/\#(\text{total cases}) \\ &= P(p) + P(q), \end{aligned}$$

which verifies axiom 3.

This leaves axiom 4. To verify it let us distinguish two cases: those p for which there are no p-cases and those for which there are some. In the first case, both $P(p)$ and $P(q/p)$ are 0. Furthermore, since every p & q-case is a p-case, $P(p \;\&\; q)$ is 0 too. Thus

$$P(p \;\&\; q) = 0 = P(p) \times P(q/p),$$

which verifies axiom 4 for this case.

In the second case,

$$P(q/p) = \#(p \;\&\; q\text{-cases})/\#(p\text{-cases}),$$

but we also have

$$\begin{aligned} P(p \;\&\; q) &= \#(p \;\&\; q\text{-cases})/\#(\text{total cases}) \\ P(p) &= \#(p\text{-cases})/\#(\text{total cases}). \end{aligned}$$

Whence we obtain

$$\begin{aligned} P(p) \times P(q/p) &= \#(p\text{-cases})/\#(\text{total cases}) \times \#(p \;\&\; q\text{-cases})/\#(p\text{-cases}) \\ &= \#(p \;\&\; q\text{-cases})/\#(\text{total cases}) \\ &= P(p \;\&\; q). \end{aligned}$$

This establishes that axiom 4 holds for both of the cases we distinguished and completes the demonstration that the classical interpretation satisfies the probability calculus.

PROBLEMS

1. Use the technique of this section to show directly that theorem 1 holds under the classical interpretation.
2. Do the same for theorem 2.

Now let us turn to the objections to this approach. Some can be overcome by enlisting the technical machinery of modern logic and mathematics, but others remain obstacles to the acceptance of even up-dated versions of the classical view.

Taking up a relatively technical problem first, the present version assumes that a definite finite set of possibilities is associated with each statement. That seems a simple enough matter when we are talking about tossing coins, spinning roulette wheels, or drawing cards. But what happens when we are talking about the probability of another world war, being successful in a career or marriage, or even of a drought? What are the relevant possibilities here? And how do we combine the possibilities needed for compound statements? How do we relate the probability that you will have a good career and the probability that you will have a happy marriage to the probability that you will have both? The classical view has no ready answers to these questions.

The situation is worse: The classical view even has problems with the simple questions which it was designed to handle. For example, how do we compute the probability of getting two heads in two tosses of the same fair coin? One computation could run as follows: There are four cases: *HH*, *HT*, *TH*, and *TT*; only one is verifying, so the probability is 1/4. *But there is another computation*: There are three cases: *HH*, *HT*, and *TT*; thus the probability is 1/3. Of course, we all know that the first computation is the right one, because it is based on the relevant set of possibilities. *But the classical view does not tell us that.* (By the way, the possibilities in the second set are mutually exclusive and exhaustive, so we cannot fault it on those grounds.)

In recent years logicians have partially solved the problem of specifying sets of relevant possibilities. They have used precise, limited, and technical languages for formulating those statements to which probabilities are to be assigned and then have defined the associated sets of possibilities in terms of statements within such languages. For example, to describe the tosses of a coin we can introduce a language with two individual constants, "*a*" and "*b*"—one to designate the side that comes up on the first toss, the other to designate the one that comes up on the second toss—and one predicate "*H*" for heads. (Tails can be expressed as "not *H*.") Then the relevant set of possibilities for two tosses of the coin can be specified as those statements of the language that affirm or deny that the coin comes up heads on either toss. These yield the four possibilities: *Ha*, *Hb*; *Ha*, not *Hb*; not *Ha*, *Hb*; not *Ha*, not *Hb*.

This method is clearly limited by our ability to construct artificial languages and specify possibilities within them. We would feel that limitation acutely were we to try to use the classical approach to assign probabilities to statements involving an indefinite or infinite number of cases. Try, for instance, to assign a probability to the statement that there will be an atomic holocaust *sometime*. It is not inconceivable, however, that technical developments in logic will solve such problems eventually.

There is another objection that technical developments cannot avert. The classical approach and its modern descendants depend on the assumption that each of the possibilities associated with a statement is *equally likely*. This is used, for example, in assigning a probability of 1/2 to getting heads on a toss of a fair coin, since the calculation assumes that we should give equal weight to getting heads and to getting tails. But how do we know that each case is equally likely? And what does "equally likely" mean in this context? We cannot appeal to the classical view and say " 'equally likely' means that they have the same probability" without involving ourselves in a circle.

Two courses are open to the classical view. The first consists in invoking some other conception of probability to explain and justify the assumption that each case is equally likely. Thus one might state that the claim that getting heads is just as likely as getting tails means that in a long series of tosses of the coin, the proportion of heads will be approximately the same as the proportion of tails. Somebody who took this position might then continue by adding that the classical view is, strictly speaking, an idealization of a more properly experimental approach to probability. Since ratios determined experimentally are often "messy," the classical approach is to be used as a shortcut approximation.

