 edited by George von Furstenberg, Kluwer ..., 1990 42

Keynes, John Maynard (1921), *A Treatise on Probability*, London: Macmillan.
Koopman, Bernard O. (1940), "The Axioms and Algebra of Intuitive Probability," *Annals of Mathematics*, 41, pp. 269-92.
Kyburg, Henry E., Jr. (1987), "Bayesian and Non-Bayesian Evidential Updating," *Artificial Intelligence*, 31, pp. 271-93.
Kyburg, Henry E., Jr. (1970), *Probability and Inductive Logic*, New York: Macmillan.
Levi, Isaac (1980), *The Enterprise of Knowledge*, Cambridge: MIT Press.
Mises, Richard von (1928), *Probability, Statistics and Truth*, London: George Allen and Unwin.
Niiniluoto, Ilkka (1977), "On A K-Dimensional System of Inductive Logic," in P. Asquith, ed., *PSA 1976*, East Lansing: Philosophy of Science Association, pp. 425-47.
Popper, Karl R. (1959), *The Logic of Scientific Discovery*, London, Hutchinson and Co.
Ramsey, Frank P. ([1931] 1950), "Probability and Partial Belief," in R. B. Braithwaite, ed., *The Foundations of Mathematics and Other Logical Essays by Frank P. Ramsey*, London: Routledge and Kegan Paul, pp. 256-57.
Rescher, Nicholas (1958), "Theory of Evidence," *Philosophy of Science*, 25, pp. 83-94.
Savage, Leonard J. (1954), *The Foundations of Statistics*, New York: John Wiley and Sons.
Shafer, Glenn (1976), *A Mathematical Theory of Evidence*, Princeton: University of Princeton Press.
Shortliffe, E. H. (1976), *Computer-Based Medical Consultations: MYCIN*, New York: Elsevier.
Teller, Paul (1976), "Conditionalization, Observation, and Change of Preference," in W. Harper and C. Hooker, eds., *Foundations of Probability Theory*, Vol. I, Dordrecht, Netherlands: Reidel, pp. 205-53.
Van Fraassen, Bas (1984), "Belief and Will," *Journal of Philosophy*, 81, pp. 235-36.
Venn, John (1866), *The Logic of Chance*, London: Macmillan; reprinted New York: Chelsea, 1963.
Zadeh, Lotfi (1979), "On the Validity of Dempster's Rule of Combination of Evidence," Berkeley, Memo UCB/ERL M79,24.

CHAPTER 5

THE UNITY OF PROBABILITY

Glenn Shafer[1]

Abstract--This article describes the ideal picture of probability, an imaginary situation in which knowledge, frequency, value, and belief are inextricably intertwined.

The article is meant as a foundation for an examination of the longstanding controversy between frequentists, who see knowledge of the long run as the true and complete foundation of probability, and subjectivists, who see rational belief as its true and complete foundation. Once we understand the interdependence of frequency and belief in the ideal picture, we can understand the shortcomings of both frequentism and subjectivism.

The mathematical theory of probability involves a circle of reasoning. There are three ideas in this circle: knowledge of the long run, value, and warranted belief. Any of the three can be taken as a starting point. From knowledge of the long run, say, we can derive an idea of fair price or value for gambles. From this, we can derive properties of warranted belief, and from these properties, we can deduce the knowledge of the long run with which we began.

Yet none of the three ideas is adequate by itself as a philosophical basis for the theory. In order to justify axioms for the fair prices of gambles, we must assume the existence of long-run frequencies and belief in these frequencies. In order to justify axioms for warranted belief, we must make arguments about knowledge and gambling. And the knowledge of the long run found in the picture includes knowledge about beliefs and odds.

---

[1]Research for this article, © 1989 by Glenn Shafer, has been partially supported by the National Science Foundation, through grant IRI-8610293 to the University of Kansas and grant BNS-8700864 to the Center for Advanced Study in the Behavioral Sciences.

The mathematical theory of probability is properly a theory about an imaginary situation in which knowledge, frequency, value, and belief are inextricably intertwined. We can call this imaginary situation the ideal picture of probability. This article examines the ideal picture verbally but in detail, with an emphasis on exactly how its different aspects are intertwined.[2]

Mathematical probability was invented in the seventeenth century as a theory of games of pure chance, and it is still here that we can most easily imagine the ideal picture. We can imagine knowing the long-run frequencies of outcomes of flips of coins or throws of dice, and we can imagine basing beliefs about outcomes and prices for gambles on these frequencies. Because our purpose in this article is to describe the ideal picture clearly, rather than to teach about the breadth of application of probability, we limit ourselves to talking about games of chance. Most of our discussion will be based on the very simplest case--the case where a fair coin is flipped repeatedly.

Today, mathematical probability has a role that reaches far beyond the realm of games of chance. It has been applied successfully to an astonishing variety of problems. But these applications are best understood as involving analogies between the domains of application and the ideal picture. The analogies vary in aptness as well as in structure, and the aptness of the analogy is one aspect of the quality of the application.

The pragmatic attitude of the preceding paragraph may seem trite. Few contemporary readers will be offended by the distance it allows between theory and application. Yet philosophical discussions of probability are still dominated by ideas that were inspired by the far less pragmatic empiricism of the nineteenth century. This empiricism demanded that the terms of a theory correspond closely to the reality that the theory models.

Nineteenth-century empiricism tore probability apart. Those who wanted to apply probability broadly in late nineteenth and early twentieth centuries could seldom find the ideal picture in reality, but they could often find aspects of it. They were driven, therefore, to seize

[2]This article is adapted from a forthcoming book by the author, *The Unity and Diversity of Probability* (Shafer, 1990). Many of the historical and mathematical facts that are mentioned here are substantiated more fully in this book.

on one or another aspect and to proclaim it the true and complete foundation for their theory. Some claimed that probability could be founded on frequency alone and hence could be applied to any situation where frequencies are observed. Others claimed that it could be founded on belief alone and hence could be applied to any situation where degrees of belief are appropriate. Thus was born the controversy between the frequency and subjective interpretations of probability. This controversy is now 150 years old.

This article is meant as a foundation for a mathematical, philosophical, and historical examination of the frequentist vs subjectivist controversy. Only when we have a clear view of the interdependence of frequency and belief in the ideal picture can we begin to dissolve this hoary controversy.

In Section I, we take a purely descriptive look at the simplest case of the ideal picture, the case where a fair coin is flipped repeatedly. The most important features of the picture are already present in this simplest case.

In Section II, we explain how our description of the ideal picture in the case of the fair coin can be refined into a circle of reasoning. There are three points in this circle: knowledge of the long run, value, and warranted belief. Any of these three ideas can be taken as the starting point of the mathematical theory. Yet none of the three is adequate by itself as a philosophical basis for the theory.

In Section III, we examine the different ways our sequence of coin flips might turn out in terms of a tree that branches as the outcomes of successive flips are considered. By studying this tree, we deepen our understanding of gambling within the ideal picture, and we set the stage for generalizing the picture.