Some adherents of the classical view have offered a second response. They have claimed that the assignment of the same weight to each possibility is simply an assumption that is so fundamental that further attempts to justify it are fruitless. Some have even gone so far as to claim that this assumption—which they have dubbed the *principle of insufficient reason*—is self-evident. I hope that the relationship of this principle to the homonymous rule for decisions under ignorance is clear. I find the principle no more compelling in the context of probability theory than I found its relative in the earlier context.

3-3b. The Relative Frequency View

One trouble with the classical view is that it is devoid of empirical content. Because its probabilities are ultimately ratios involving abstract possibilities, a classically interpreted probability statement has no implications concerning actual events and can be neither confirmed nor refuted by them. The relative frequency view is an objective and empirical view that was developed in response to this need; it defines probability in terms of actual events.

To state this view with some precision we must assign probabilities to either classes, kinds, or properties of events rather than to statements. Furthermore, we must view all probabilities as implicitly conditional. This entails modifying our presentation of the probability calculus by replacing statements of the respective forms

$P(S) = a$ and $P(S/W) = b$

with ones of the forms

$P_R(S) = a$ and $P_R(S/W) = b$,

and reading these as

the probability that an R is an S equals a;
the probability that an R is an S given that it is a W equals b.

With this modification probability statements no longer make assertions about statements but instead make assertions about classes, properties, or kinds of events. For example, instead of saying the probability assigned to the statement that this coin will land heads on the next toss is 1/2, the frequentist says the probability that a toss of this coin (an event) is one in which it lands heads (another event) is 1/2. This variation on our original approach to the probability calculus is not sufficient to count for or against the frequency approach.

Relative frequentists hold that probabilities are proportions or relative frequencies of events of one kind to those of others. Their interpretation of probability is thus

$P_R(P) = a$ means the proportion of Rs that are Ps is a,
i.e., $\#(P \ \& \ R\text{s})/\#(R\text{s}) = a$.

The interpretation of conditional probability is just

$$P_R(Q/P) = \#(P \ \& \ Q \ \& \ R\text{s})/\#(P \ \& \ R\text{s}).$$
$$= 0 \text{ if nothing is both a } P \text{ and an } R.$$

To illustrate this conception of probability consider the statements:

1. The probability that an airplane flying from New York to London crashes is 1/1,000,000
2. The probability that an airplane flying from New York to London crashes given that it has engine failure is 1/10.

The first predicts that if we were to inspect the record of flights from New York to London we would find that only one in a million crashes, while the second predicts that if we were to keep a record of those flights from New York to London that also experienced engine failures, we would find that the crash rate increased to 1 in 10. Plainly, this approach to probability is very different from the classical approach.

Verifying that the relative frequency interpretation satisfies the axioms of the probability calculus follows the model set earlier for the classical view. The relative frequency interpretation of $P_R(P)$ is concerned with the proportion of events of kind P among those of kind R; it is thus a ratio between 0 and 1 inclusively. Since the same is easily shown for $P_R(Q/P)$ as well, the interpretation satisfies axiom 1. Turning to axiom 2, if every R is certain to be a P, then the ratio of Ps to Rs is 1 and $P_R(P) = 1$.

To verify axiom 3 we must show that

$$P_R(P \text{ or } Q) = P_R(P) + P_R(Q)$$

when no event can be both a P and a Q. But this just means showing that the proportion of Ps among the Rs plus the proportion of Qs among the Rs is just the proportion of the (P or Q)s among the Rs. And that must certainly be the case when no R can be both P and Q.

The verification of axiom 4 parallels the verification of the classical interpretation of that axiom. I leave it as an exercise.

PROBLEMS

1. Verify axiom 4.
2. Do the same for theorem 2.

Given the way in which it has been specified, the relative frequency view is bound to have problems with indefinite and infinite totalities. Even determining the frequency of heads in the tosses of the dime on my desk presents difficulties, for no one can currently specify its total number of tosses. However, mathematical improvements on the relative frequency view can handle cases like this one and give them empirical content. Very roughly, we toss the coin again and again and after each toss note the proportion of heads among the tosses to that point. If these ratios appear to be tending to a limit point, then that limit is identified with the probability of getting heads on a given toss of the coin. We might call this "the long-run frequency" approach to probability.

Either frequency approach has the obvious advantage over the classical view of not being circular. Nor does the frequency approach require dubious assumptions, such as the principle of insufficient reason, and it extends to the indefinite/infinite case with greater ease than does the classical approach. On the other hand, the relative frequency approach has a less firm grip on "true" probabilities than the classical approach. To see what I mean, consider the question again of the frequency of heads in tosses of the dime on my desk. A theorem of the probability calculus entails that if the probability of heads is 1/2, the frequencies of these tosses will converge to 1/2 in the long run. However, another theorem entails that ultimate convergence to 1/2 is compatible with any initial finite sequence of tosses consisting entirely of *heads*. Now suppose my dime were tossed 10,000 times and it came up heads 90% of the time. Most relative frequentists would be willing to stop the tossing, proclaim that the coin is most probably biased and that the true probability of heads is close to 9/10. But it could turn out that after the first 10,000 tosses heads start to appear so often that the long-run frequency is 1/2. On the frequency view, 1/2 would be the true probability, although the initial observations belied it. Putting the point more generally, *there is simply no guarantee that the frequencies observed to date are even close to the long-run frequency.*

The classical view does not face this problem because it cuts itself off from observation. But that does not seem to be much of an advantage either; for if my dime did turn up heads on every one of 10,000 tosses, even fans of the classical view would be hard pressed to justify taking 1/2 as the probability of getting heads. (But they would not be without any defense. They could recheck their analyses and stand fast, since on any view of probability such a run of heads on tosses of a fair coin is *possible* though not very probable.)

Since the frequency view specifies probabilities in terms of proportions, it cannot make sense of assigning probabilities to single events or statements. Thus if you ask a frequentist what is the probability that you will pass your final exam in English, he will respond that he can only speak to you of the percentage of students in the course that will pass, or of the percentage of those tests you take that you pass, or of the percentage of those like you in certain relevant respects who will pass, and so on. But he will refuse to assign a probability to the single case of your passing.

Several of these problems can be averted while retaining an empirical account of probability by turning to the *propensity* interpretations of probability. One version of this view construes probabilities as the frequencies predicted by the relevant theoretical laws. Since predicted frequencies may differ from those actually observed, this account is not thrown by my dime coming up heads on the first 10,000 tosses. Of course, that happening with most dimes would strongly signal something wrong with any theory that predicts a frequency of 1/2, and we would take steps to revise it. The propensity view responds to observation without following it slavishly.

The so-called *single-case propensity* interpretation even countenances assigning probabilities to single events. Consider my dime again. By taking advantage of its symmetry and the laws of physics one should be able to design a device for tossing it that would favor neither heads nor tails. Such a device would, we might say, have a propensity of 1/2 to yield heads on its tosses. More important, however, it would have a propensity of 1/2 to yield a head *on any particular toss*. By identifying probabilities with such single-case propensities we could make sense of assigning probabilities to single events.

Unfortunately, there are limits to the applicability of the propensity approach too. Often we do not know enough to discern propensities. Physicians, for instance, know that heart disease is much more frequent among heavy smokers, but currently they have no way of knowing whether any individual heavy smoker has a higher propensity for developing heart disease than some other one does. Also it does not always make sense to speak of propensities where we can significantly speak of probabilities. Thus wondering about the probability of the truth of the theory of relativity seems to make sense, but wondering about its propensity to be true does not.

3-3c. Subjective Views

The logical approach to probability fails in situations where we lack the analytic resources it presupposes. The frequency approach breaks down on single-case probabilities, whereas the propensity approach fails to cover cases where propensities are not known or do not make sense. The subjective approach to probability is an attempt to develop a notion of probability that meets all these challenges. Subjective probabilities are personal assessments. Since we can have and often do have our own estimates of the probability that something is true even when that thing is a single case or when we lack any theory or logical analysis concerning it, the subjective approach bypasses the impediments to the previous views.

Offhand, however, it would seem that a subjective view of probability would immediately encounter insurmountable difficulties. How can personal assessments be subjected to critical evaluation? How can they produce a concept of probability of use to scientists and decision makers? How can they be measured—at all—or with enough accuracy to furnish appropriate numerical inputs for the probability calculus? In the last sixty years logicians, mathematicians, and statisticians have made remarkable progress toward dealing with these questions.

The connection between belief, desire, and action is well known to psychologists and philosophers. There are boundless illustrations. If you believe your water supply has been poisoned, you will resist attempts to make you drink from it even though you may be quite thirsty. If you cross a street, we can reasonably infer that you want to get to the other side and believe it is safe to cross. Frank Ramsey was the first theorist to use these connections to construct a subjective theory of probability. Ramsey realized that our degrees of belief (or confidence) in statements are connected with certain of our actions—the bets we make. If, for example, you believe a certain horse is very likely to win a race, you are likely to accept a bet at less than even money. The more likely you think the horse is to win, the less favorable odds you will accept. Now if we identify your personal probabilities with the odds you are willing to accept, by asking you about the various odds you would accept, we may be able to measure your personal probabilities. Ramsey managed to parlay this into a full case for subjective probabilities.

We are used to betting on the outcome of events, such as races, football games, or elections. But there is no reason in principle why we cannot bet on the truth of statements too. For instance, instead of betting on Fancy Dancer to win in the third race, I can bet that the statement *Fancy Dancer wins the third race* is true. If I am willing to set odds on enough statements and do so in a certain way, it can be shown that my odds constitute probability assignments to those statements and obey the probability calculus. I will present Bruno DeFinetti's proof of this rather than Ramsey's, since the latter's is intertwined with his treatment of utility.