In Section IV, we spell out the generalization. In the special case of the fair coin, the same experiment is repeated over and over, and this experiment has only two possible outcomes, heads and tails. In the general ideal picture, we still perform a long sequence of experiments, but the experiments may not all be the same, some or all of them may have more than two possible outcomes, and which experiment is performed at a given step may depend on the outcomes of the preceding experiments.

1. FLIPPING A FAIR COIN

In the foreground of the simplest version of the ideal picture of probability is a person who flips a coin many times, always hoping for heads. Since he is trying to get heads, the successive flips are called trials. In the background of the picture are spectators who watch the trials and gamble on their outcomes.

Knowledge of the Long Run

The spectators have a peculiarly circumscribed chunk of knowledge. They know that the coin will land heads about half the time in the whole sequence of trials, but they know nothing further that can help them predict the outcome of any single trial or group of trials. They cannot identify beforehand a group of trials in which the coin will land heads more than half the time, and the outcomes of earlier trials are of no help to them in predicting the outcomes of later trials.

The spectators gamble on the outcomes. Just before each trial, each spectator has an opportunity to make a small even-money bet on heads or on tails. But since they are unable to predict the outcomes of the trials, the spectators are unable to take advantage of these opportunities with any confidence. Each spectator knows that he or she will lose approximately half the time.

A spectator who makes many small bets may end up with a net gain even though he or she does lose approximately half the time. A spectator who bets a dollar on each of a million trials, for example, might conceivably win 501,000 times and lose 499,000 times, for a net gain of $2,000. But a comparable net loss is also possible. The spectator has no way of knowing whether he or she will gain or lose.

It does not matter whether the spectator takes the outcomes of preceding trials into account in deciding whether, how, and how much to bet on each flip. No plan or strategy based on earlier outcomes can assure a net gain.

Even if the spectator does come out on the plus side after a long sequence of bets, he or she may fall behind at some earlier point. So unless the spectator has a substantial stake, the attempt to make a long sequence of even-money bets of fixed size may be interrupted by

bankruptcy. The spectator can always avoid bankruptcy by making the even-money bets smaller when reserves get small, but this will make it even harder to recover lost ground. Smaller bets can yield only smaller gains.

Because of the danger of bankruptcy, a spectator can hope only for gains comparable in size to his or her initial stake. The spectator knows that no strategy will enable him or her to parlay a small stake into a large fortune.

Fair Bets and Fair Prices

Knowing that none of their fellows can take advantage of even-money bets on the individual trials, the spectators consider such bets fair. They consider it fair for Peter and Paul each to put 50¢, say, on the table, with the understanding that Peter will get the whole $1 if the coin lands heads on a given trial, and Paul will get it if the coin lands tails. In other words, they consider 50¢ the fair price for a ticket that will pay $1 if the coin lands heads and nothing if it lands tails. This is the fair price on the given trial when the whole sequence of trials begins, and it remains the fair price until that trial is performed. The fair price on a particular trial does not change as the spectators see the outcomes of earlier trials.

The spectators also bet on more complicated events, events that involve more than one trial. They may bet, for example, on the event that the coin comes up heads on both of the first two trials, on the event that it comes up heads on at least three of the first ten trials, or on the event that it comes up heads on approximately half of the first thousand trials. They agree on fair odds for all these events.

Consider, for example, the event that the coin lands tails at least once on the first two trials. The spectators agree that 3 to 1 odds are fair for this event. They consider it fair for Peter to put 3 times as much money on the table as Paul, with the understanding that Peter will get it all if at least one of the two trials produces tails, and Paul will get it all if both produce heads. (In general, when the odds are k to 1 for an event, the person betting on the event must put up k times as much as the person betting against it. Similarly, when the odds are k to 1 against, the person betting against must put up k times as much as the person betting for.)

We can always translate talk about fair odds into talk about fair prices. In the case of the 3 to 1 odds for the coin landing tails at least once, the spectators consider it fair for Peter to put 75¢ on the event while Paul puts 25¢ against it. If the coin does land tails at least once, then Peter gets to keep his 75¢, and he also gets Paul's 25¢, for a total of $1. So, in effect, Peter has paid 75¢ for a ticket that pays $1 if the event happens. Paul, on the other hand, is betting that the coin will land heads both times. If this happens, he gets to keep his 25¢, and he also gets Peter's 75¢. So Paul has paid 25¢ for a ticket that pays $1 if the coin lands heads both times. The spectators consider these ticket prices fair.

At the end of the next section (Section II), we will explain more thoroughly the arithmetic involved in translating between odds and prices.

A Short-Run Aspect of Fairness

As we have said, the spectators consider the odds and prices they have agreed on fair because no one can take advantage of these odds and prices. A spectator who makes small bets on many different trials will break even, a spectator cannot hope to parlay a small stake into a large fortune, and a spectator cannot even be confident of a modest gain. The first of these three statements is strictly a statement about the long run, but the latter two statements apply to the short run as well.

One aspect of the short-run fairness of ticket prices is that they are related to each other in such a way that a spectator cannot make money simply by buying and selling, without regard to how the trials come out.

Here is an example of how Peter might try and fail to make money simply by buying and selling. He deals in four $1 tickets:

#1: a $1 ticket on heads on the first trial;
#2: a $1 ticket on heads on the second trial;
#3: a $1 ticket on heads on both the first and second trials;
#4: a $1 ticket on heads on at least one of the first two trials.

First Peter buys tickets #1 and #2. He realizes that he will get $2 from these tickets if both trials produce heads, $1 if only one produces heads, and $0 if neither produces heads. Counting on these proceeds to cover his obligations, he now sells ticket #3 (this obligates him to pay $1 if both trials produce heads), and ticket #4 (this obligates him to pay $1 unless both trials produce tails).

As Table 1 shows, the proceeds Peter will get from tickets #1 and #2 will exactly cover the obligations he acquires by selling tickets #3 and #4, no matter how the two trials come out. Peter has simply repackaged and resold what he bought. He bought #1 and #2, and he resold them as #3 and #4. Did he make any money by this maneuver? No. Since the tickets are priced at 50¢, 50¢, 25¢, and 75¢, respectively, he paid 50¢ + 50¢ = $1 for #1 and #2, and he received 25¢ + 75¢ = $1 when he sold #3 and #4. He broke even.

Table 1

Peter's Payoffs

		#1	#2	#3	#4	Net
Outcome	HH	+$1	+$1	-$1	-$1	0
	HT	+$1	0	0	-$1	0
	TH	0	+$1	0	-$1	0
	TT	0	0	0	0	0

Changes in Price

We have said that the fair price for a ticket on heads or tails on any particular trial remains unchanged until that trial is performed. How does it change when the trial is performed?

The change is very simple. Consider a ticket that pays $1 if the coin lands heads on the fourth trial. Until the fourth trial is performed, the price of this ticket is 50¢. When the fourth trial is performed, the price changes, and the change depends on the outcome of the trial. If the coin lands tails, the ticket will be worthless; the new price will be $0. If the coin lands heads, the ticket will be worth a dollar; the new price will be $1. The new price will be permanent; it will not change again.

Prices for tickets that involve several trials can change more often. Consider, for example, the event that the first two trials both produce heads. Initially, the price for a $1 ticket on this event is 25¢. After the first trial, it is either zero or 50¢. If it is zero, it remains zero. If it is 50¢, then it changes again after the next trial, to either zero or $1.