DeFinetti's reasoning deals with an agent in a situation in which he must place a series of bets on a certain set of initial statements as well as all negations, conjunctions, disjunctions, and conditional bets that can be formed using these statements. Let us suppose, for example, that you are the agent in question and the initial statements are

Jones will win the match.
Smith will win the match.
The crowd will be large.

Then you will be expected to bet not only on those three statements but also on the statements

Jones will not win the match.
Smith will win the match or Jones will win the match.
Jones will win the match and the crowd will not be large.

In addition you will be expected to take conditional bets such as

Jones will win given that the crowd is large,
Jones will not win given that the crowd is large and Smith does not win,

and so on, for all of the infinitely many bets that are constructible from the initial set.

For future reference, let us call the set of statements on which the agent is expected to bet the *DeFinetti closure* of the initial set of statements. Given a set of statements A the DeFinetti closure of A, $DC(A)$, may be formally defined as follows:

1. Every statement in A is also in $DC(A)$.
2. If the statements S and W are in $DC(A)$, so are the statements not S, (S or W) and (S & W).
3. If S and W are in $DC(A)$, so is the phrase S given W.
4. Nothing is in $DC(A)$ unless its being so follows from 1-3.

With the DeFinetti closure behind us let us return to the agent, you. You are expected to place bets on every statement in the DeFinetti closure of your initial set with a "bookie." But here the situation changes dramatically from the usual betting situation. For you must post the odds on *all* the statements and conditional bets in the DeFinetti closure, and that is all you are permitted to do. Once you set the odds, *the bookie determines all the other features of the bet*, including the amount at stake on the various bets and who bets for or against a given statement.

The situation with respect to a particular bet on a single statement p can be summarized by means of table 3-6. The entries under the statement p are

3-6

p	Payoff for p	Payoff against p
T	$(1-a)S$	$-(1-a)S$
F	$-aS$	aS

simply the truth values, true and false; the other entries tell how much you (or the bookie) win or lose for the various outcomes. Thus if you are betting for p and p is false, you "win" $-aS$ and the bookie wins aS. Notice that the entries in a row under "for" and "against" are the negatives of each other, so that the person "for" always wins (or loses) an amount equal to that lost (or won) by the person "against." S is the stake for the bets, which is always some positive amount of money, aS and $(1-a)S$ are portions of the stake, and a is a number, called *the betting quotient* for p, the ratio of a to $1-a$ constitutes the odds you set for the statement p. (Note that $aS+(1-a)S=S$.) When you set the odds for the statement p at a to $1-a$, you must be prepared to lose a portion a of the stake S if p turns out to be false. The higher you set the betting quotient a, the greater portion of the stake you risk losing; so, presuming you are rational and prudent, you will not set high odds for a statement unless you are quite confident it is true.

The only feature of table 3-6 you control is the odds a to $1 - a$. The bookie not only fixes the amount S at stake but also decides who is to be for p and who is to be against p. (However, he cannot bet for p and also against p, nor can he force you to do so.) To complicate matters, remember that you must post odds for many statements and take many other bets at the terms the bookie sets.

Before considering the rest of DeFinetti's argument, let us use an example to relate our concepts to conventional betting odds. Suppose the Golden Circle Race Track posts odds of 99 to 1 on Fancy Dancer to win. This means that if the horse wins, the track will pay us \$100 for every \$1 "to win" ticket for that horse that we have purchased. The total stake for a \$1 ticket is \$100. We risk 1/100 of it, the track risks 99/100 of it. In our terms the track's betting quotient is 99/100, and the odds are 99/100 to 1/100. More generally, suppose that in a conventional betting situation someone offers odds of a to b on a given outcome. Then they are willing to risk losing a portion $a/(a+b)$ of a stake in case the outcome fails to obtain so long as we are willing to risk losing the portion $b/(a+b)$ in case the outcome does obtain. In our terms the odds are $a/(a+b)$ to $b/(a+b)$ and the betting quotient is $a/(a+b)$.

Returning to DeFinetti's work, suppose you were in the kind of situation with which DeFinetti is concerned. Then a clever bookie might be able to arrange the bets so that he was bound to have a net gain no matter what happened. For instance, suppose you posted odds of 9/10 to 1/10 on a statement p and odds of 1/2 to 1/2 on its negation not p. Although the bookie cannot force you to bet for p and also against p, he can force you to bet for p and for not p. Suppose that he does and fixes the stake at \$1. Then if p is true, he loses \$.10 on his bet against p and wins \$.50 on his bet against not p. That is a net gain of \$.40. (Check and you will see that you will have a net loss of \$.40.) On the other hand, if p is false, the bookie wins \$.90 on his bet against p and loses \$.50 on his bet against not p—again a net gain of \$.40. The bookie has made a *Dutch Book* against you. You could have prevented this by posting odds of 1/10 to 9/10 on not p or odds of 1/2 to 1/2 on p, or any other combination of odds under which your betting quotients for p and not p summed to 1. In short, so long as your betting quotients for p and not p sum to 1, you have protected yourself against this type of Dutch Book. DeFinetti generalized this to prove the following:

> DUTCH BOOK THEOREM. *Suppose that no Dutch Book can be made against an agent using the odds he posts on the DeFinetti closure of set of statements. Then his betting quotients for the DeFinetti closure in question satisfy the probability calculus.*

This means that the agent's betting quotients form a satisfactory interpretation of the probability calculus.