The fact that prices can change repeatedly gives Peter another opening to try to make money by buying and selling tickets. He can buy cheap tickets now, sell the ones whose prices go up, and use the proceeds to buy better tickets. Here, as before, however, the prices are such that he cannot make money in this way.

Here is an example, so simple as to be silly, yet adequate to make the point. Consider again the event that both of the first two trials produce heads. Peter sells Paul a $1 ticket on this event, and he uses the 25¢ from the sale to buy an identical ticket from Mary. This might seem to accomplish nothing, but Peter has a crafty plan for the ticket he has bought from Mary. If the first trial produces heads, then this ticket will be worth 50¢, and he plans to sell it and buy a better ticket. He will buy a $1 ticket on heads on the second trial.

The crafty plan is to no avail, of course. His new $1 ticket on heads on the second trial will pay $1 if the second trial produces heads, but he will have to use this $1 to pay Paul. Just as before, his whole maneuver will break even, both in proceeds from the buying and selling and in payoffs from the tickets, no matter what happens.

The example is silly because it is so obvious that nothing can be accomplished, after the first trial produces heads, by selling the original ticket on heads on both trials and buying a new ticket on heads on the second trial. Since the first trial has already produced heads, the original ticket is now, for all practical purposes, already a ticket on heads on the second trial.

Back to the Long Run

As it happens, the spectators give very great odds that the coin will land heads on approximately half of any large number of trials.

They will give 600 to 1 odds, for example, that the number of heads in the first thousand trials will be between 450 and 550.

The spectators also give very great odds against the success of any betting strategy that attempts to increase initial capital by more than a few orders of magnitude. They give at least 1,000 to 1 odds, for example, against any attempt to parlay $20 into $20,000.

This brings us back to our earlier comments about the spectators' knowledge of the long run. We said that they know that about half the trials will be heads, and we said that they know that no strategy can enable them to parlay a small stake into a large fortune. Knowledge, it appears, goes together in the ideal picture with giving very great odds.

Once we have said that knowledge goes with very great odds, it is natural to say also that guarded belief goes with less great but substantial odds. It is natural to say that the spectators' degree of belief in an event is measured by the odds they give. Knowledge is simply the limiting case of very great belief, or practical certainty.

The odds that the spectators give are warranted by their knowledge of the long run. So the belief measured by these odds is not mere whim or fantasy. It is warranted belief.

Once we have interpreted odds as a measure of warranted belief, we can quantify how warranted belief becomes knowledge or practical certainty as the risky shot is stretched into the long shot, or as the short run is stretched into the long run.

Indeed, the spectators' knowledge that long shots, or very ambitious gambling strategies, will fail is simply the limiting case of their skepticism about all gambling strategies, more ambitious and less ambitious. They give two to one odds against doubling initial capital, thousand to one odds against increasing initial capital a thousandfold, and million to one odds against increasing initial capital a millionfold. The spectators will give odds of at least k to 1 against any strategy for multiplying initial capital by k (Feller, 1968).

Similarly, the spectators' knowledge that the percentage of heads will be very near 50% in the long run is the limiting case of a belief that it will be relatively close to 50% in the shorter run. The belief that it will be close to 50% and the degree of closeness expected both increase steadily with the number of trials. This point is illustrated by Tables 2 and 3.[3]

Table 2 shows how the odds for the percentage of heads falling between 45% and 55% increases with the number of trials. The first column gives the number of trials in the sequence, the second column gives the number of heads needed in order for the percentage to fall between 45% and 55%, and the third column gives the odds, to one significant figure.

Table 2

Odds for 45% to 55% Heads

Number of Trials	Number of Heads	Approximate Odds
100	45 to 55	2 to 1
1,000	450 to 550	600 to 1
10,000	4,500 to 5,500	$6x10^{22}$ to 1

The odds in the last row of Table 2, $6x10^{22}$ to 1, are far too great to have any place in real gambling. Numbers this great do serve, however, to dispel misgivings we might have about identifying knowledge and practical certainty with very great odds. These odds are surely great enough to match the certainty we have in any of our everyday knowledge.

Table 3 shows how the spectators' range of uncertainty for the percentage of heads tightens around 50% as the number of trials increases. Each row of this table describes an event for which the spectators give about 20 to 1 odds. In the case of 10,000 trials, for example, the spectators give about 20 to 1 odds that the number of heads will be between 4,900 and 5,100, or between 49% and 51%.

[3]These tables are based on the normal approximation to the binomial distribution. Tables of the normal distribution can be found in most probability and statistics textbooks. The extreme odds shown in Table 2 cannot be found in standard tables, but they can be found using an approximation given by Feller (1968, p. 175).

Table 3

Ranges for Approximate 20 to 1 Odds

Number of Trials	Number of Heads	Percentage of Heads
100	41 to 59	41% to 59%
1,000	470 to 530	47% to 53%
10,000	4,900 to 5,100	49% to 51%
100,000	49,700 to 50,300	49.7% to 50.3%
1,000,000	499,000 to 501,000	49.9% to 50.1%

We mentioned earlier that a spectator might win as much as $2,000 by betting a dollar on each of a million trials. The last row of Table 3 tells that so large a win, though possible, is unlikely. The odds are 20 to 1 that the million bets will result in a gain or loss of less than $2,000.

II. A CIRCLE OF IDEAS

Our description of the ideal picture traced a circle. We started with knowledge of the long run. Then we talked about fair odds. Then, by relating odds to belief and very high odds to knowledge, we came back to knowledge of the long run.

This circle of description can be cast as a circle of reasoning. The spectators reason from their knowledge of the long run to the assignment of fair odds to events. Then they reason that these odds measure warranted belief. Then they deduce from these odds, or warranted beliefs, the knowledge of the long run with which they began.

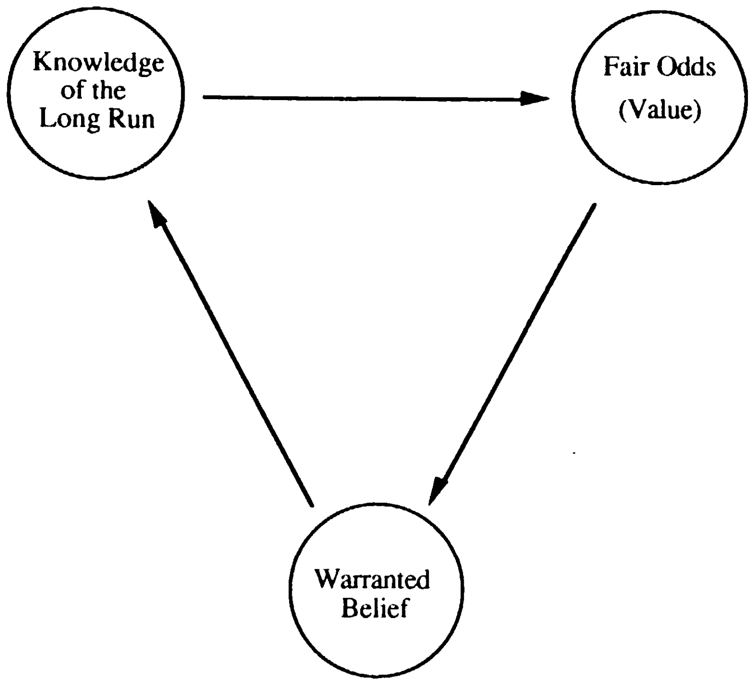

Figure 1. The Circle of Probability Ideas

This circle is depicted in Figure 1. The spectators' reasoning as they move around the circle is not airtight, but it is persuasive. It is the raw material for a rich mathematical theory.

Five Steps

We can break the circle of reasoning into five steps. Three of these steps are represented by the three arrows in Figure 1; the other two can be located within the circle labelled "Fair Odds."

The spectators take the first step, represented by the arrow from "Knowledge of the Long Run" to "Fair Odds," when they reason from their knowledge of the long run to an assignment of odds to the individual trials. They reason that this assignment of odds is sensible and fair, because it takes all their relevant knowledge into account, and because someone who makes many bets at these odds will approximately break even. We can quibble with their reasoning. We can point out that if many bets are made, a person can have substantial losses even while approximately breaking even. And we are certainly entitled to decline to risk our own capital in this way. But after quibbling, we will concede that the spectators' knowledge of the long run provides a good justification for calling even odds on each trial fair, as good a justification for calling odds fair as we could ever find.

The second and third steps take place within the circle labelled "Fair Odds."

The second step is from fair odds on individual trials to fair odds for all events. We have not yet spelled out the reasoning in support of this step, but the germ of the argument is contained in the examples where Peter fails to make money by buying and selling tickets on events. Once we agree on even odds on individual trials, and once we agree that these odds are not affected by the results of earlier trials, there is exactly one way of assigning odds to more complicated events so that Peter cannot make money by buying and selling tickets.[4] The fact that there is a way at all strengthens the case for calling the even odds on individual trials fair. The fact that there is only one way provides an argument for calling the resulting odds on more complicated events fair.

The third step is to deduce that the odds we have assigned to complicated events include very great odds that the coin will land heads approximately half the time and that schemes for parlaying small sums into large ones will not succeed.

The spectators take the fourth step, represented by the arrow from "Fair Odds" to "Warranted Belief," when they interpret fair odds as measures of belief. They point out their own willingness to bet at the odds they call fair. Appealing to the natural tie between action and belief, they conclude that these odds reflect or measure their beliefs.

[4]A proof of this assertion can be extracted from the more general reasoning found in de Finetti (1937).

Finally, the spectators take the fifth step, represented by the arrow from "Warranted Belief" to "Knowledge of the Long Run," when they interpret very great belief as practical certainty or knowledge. This step offers great scope for philosophical argument; we can debate at length the nature of knowledge. But the step is reasonable if interpreted modestly. It is reasonable to use the word know to express very great belief or willingness to take very great risks.

Making the Circle into Mathematics

The reasoning we have just described does not qualify as formal mathematics. Much of it is rhetorical, not deductive. And it goes in a circle. These are characteristics of informal mathematical reasoning. They are eliminated when mathematics is formalized.

Informal mathematics always involves both rhetorical and deductive reasoning. When it is formalized, the rhetorical reasoning becomes the justification for definitions, and the deductive reasoning becomes the justification for theorems.

Though it is often a sign of confusion in debate, circular reasoning is a sign of vigor in informal mathematics. It is a sign that our ideas are connected in multiple and subtle ways. It is also a sign of freedom of choice in formalization. Formalization requires taking one idea as an axiomatic starting point. When ideas lie in a circle of reasoning, all of them are candidates for this starting point. The circle can be broken at any point.

In Figure 1, the arrows represent the major rhetorical steps, and hence the major potential definitions. We can define value for gambles in a way that accords with long-run knowledge, we can define warranted belief in terms of value, and we can define knowledge as very great warranted belief. The circles joined by the arrows represent the potential starting points. The mathematical theory of probability can be based on axioms for knowledge of the long run, axioms for the fair prices of tickets on events, or axioms for warranted belief.

In our informal exposition, we began with knowledge of the long run. Formal axiomatizations generally begin with axioms for the fair prices of tickets on events or axioms for warranted belief.

Probability

We have delayed using the word probability within our description of the ideal picture of probability, for fear that a clear view of the roles of knowledge, frequency, belief, and value might be obscured by confusion about the pertinence of this word to these different roles. As Porter reminds us in an earlier chapter in this volume, the word probability was first associated with warranted degree of belief, yet it has become a synonym for frequency in the minds of many.

Yet the word probability is too well established and useful to ignore indefinitely. We will use it. We will use it primarily to mean warranted belief, but we will also use it to mean frequency. We will treat it as a symbol of the unification of these ideas within the ideal picture.

We say that the probability of the coin landing heads on a given trial is 0.5. This number is simultaneously a frequency (the frequency of heads in the whole sequence of trials), a degree of belief (another way of expressing the belief that the spectator expresses by talking about even odds), and a price (the price of a $1 ticket on the event).

More generally, we refer to the price in dollars of a $1 ticket on an event as the probability of the event. This is always a number between zero and one. It is a price, but it is also a measure of belief. It measures the same belief that is measured by the odds for the event, just as a Celsius thermometer measures the same temperature that is measured by a Fahrenheit thermometer.

It is only in the case of the elementary events, "the coin lands heads" and "the coin lands tails," that the word "probability" directly unifies frequency and belief. These elementary events have an opportunity to happen on each trial, and their probability is the frequency with which they happen. Other events are not repeated as the coin is flipped, and hence their probabilities cannot be interpreted directly as frequencies. These probabilities are merely warranted beliefs, not frequencies. But in saying this, we are not giving belief primacy over frequency in our understanding of probability. In the ideal picture, the belief measured by probability is always warranted belief, and the warrant comes from our knowledge of the long run, including our knowledge of the frequencies of the elementary events.

The Interdependence of Frequency, Value, and Belief

The fact that knowledge of the long run, value, or warranted belief can each be used as a starting point for the mathematical theory of probability should not be taken to mean that any one of these ideas is sufficient for grounding the theory. As it turns out, the axioms that we need in order to begin in any one of these domains can be understood and justified only by reference to the other aspects of the picture. The three aspects of the ideal picture are inextricably intertwined.

We have just noted one of the problems with an exclusive emphasis on frequency. Since only some of the events in the ideal picture can be considered repetitive, not all the probabilities in the picture can be interpreted directly as frequencies.

A deeper problem is that the knowledge of the long run that provides warrant for belief involves more than frequency. It involves the relationship between frequency and knowledge. In addition to knowing frequencies, the spectators know that these frequencies, and the odds and fair prices determined by them, leave no openings for chicanery. They know that no strategy can assure modest gains or give a realistic chance of substantial gain. It might be thought that this negative knowledge could be deduced from the fact that one knows frequencies and nothing more, but this seems not to be true. Attempts to make frequency a formal foundation of probability have foundered on the problem that the specification of long-run frequencies cannot by itself rule out strategies that assure modest gains.[5]

When we broaden our foundation from frequency alone to include other aspects of our knowledge of the long run, we find ourselves talking about odds or prices. We have brought value into the picture. But value alone is also an inadequate foundation.