To establish the Dutch Book theorem, we will set

$P(p)=$ the agent's (your) betting quotient for $p=a$

and verify that $P(p)$ (i.e., a) satisfies the axioms of the probability calculus provided that no Dutch Book can be made against you.

Turning to axiom 1a, we must show that if no Dutch Book can be made against you,

$$0 \leq a \leq 1.$$

We will use an indirect proof to do this, however. We will first assume that $a < 0$ and show how to make a Dutch Book against you, and then assume that $a > 1$ and again show how to make a Dutch Book against you. This will be our general strategy: In each case, we will show that if your betting quotients violate one of the axioms of the probability calculus, a Dutch Book can be made against you.

Suppose then that $a < 0$. Then $-a > 0$, $(1-a) > 1$, and both are positive. Now reconsider table 3-6 for the bets involving p. Since both the payoffs $-aS$ and $(1-a)S$ are positive, the bookie can guarantee himself a net gain by betting for p at any positive stake S. (For simplicity, we will assume that the stake is 1.) Whether p is true or false, his payoffs are positive. Yours are negative, since you must bet against p.

On the other hand, if $a > 1$, then $1 - a < 0$ and both $-(1-a)$ and a are positive. So the bookie can make a Dutch Book against you by betting against p. His payoffs, being in the column under "Payoff against p," are bound to be positive.

Let us deal with axiom 2 next. Let us suppose that p is certain and show that if your betting quotient for p is less than 1, a Dutch Book can be made against you. (We have already established that it cannot be greater than 1.)

Since p is certain, we know that the bottom row of table 3-6 will never apply. So we need only consider table 3-7. If $a < 1$, then $(1-a)$ is positive;

3-7 p	Payoff for p	Payoff against p
T	$(1-a)$	$-(1-a)$

clearly the bookie can make a Dutch Book against you by betting for p.

Stepping up in difficulty, let us turn to axiom 3. Here we are concerned with the probabilities of p, q, and their disjunction "p or q." Let a be your betting quotient for p, b the one for q and c that for "p or q." We must now use the next betting table (table 3-8) with three bets—one for each betting quotient. The payoffs under p and q are determined by referring to the columns under "p" and "q" to determine whether they are true or false, and then applying our original table (3-6) for a single statement. The payoffs under "p or q" are determined in the same way, but since it is true in the first three rows, its payoffs are the same in those rows. I have set the stakes at 1 throughout table 3-8.

3-8

		p		q		p or q	
p	q	For	Against	For	Against	For	Against
T	T	$1-a$	$-(1-a)$	$1-b$	$-(1-b)$	$1-c$	$-(1-c)$
T	F	$1-a$	$-(1-a)$	$-b$	b	$1-c$	$-(1-c)$
F	T	$-a$	a	$1-b$	$-(1-b)$	$1-c$	$-(1-c)$
F	F	$-a$	a	$-b$	b	$-c$	c

Axiom 3 states that if p and q are mutually exclusive, $P(p \text{ or } q) = P(p) + P(q)$. Thus we must show that if p and q are mutually exclusive and $c \neq a+b$, a Dutch Book can be made against you. Let us assume then that p and q are mutually exclusive. This means that the first row of betting table 3-8 never applies and can be ignored. Also assume that $c \neq a+b$. Then either $c < a+b$ or $c > a+b$. I will show how to construct a Dutch Book against you for the first case and leave the second case to you.

Since $c < a+b$, $(a+b)-c$ is positive. If the bookie bets against p and against q but for "p or q" his total payoffs for the last three rows of the table all equal $(a+b)-c$. (In the second row he is paid $-(1-a)$, b, $1-c$; these sum to

$$-1+a+b+1-c=(a+b)-c.$$

Check the other rows.) Thus by betting as indicated, the bookie can guarantee himself a positive net gain no matter what the truth values of p and q turn out to be.

Before we can handle axioms 1b and 4, we must interpret $P(p/q)$ in terms of betting quotients. This involves the use of *conditional bets*, such as a bet you might make that a horse will win *given* that the track is dry. The bet is off if the track is not dry and nobody wins or loses. Similarly, we will construe a bet on "q given p" as on only when p is true, and then as won according to whether q is true or false. Using odds of a to $1-a$ for the conditional bet "q given p," this leads to table 3-9—(we use payoffs of 0 to handle cases in which the bet is off).

3-9

p	q	For "q given p"	Against "q given p"
T	T	$1-a$	$-(1-a)$
T	F	$-a$	a
F	T	0	0
F	F	0	0

It is easy to see that if $a < 0$, the bookie can make a Dutch Book against you by betting for "q given p." It is also easy to show that he can make a Dutch Book against you if $a > 1$. Thus we know that axiom 1b must hold if we interpret "$P(q/p)$" as your betting quotient for "q given p."