[5]Martin-Löf (1969) gives an excellent review of these attempts. The best-known attempt to found probability on frequency was made by von Mises. His ideas, which were first published in 1919, are set out most comprehensively in von Mises and Geiringer (1964). Wald (1937) demonstrated that von Mises's axioms for frequency were noncontradictory, but Ville (1939) showed that they failed to rule out schemes for modest gains.

It is sometimes argued that value is adequate. It is asserted that rationality requires a person to set prices at which he or she will buy and sell tickets on events.[6] If a person does set such prices, the person will want the prices to be related to one another in such a way that Peter cannot make money from him or her merely by buying and selling. This is enough, as it turns out, to lead to the general version of the ideal picture that we will look at in Section IV.

Yet the first step in this argument is a non sequitur. Rationality does not require a person to set a whole schedule of odds and allow others to choose the bets. Such behavior would be irrational in most circumstances. It is precisely the spectators' knowledge of the long run that makes it rational, or at least acceptable, in the ideal picture.

Trying to ground probability on warranted belief alone is even more hopeless. The axioms for warranted belief that arise in the mathematical theory of probability are simple and mathematically attractive, but they can hardly be explained or justified by appealing to the concept of warranted belief alone. Their justification requires appeals to frequency and value.

Translating Between Odds and Probability

Just as we can translate between Fahrenheit and Celsius thermometer readings, we can translate between odds and probability. Given odds of r to s in favor of an event, the probability is $r/(r+s)$. Given a probability p for an event, the odds in favor of the event are in the ratio of p to 1-p.

We have already seen examples of the translation from odds to probability. If the odds are even, Peter and Paul both put up 50¢, and the winner gets the whole dollar. Thus Peter pays $0.50 for a $1 ticket. So even odds means a probability of 0.50. If Peter gives Paul 3 to 1 odds, then when Peter puts up 75¢, Paul only needs to put up 25¢. Thus Peter pays $0.75 for a $1 ticket. So 3 to 1 odds means a probability of 0.75.

[6]This claim was advanced forcefully by de Finetti (1937) and Savage (1954), and it has become widely accepted among subjectivists in recent decades.

In general, if Peter gives Paul the odds r to s, then Peter's \$r will be matched by Paul's \$s, and the winner will get \$(r + s). If they want to set the total stakes at \$1, then Peter will put up \$r/(r + s), and Paul will put up \$s/(r + s). Thus putting odds r to s on an event means pricing a \$1 ticket on the event at \$r/(r + s). This is how we get 75¢ in the case of 3 to 1 odds: \$3/(3 + 1) = 75¢.

It is even easier to translate probability into odds. If the probability is 0.75, then we pay 75¢ for a \$1 ticket. The person we are betting against must be putting up the other 25¢. So the odds are 75 to 25, or 3 to 1. In general, when the probability is p, we are paying \$p for a \$1 ticket, and hence the other person is putting up the other \$(1-p). The odds are thus p to 1-p. The numbers p and 1-p are usually fractions (0.75 and 0.25 in the example), but it is legitimate to multiply them to get whole numbers, or to divide them to simplify the whole numbers, because the meaning of r to s odds is simply that the stakes put up should be in the proportion r to s. The proportion is unchanged when r and s are multiplied or divided by the same number.

The extremes require special comment. When the probability of an event is one, we usually say that the odds in favor of the event are infinite. When the probability is zero we say that the odds against are infinite.

Saying the odds in favor of an event are infinite is a way of saying that no one will bet against the event; in order to get Paul to put up \$1 against the event, Peter would have to put up an infinite amount for it. This talk about infinity might not seem to agree with the general rule that probability p means odds p to 1-p. But actually it does. The numbers 1 and 0 are in the same proportion as the numbers ∞ and 1.

III. THE TREE OF SITUATIONS

A tree of situations is a graphical representation of the possible ways a sequence of experiments might turn out. Though it is not strictly necessary for understanding the ideal picture in the special case of a fair coin, the tree of situations clarifies the picture and facilitates its generalization.

In this section, we use the tree of situations to describe events, expectations (tickets formed by compounding tickets on different events) and their fair prices, and strategies.

Situations

Figure 2 shows the eight ways three flips of a fair coin can come out. Each of the eight ways is represented by a path down the figure, from the circle at the top to one of the eight stop signs at the bottom.

We call Figure 2 a tree of situations. Each circle and stop sign is a situation that can arise in the course of the three flips. The circle at the top is the initial situation, the situation at the beginning. The stops signs are the possible situations at the end. The circles in between are possible situations in which only one or two of the flips have been completed. Inside each situation are directions for what to do in that situation.

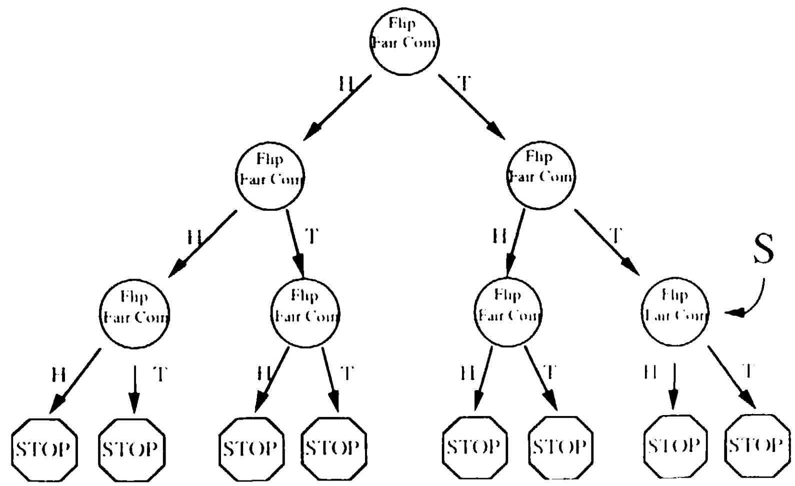

Figure 2. Three Flips of a Fair Coin

A situation can be identified by saying what has happened so far. The situation labelled S in Figure 2, for example, is the situation after the first two flips have come up tails.

The fair coin is flipped many times in the ideal picture, thousands and thousands of times. So the tree of situations for the ideal picture is like Figure 2 but much deeper. It goes down the page thousands of steps, branching at each step.

This tree exists only in our imagination, of course. We could not possibly draw it on paper. The tree for only a hundred flips would already contain 2^{100} stop signs. This is more than 10^{30}, or one nonillion, a number immense compared even to the age of the universe in seconds.

For the sake of simplicity, we will assume that all the stop signs in our tree of situations lie at the same depth. This means that all paths down the tree--all paths from the initial situation to a stop sign-- have the same length. In other words, every path down the tree involves the same number of flips. This number is fixed, as it were, at the outset. The person flipping the coin does not use the outcomes of the flips to decide when to stop.

The assumption that all stop signs lie at the same depth will allow us to give a simple recipe for computing (in principle!) probabilities and values of expectations in the tree. The assumption is of no conceptual importance, however. We will drop it in the next section, when we study the general ideal picture of probability. In the general picture, flipping a fair coin is not the only experiment that can be performed in a given situation, and what experiment to perform, or whether to stop, can depend on the situation.

Events

We already know what an event is. An event is something that happens or fails in the course of the flips. Getting heads on the first flip is an event. Getting exactly two heads in the course of the first three flips is an event. And so on.

In terms of the tree of situations, we can identify an event by listing a set of stop signs. By the time we get to a stop sign, the event has happened or failed. So we can identify the event by identifying the stop signs in which it has happened.

To illustrate this, we letter the stop signs in Figure 2. This produces Figure 3. The event that we get heads on the first flip corresponds, here, to the set {a,b,c,d} of stop signs. The event that we get exactly two heads corresponds to the set {b,c,e}. And so on.

When we go beyond a few flips to the long sequence of flips involved in our ideal picture, it becomes impractical to identify events by listing stop signs. There is no practical way, for example, that we could list the ways of getting between 45% and 55% heads in a thousand flips. The conceptual identification of events with sets is extremely useful, however, because sets are central to modern mathematics.

Expectations

By betting on several events, a spectator can arrange matters so that he may receive any of several different amounts of money, depend-

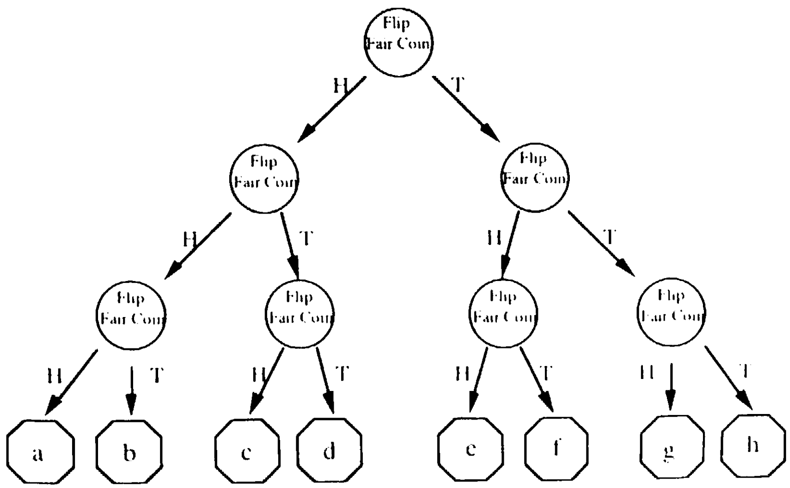

Figure 3. Events as Sets of Stop Signs

ing on how events come out. When the spectator does this, we say that he has bought an *expectation*.

If the sequence of coin flips is short enough for us to draw the tree of situations, then we can display an expectation by writing the amount it pays in each stop sign. We have done this in Figure 4 for an expectation that pays a dollar for every head in three flips.

There are many different ways a spectator can compound simple bets, or simple tickets on events, in order to obtain the expectation displayed in Figure 4. One way is to buy a $3 ticket on {a}, a $2 ticket on {b,c,e}, and a $1 ticket on {d,f,g}. Another way is to buy a $1 ticket on {a,b,c,d,e,f,g}, a $1 ticket on {a,b,c,e}, and a $1 ticket on {a}.

We know from general principles that any two ways of buying an expectation will have the same total cost. Were this not so, a spectator could make money buying the expectation in the one form and selling it in the other.

So expectations have well-defined fair prices. We call these prices fair for the same reason we call the prices of tickets on arbitrary

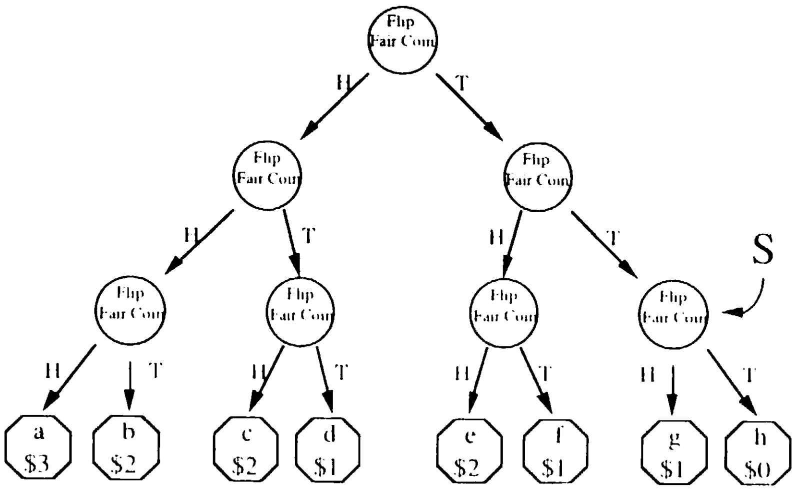

Figure 4. An Expectation

events fair. These prices do not allow a spectator to make money simply by buying and selling, and they do not give a spectator any real chance of parlaying a small initial stake into a large fortune.

As it turns out, there is simple recipe for the fair price of an expectation. Provided, as we have assumed, that the stop signs all lie at the same depth, the fair price of an expectation is simply the average of its payoffs in these stop signs. The fair price of the expectation in Figure 4, for example, is

$$(\$3+\$2+\$2+\$1+\$2+\$1+\$1+\$0)/8 = \$1.50.$$

This is the fair price at the outset, in the initial situation.

More generally, the fair price in an arbitrary situation is the average of the expectation's payoffs in the stop signs below the situation. In situation S in Figure 4, the fair price of the expectation displayed in the figure has gone down to 50¢, the average of $1 and $0.

In the case where the situation is a stop sign, this recipe simplifies. There is no need to average, since there is only one payoff possible in the situation. This payoff is the value of the expectation. The value of Figure 4's expectation in stop sign e, for example, is $2.

Any assignment of payoffs to stop signs defines an expectation. We require only that these payoffs be non-negative amounts of money, measured in dollars.

The simplest expectations are tickets on events. The expectation in Figure 5, for example, is a $1 ticket on the event that we get exactly two heads. It is worth 37.5¢ at the outset and nothing in situation S. Thus the probability of this event is 0.375 at the outset and zero in situation S.

As this example illustrates, our recipe for the fair price of an expectation--take the average of the payoffs--specializes to a recipe for the probability of an event. The probability of a given event in a given situation is an average of ones and zeros assigned to stops signs below the situation--a one for every stop sign in which the event has happened, and zero for every stop sign in which it has failed. This average can also be described as a ratio. It is the ratio of the number of stop signs in which the event has happened to the total number of stop signs, counting in both cases only those stop signs that lie below the situation.