Axiom 4 states that $P(p \,\&\, q) = P(p) \times P(q/p)$. To establish it, assume your betting quotient for "p & q" is c, that for p is a, and that for "q given p" is b. Here is the relevant betting table (3-10). (There are zeros under "q given p" because it is a conditional bet.)

3-10

		p		q given p		p & q	
p	q	For	Against	For	Against	For	Against
T	T	$1-a$	$-(1-a)$	$1-b$	$-(1-b)$	$1-c$	$-(1-c)$
T	F	$1-a$	$-(1-a)$	$-b$	b	$-c$	c
F	T	$-a$	a	0	0	$-c$	c
F	F	$-a$	a	0	0	$-c$	c

We want to prove that if $c \neq ab$, a Dutch Book can be made against you. But no payoff is the product of any of the other payoffs, so our previous strategy for making Dutch Books does not seem applicable. But remember that the bookie is free to choose the stake for each bet. Until now we have let this equal 1 for the sake of simplicity. But now let us have the bookie set the stake at b for the bets on p. That changes the payoffs under p to

$$\begin{array}{ll} (1-a)b & -(1-a)b \\ (1-a)b & -(1-a)b \\ -ab & ab \\ -ab & ab. \end{array}$$

Now suppose $c \neq ab$. Then as before $c < ab$ or $c > ab$. Suppose that the first case holds. Then $ab - c$ is positive. So if the bookie bets against p, against "q given p," and for "p & q," he will be paid $ab - c$ no matter what. (Check this.) This means that he can make a Dutch Book against you. The case where $c > ab$ is handled similarly and is left to you as an exercise.

PROBLEMS

1. Return to the Dutch Book argument for axiom 3 and complete the case for $c > a + b$ that I left as an exercise.
2. Construct Dutch Book arguments to show that $0 \leq a \leq 1$, where a is your betting quotient for "q given p."
3. Carry out the check I asked you to make in the argument for axiom 4.

4. Complete the argument for axiom 4 for the case when $c > ab$.
5. Give a Dutch Book argument to establish directly each of the following:
   a. If p is impossible, $P(p) = 0$.
   b. $P(p) + P(\text{not } p) = 1$.
   c. If p logically implies q, $P(p) \leq P(q)$.

3-3d. Coherence and Conditionalization

DeFinetti called an agent's betting quotients *coherent* in case no Dutch Book can be made against him. We can summarize DeFinetti's theorem as showing that if an agent's subjective probabilities (betting quotients) are coherent, they obey the probability calculus. Now coherence is a very plausible condition of (idealized) rationality, since few of us think that it would make sense to place ourselves in a position where we were bound to suffer a net loss on our bets. If we accept coherence as a condition of rationality, DeFinetti's conclusion can be paraphrased as *the laws of probability are necessary conditions for rationality.*

Since DeFinetti first proved his theorem others have proved that having subjective probabilities that obey the probability calculus is also sufficient for coherence. In other words, if an agent picks his betting quotients so as to satisfy the laws of probability, no Dutch Book can be made against him. This is called the *converse Dutch Book theorem.* I will leave it as additional reading for those of you who are studying to be bookies. (Note: We proved that if the agent did *not* pick his betting quotients thus, a Dutch Book can be made against him.)

Rational people modify their degrees of belief (subjective probabilities) in the light of new data. Whereas it would be reasonable for me to be quite confident that my spanking new car will not break down before I get it home, I would be much less confident about this if someone told me my new car was actually a "lemon" that had just been returned to the dealer. In view of this we might ask whether there are any rules for modifying subjective probabilities in the light of new data. One very popular proposal is that we use the following *rule of conditionalization*: Suppose that D is the conjunction of the statements describing the new data. Then for each statement p, take $P_N(p) = P_O(p/D)$. In other words, let the new probabilities be the old probabilities conditional on the data. Notice that since $P(D/D) = 1$, this rule has the effect of assigning the probability of 1 to the data.

If we replace "$P(p)$" with "$P(p/D)$" and "$P(q/p)$" with "$P(q/p \,\&\, D)$" in every axiom of the probability calculus, the formulas we obtain will be laws of probability too. This means that if we use the rule of conditionalization to modify our subjective probabilities, the new ones will obey the laws of probability too, and thus by the converse Dutch Book theorem will be coherent. A reason favoring using conditionalization to change our probabilities, then, is that it guarantees the coherence of our new probabilities.

There is also a Dutch Book argument supporting conditionalization, but it depends on our willingness to bridge our old and new probabilities with bets. To see how it runs, suppose you have revised the probability of D to 1, that of p to a, and that your old $P(p/D)$ was b. Now if $a \neq b$, and if the bookie knows this in advance, he can make a Dutch Book against you. Here is how. Before the truth of D is known, he places two bets that are not to take effect until D is verified: one for p—if $a < b$—or against p—if $a > b$; the other placed oppositely on "p given D." He sets the stakes at \$1. Then after D is verified, his payoffs on each bet will be determined according to whether p is true. Thus he will win one bet and lose the other, but because of his choices and your bad odds he will have a net gain. For example, if $a < b$, and p is true, he is paid $1 - a$ from the bet for p and $-(1 - b)$ from the bet against "p given D." His net is $b - a$, which is positive. He receives the same net if p is false, since the first bet pays $-a$ and the second b.

If D does not come true, the bookie neither gains nor loses since none of his bets are in effect. But he can assure himself of a gain even then by placing a side bet against D. Suppose that c is your betting quotient for D and d is 1/2 the absolute difference between a and b. Then the bookie sets the stake for the bet on D at $d/(1 - c)$. If D is false, he is paid $c[d/(1 - c)]$, which is positive. If D is true, he is paid $-(1 - c)[d/(1 - c)]$ (i.e., $-d$) from the bet on D and the absolute difference between a and b (i.e., $2d$) from his other two bets. Although $-d$ is negative, his net winnings are d and that is positive.

Let me summarize what has been established so far. If an agent's subjective probability assignments are rational in the sense of being coherent, then it follows, by the Dutch Book theorem, that they obey the laws of the probability calculus. Furthermore, by the converse Dutch Book theorem, obeying the laws of probability is sufficient for having coherent subjective probabilities. What is more, using the rule of conditionalization will allow an agent to form new coherent probability assignments as he learns new data and will protect him against future Dutch Books. Finally, since subjective probabilities are defined in terms of publicly observable betting behavior, they can be measured and objectively known.

PROBLEMS

1. Show that we obtain a law of the probability calculus when $P(p/D)$ is substituted for $P(p)$ in axiom 1a.
2. Do the same for axiom 2.
3. Do the same for axiom 3.

Have we made much of a case for the subjective approach? Is not coherence such a loose requirement that probability assignments based on it are useless for scientists and decision makers? If I am coherent in my betting at a racetrack, even if I bet on all the horses, the track will not be assured of a profit—no matter what. But I might lose my shirt, all the same. If prior to a match race between a Kentucky Derby prospect and my old nag, I assign a 90% chance of winning to the old nag and a 10% chance to the racehorse, then I am coherent. Yet unless I know something very unusual about the race (such as that the racehorse has been given a horse-sized dosage of sleeping pills), I would be foolish to take bets at those odds.

This example suggests that coherence is not a strong enough requirement for rationally assigning probabilities. There is a sense in which it would be irrational for me to set those odds on the old nag. (Of course, I might belie my true beliefs because I did not want to let the old fellow down by showing my lack of confidence in him. But I am assuming that such considerations do not intrude here.) The subjectivist has a response to this objection, however. The irrationality we sense in this example will be a genuine case of incoherence when we consider the other information I possess about horses and horseraces and my probability assignments to other statements. Since I know a fair bit about horses, I know that healthy young racehorses such as this one, participating in a normal match race against an old tired horse, are (virtually) certain to win. Given that, I should assign a probability of (nearly) 1 to the racehorse winning. My current probability assignment would not cohere with the probabilities I assign to other statements about horses and horseraces.

But what about those who know nothing about horses? What should they do at the track? Says the subjectivist: First of all, they should bet coherently. Then they should try to learn more about the horses and about horses in general. As they acquire new information, they should modify their probability assignments. If they use conditionalization, they can be assured of preserving coherence. Furthermore, the convergence theorems we mentioned in connection with Bayes's theorem may apply. If so, in the long run their odds are likely to be as good as those of the experts.

More boldly put, the convergence theorems show that in a wide class of cases, a person can start with completely heretical probability assignments, modify them by conditionalizing, and be assured in the long run of converging to the probabilities used by the experts. Subjectivists appeal to this to argue that subjective probabilities are suitable for scientists and decision makers. Convergence implies that those who derive their probabilities from observed frequencies will tend to agree in the long run with those who initially use hunches—provided, of course, that the latter use the data furnished by their frequentist colleagues. Consequently, subjectivists can provide empirically based probabilities whenever anyone can. That should satisfy scientists. But, it can be argued, subjectivists can do more: They can assign probabilities where frequencies, propensities, or logically based measures are unavailable or in principle unattainable.

Decision makers ask for nothing more than probabilities furnished by the "experts," and, in the long run, subjectivists can give them that. Moreover, and this can be crucial, they can supply probabilities in cases where frequentists, logicists, or propensity theorists must stand by with their hands in their pockets.

The capstone of the subjectivist defense consists in noting that subjective elements intrude in the ascertainment of so-called objective probabilities too. Under the classical view, we must choose the "equiprobable" possibilities to assign to a statement. This choice is based on a personal and subjective judgment that we have the right set of possibilities and that they are *equiprobable*, since the classical view fails to supply any methods for associating a set of equiprobable possibilities with a given statement. Similarly, when we decide, after seeing a coin land heads on 511 of 1,000 tosses, that the long-run relative frequency of heads is 1/2, we are making a personal assessment of the *probable* future behavior of the coin. Also we must decide whether a theory is sufficiently *probable* given the evidence for it before we can accept the claims it makes about propensities. In short, the application of objective probabilities is dependent on subjective ones. Subjective probabilities are thus both fundamental and unavoidable.