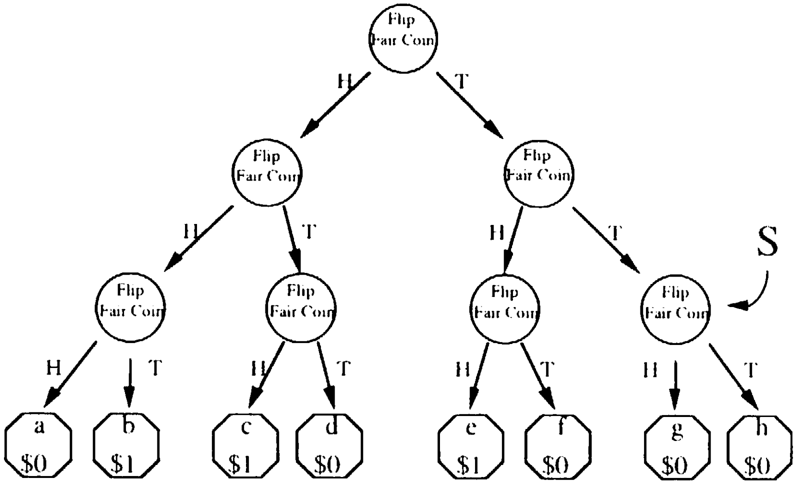

Figure 5. A Ticket on an Event is also an Expectation

Returning to Figure 4, we may now check that the price we assigned the expectation there, \$1.50, is also the total we pay when we buy the expectation by compounding tickets. Applying our recipe to the individual tickets, we find that a \$3 ticket on {a} costs 37.5¢, a \$2 ticket on {b,c,e} costs 75¢, and a \$1 ticket on {d,f,g} costs 37.5¢, for a total of \$1.50. Similarly, a \$1 ticket on {a,b,c,d,e,f,g} costs 87.5¢, a \$1 ticket on {a,b,c,e} costs 50¢, and a \$1 ticket on {a} costs 12.5¢, again for a total of \$1.50.

Notice also that our recipe for the probability of an event gives the right answer for the probability of heads or tails on any individual trial. If we look, in Figure 4, at any situation that is not a stop sign, at the two arrows below it, and at the stop signs below these arrows, we see that half the stop signs lie below the arrow labeled H for heads, and the other half lie below the arrow labeled T for tails. So the probability for both heads and tails on that trial is 0.5; the odds are even.

We hasten to emphasize that our recipes are only conceptual in the case of a long sequence of flips. In the case of three flips, we can specify events by listing the stop signs to be included, we can specify expectations by telling their payoffs in every stop sign, and we can compute the probabilities of events or the fair prices of expectations by going through the stop signs one by one. None of this is practical if there are many flips.

We do not really look at arbitrary events and expectations in the ideal picture. We only look at events and expectations that can be described in some simple way. It may be difficult, moreover, to compute probabilities and fair prices even for events and expectations that we can describe simply. It is difficult, for example, to compute the probability that the frequency of heads in a long sequence of flips will be between 45% and 55%. The numbers we gave in Tables 2 and 3 have been known for a long time, but they were the product of over a century of mathematical research (Stigler, 1986).

Since we are concerned primarily with the philosophical understanding of probability, we will not investigate clever ways of computing probabilities of events and values of expectations. We should note, however, that this task has always been the central concern of mathematical probability, from the time of Pascal and Fermat right up to the present day.

We should also note that our use of the word expectation is very old-fashioned. Readers of De Moivre and Bayes will be familiar with this usage (Shafer, 1982; 1985), but readers of modern work in probability might have expected us to write instead about random payoffs or random variables. Unfortunately, random, like probability, is a loaded word. For the moment, we avoid it in favor of the possibly less divisive older word.

Strategies

A spectator is free to make new gambles, or to buy and sell expectations, at each step as the sequence of flips proceeds. In terms of the tree of situations, this means that the spectator can buy and sell expectations in each situation. The only restrictions are those imposed by the spectator's means and obligations. The spectator cannot pay more for an expectation in a given situation than he or she has in that situation, and the spectator cannot sell an expectation in a given situation if there is a stop sign below that situation in which he or she would not be able to pay off on this expectation together with any others he or she has already sold.

A strategy is a plan for how to gamble as the flips proceed. To specify a strategy, we specify what expectations to buy and sell in each situation, subject to the restrictions just stated.

When we first used the word strategy, in Section I, we explicitly said that a strategy could take the outcomes of preceding flips into account. This is implicit when we talk about situations. A situation S is defined by how the flips have come out so far, so when we say that the spectator is to buy and sell given expectations in S, we are specifying what the spectator is to do when the preceding flips have come out in this way.

As we have said repeatedly, it is a feature of the ideal picture that strategies are to no avail. Nothing can be gained by clever buying and selling.

This means that a strategy, no matter how complicated, is always equivalent to the purchase of some expectation at the outset. The spectator has some initial capital at the outset, and for each stop sign, the strategy will determine the spectator's net payoff if he or she arrives at that stop sign. So the strategy boils down to using the initial capital to purchase an expectation with these payoffs. And since the strategy can gain nothing, the initial capital must be the fair price for this expectation.

IV. THE GENERAL IDEAL PICTURE

Once we understand the tree of situations for the fair coin, it is easy to generalize. Instead of requiring that a fair coin be flipped in each situation, we permit a variety of experiments, we permit experiments with more than two possible outcomes, and we permit different odds and even different experiments in different situations.

As we explain in this section, knowledge of the long run, value, and belief are intertwined in this more general picture just as they are intertwined in the case of the fair coin. The odds specified for the

outcome of a given experiment are merely the odds in the situation where the experiment is performed, but together these odds determine odds for all events in all situations and prices for all expectations in all situations. The prices are fair in the same way that prices are fair in the case of the fair coin. The spectators cannot make money merely by buying and selling at these prices, they know that they will approximately break even if they buy a long sequence of expectations depending on successive trials, and they know that no strategy for buying and selling expectations will enable them to substantially multiply their initial capital.

An Example

Figure 6 is an example of a tree of situations that permits several different experiments. The first experiment in this tree is a flip of a fair coin, but the second experiment is another flip of a fair coin only if the first flip produces a head. If the first flip produces a tail, then the second experiment is a flip of a coin that is biased 3 to 1 for heads. Depending on how earlier flips come out, later steps may include flipping another fair coin, flipping a coin that is biased 4 to 1 for heads, or throwing a fair die. Altogether, there will be three or four experiments, depending on the course of events.

The tree of situations in the general ideal picture is much deeper than the tree in Figure 6, of course. We assume that all paths down the tree are long. No matter what turns events take, the experimenter must take many steps before coming to a stop sign.

The tree of situations specifies the experiment to be performed in each situation. This includes specifying the possible outcomes for the experiment and, explicitly or implicitly, the odds for these outcomes. In Figure 6, the odds for the outcomes are specified by specifying the bias or lack of bias for each coin, and by saying that the die is fair.

When we discussed the ideal picture for the fair coin, we said that the odds on a particular flip are not affected by the outcomes of earlier flips. We cannot say this about Figure 6; there the odds for a flip sometimes do depend on earlier outcomes. Even the experiment performed may depend on what has happened before. Both the experiment performed and the odds for its outcomes depend on the situation.