This is a persuasive argument—one that should make us seriously consider the merits of the subjective approach—but it is not airtight. Subjectivists are right, I think, in claiming that in applying the objective theories of probability, we will often be faced with decisions that are not clear-cut. That in turn could lead different people to make different probability assignments. But it does not follow from this that it is never possible to objectively ground probability assignments, and that is what subjectivists must show for their point to stick.

Furthermore, the subjective approach faces some problems of its own that are not unlike those faced by the previous views. First, subjectivists' assurances about long-run convergence may fail to work out in practice. If the initial probabilities are widely divergent and the data either sparse or inconclusive, convergence may not come until long after a decision based on the probabilities in question is due.

Second, there is a plethora of statements for which we can set betting quotients at values that do not comply with the probability calculus without having to worry about being trapped in a Dutch Book. A bet on whether the heat death of the universe will ever occur is a case in point. If it does occur, no one will be around to be paid; if it is not going to occur, we might not know that with enough certainty to settle the bet. Another example would be a bet on whether there is life outside the observable portion of the universe. "Can't we settle those bets by appealing to science?" subjectivists might respond. Here we cannot. Science has no definite answers to these questions; it tells us only that each of the outcomes is possible.

A related problem is that the betting approach presumes that once we make a bet it will remain clear what winning means until the bet is settled. But that it not always so. Frequently, we find ourselves in situations where we initially wonder whether something is true and then find that our very question does not really make sense or fails to have a definite answer. Suppose, for example, that you bet that the fellow in the corner drinking gin knows how to play bridge and I bet that he does not. Next suppose we learn that no one in that corner is drinking gin, although one of the men is drinking vodka—and he does play bridge. Is our bet still on? You say it is and claim that it was clear that we meant *that fellow*—the one over there. I say that it is not and that I meant *the gin drinker*, since I would never bet on a vodka drinker's being able to play bridge. I do not think that you could hold me to the bet. If we turn to the history of science and mathematics, we can even find cases where our conception of what counts as an "answer" to a question changed in the course of finding one.

This and the previous point show that we can no more count on the subjective approach to be universally applicable than we could count on the previous approaches we considered. In addition, subjectivists face a difficult dilemma concerning the revision of probabilities. As we have seen, coherence is the sole condition that subjectivists impose on anyone's initial probability assignments. Now suppose I, who know nothing of abstract art, have become friendly with some art critics and start to visit museums of modern art with them. I know what I like, but I try to guess at what the critics like. At first, I use my best hunches to assign probabilities to their liking various paintings and other art objects, being sure, of course, to be coherent. Later, when I learn what the critics have to say, I see that my probabilities must be modified. How should I do this?

An obvious proposal is to use the rule of conditionalization. But, let us suppose, I find that that rule yields probabilities that are still far from those I am now inclined to hold. (One way this could occur is if the conditional probabilities I use are, themselves, "way off"; another is through my initial assignments being wildly at variance with the critics' judgments.) What should I do? Stick with the rule and hope that convergence will soon come? Or give up my initial priors and start with a new set? Some subjectivists would advise me to stick by the rule no matter what. That opens them to the charge of stubborn tenacity. Others would allow me to pick a new set of priors whenever I wish with coherence being the only requirement I must meet. And that seems flippant. The challenge for the subjectivist is to propose a middle course, that is, a rational method for the modification of probabilities that is wedded to neither unwarranted tenacity nor flippancy.

Solutions to this problem have been proposed, but I will not delve into them here. I will only remark that they have been tied to the abandonment of a unique prior probability assignment in favor of sets of assignments or intervals between upper and lower probabilities. This amounts to rejecting the original subjectivist view—at least in its full particulars.

Our discussion of decisions under ignorance left several important issues unsettled; that was to be expected since we had reached the philosophical frontiers of the subject. The same has now happened with our discussion of probability, and I am forced to conclude it with no decision as to which conception is best. However, at least you should have a clear idea of the difficulties a successful view must overcome and of the advantages and drawbacks to the major views that have been proposed so far.

3-4. References

Skyrms 1975 contains a good introduction to probability theory and the interpretations of probability. *Chernoff and Moses* expound somewhat more advanced material. *Raiffa* presents an excellent discussion of the use of Bayes's theorem in decision theory as well as an introduction to subjective probability. *Carnap* is a classic on the logical and philosophical foundations of probability. The papers by DeFinetti and Ramsey are reprinted in *Kyburg and Smokler*. Also see

Savage and *Eells* for defenses of the subjective approach. *Kyburg* presents an account of probability that fails to obey the entire probability calculus; it also contains a number of critical papers on the various interpretations of probability as well as further references. See *Levi* for another novel approach and further criticism of standard Bayesianism.