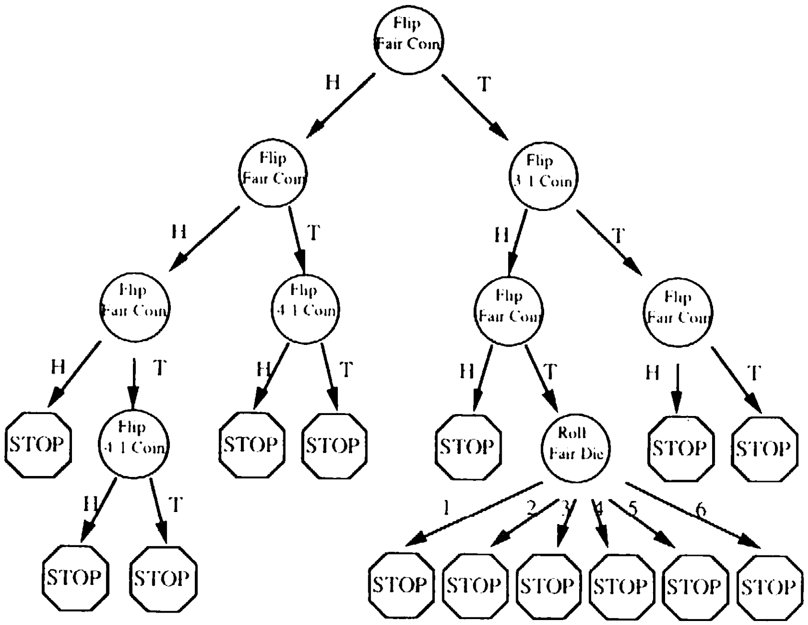

Figure 6. A General Tree of Situations

Knowledge of the Long Run

In the case of the fair coin, the spectators know that about half the flips will come up heads, and they know nothing that can enable them to take advantage of even-money bets on successive flips. No strategy for selecting even-money bets or for buying and selling expectations can assure the spectator even modest gains, and no strategy can give the spectator any real hope of substantially multiplying initial capital.

We can make similar statements in the general ideal picture, except that here we are talking about bets at the odds specified by each experiment, rather than always about even-money bets. A spectator knows that if he or she makes many small bets on the outcomes of successive experiments, at the odds specified by these experiments, he or she will approximately break even. Even if large bets and complicated expectations are allowed, no strategy can assure the spectator modest gains or give the spectator any real hope of substantially multiplying initial capital.

The preceding paragraph is cast entirely in terms of gambling, whereas in the case of the fair coin, we were able to make a statement purely about frequency. We said that the coin would land heads about half the time in the whole sequence of trials. Can we make a statement about frequency in the general case?

Yes, we can, with some effort. We can get a statement about frequency out of the statement that the spectator will break even on many small bets.

Suppose that at the outset, before any experiments are performed, the spectator adopts a betting strategy. He or she chooses an outcome from the experiment in each situation, and plans to bet on that outcome if that situation arises. The spectator may decline to bet in some situations, but he or she does specify bets in enough situations to be sure to make many bets no matter which path down the tree is taken by the course of events. The bets are all of a specific form. If the situation arises, the spectator buys a \$1 ticket on the outcome chosen. In other words, the spectators puts \$p on the outcome, where p is its probability.

Under these circumstances, the statement that the spectator approximately breaks even becomes a statement about the frequency with which the outcomes he or she chooses happen in the experiments actually performed: *This frequency approximately equals the average of the probabilities for the chosen outcomes.*

To see that this is true, suppose n experiments are performed, and the spectator chooses correctly in k of them. Then the amount the spectator spends on tickets is the total of the probabilities of the n outcomes, and the amount he or she gets back is \$k. If the spectator breaks even, these amounts are approximately equal. Divide both by n. When we divide the total of the probabilities by n, we get the average probability. When we divide the \$k by n, we get k/n, the frequency with which the chosen outcome happens. So the frequency approximately equals the average probability. (See Dawid, 1982; Shafer, 1985.)

In the case of the fair coin, this general statement reduces, of course, to the statement that the coin will land heads approximately half the time. The spectator always bets on heads, the probabilities are all 0.5, and so their average is 0.5.

The Circle of Reasoning

The circle of reasoning for the general ideal picture proceeds as in the special case of the fair coin.

First, from their knowledge of the long run, the spectators argue that the odds specified for the outcomes of each experiment are fair for the situation in which that experiment is performed. These odds are fair because it is common knowledge among the spectators that no matter how they choose among bets at these odds, they will approximately break even, they cannot count on even modest gains, and they cannot even hope for substantial gains.

Second, the spectators recognize that once they have assigned odds to the outcomes of the individual experiments, there is just one way to extend this assignment to an assignment of odds to all events in all situations so that Peter cannot make money by buying and selling tickets. Again, the fact that there is at least one way strengthens the case for calling the individual odds fair, and the fact that there is only one way provides an argument for calling the resulting odds on more complicated events fair.

Third, the spectators deduce that the odds they have assigned include, in the initial situation, very great odds that any betting strategy will approximately break even, very great odds against any strategy multiplying initial capital by more than a few orders of magnitude, and no particularly good odds for even modest gains.

Fourth, the spectators interpret their fair odds as measures of warranted belief.

Fifth, the spectators interpret very great warranted belief as practical certainty or knowledge. This brings them back to the knowledge of the long run with which they began, except that this knowledge is now expressed in quantitative form.

In the general case, as in the special case of the fair coin, frequency, belief, and value are intertwined. We can argue from one to the another of the three, but none of the three can stand on its own.

REFERENCES

Dawid, A. P. (1982), "The Well-Calibrated Bayesian," *Journal of the American Statistical Association*, 77, pp. 605-13.

Feller, William (1968), *An Introduction to Probability Theory and Its Applications, Volume I*, New York: Wiley, 3rd Ed.

de Finetti, Bruno (1937), "La Prévision: ses Lois Logiques, ses Sources Subjectives," *Annales de l'Institut Henri Poincaré*, 7, pp. 1-68.

Kyburg, Henry E., Jr. and Smokler, Howard E., eds. (1980), *Studies in Subjective Probability*, Huntington, New York: Robert E. Krieger.

Martin-Löf, Per (1969), "The Literature on von Mises' Kollektivs Revisited," *Theoria*, 35(1), pp. 12-37.

von Mises, Richard and Geiringer, H. (1964), *The Mathematical Theory of Probability and Statistics*, New York: Academic Press.

Savage, Leonard J. (1954), *The Foundations of Statistics*, New York: Wiley.

Shafer, Glenn (1982), "Bayes's Two Arguments for the Rule of Conditioning," *The Annals of Statistics*, 10(4), pp. 1075-89.

Shafer, Glenn (1985), "Conditional Probability," *International Statistical Review*, *53*, pp. 261-77.

Shafer, Glenn (1990), *The Unity and Diversity of Probability*, forthcoming.

Stigler, Stephen M. (1986), *The History of Statistics: The Measurement of Uncertainty Before 1900*, Cambridge, MA: The Belknap Press of Harvard University Press.

Wald, Abraham (1937), "Die Wiederspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung," in K. Menger, ed., *Ergebnisse eines mathematischen Kolloquiums*, Vol. 8, Leipzig and Vienna: F. Deuticke, pp. 38-72.

Ville, Jean (1939), *Étude Critique de la Notion de Collectif*, Paris: Gauthier-Villars.