Can the various meanings of probability be reconciled?[1]

Glenn Shafer[2]

Abstract

The stand-off between the frequentist and subjectivist interpretations of probability has hardened into a philosophy. According to this philosophy, probability begins as pure mathematics. The different meanings of probability correspond to different interpretations of Kolmogorov's axioms.

This chapter urges a slightly different philosophy. Probability begins with the description of an unusual situation in which the different meanings of probability are unified. It is this situation—not merely the mathematics of probability—that we use in applications. And there are many ways of using it.

This philosophy reconciles the various meanings of probability at a level deeper than the level of axioms. It allows us to bring together in one framework the unified eighteenth-century understanding of probability, the frequentist foundations of von Mises and Kolmogorov, and the subjectivist foundations of de Finetti. It allows us to recognize the diversity of applications of probability without positing a myriad of incompatible meanings for probability.

1 An agreement to disagree

For over fifty years, there has been a consensus among philosophers, statisticians, and other probabilists about how to think about probability and its applications. According to this consensus, probability is first of all a theory in pure mathematics, based on Kolmogorov's axioms and definitions. Different interpretations of these axioms are possible, and the usefulness of each interpretation can be debated, but the mathematical theory of probability stands above the debate. As the historian Lorraine Daston puts it, “The mathematical theory itself preserves full conceptual independence from these interpretations,

---

[1] To appear in *Methodological and Quantitative Issues in the Analysis of Psychological Data,* Second Edition, edited by Gideon Keren and Charles Lewis, and published by Lawrence Erlbaum, Hillsdale, New Jersey.

[2] Ronald G. Harper Distinguished Professor of Business, School of Business, Summerfield Hall, University of Kansas, Lawrence, Kansas 66045. Research for this article has been partially supported by the National Science Foundation through grant IRI8902444 to the University of Kansas and grant BNS8700864 to the Center for Advanced Study in the Behavioral Sciences. The author has benefited from conversations with Robert Fogelin, David Israel, Ali Jenzarli, Don Ylvisaker, and Joe VanZandt.

however successful any or all of them may prove as descriptions of reality" (Daston 1988, pp. 3-4).

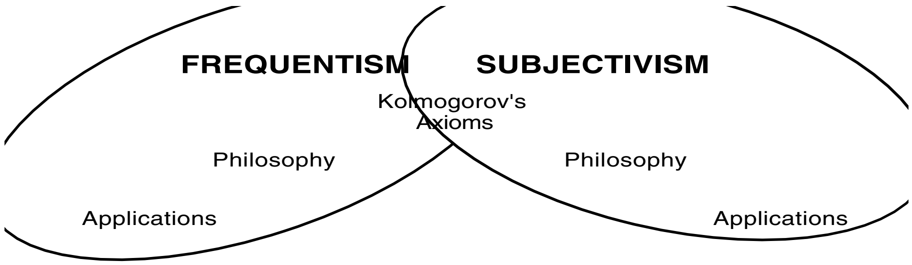

Figure 1. The consensus

The consensus is depicted in Figure 1. The subjectivists, who interpret probability as degree of belief, and the frequentists, who interpret it as relative frequency, have only the purely mathematical theory as common ground. Both subjectivists and frequentists find applications for probability, but these applications are separated from the common ground by the opposing philosophies. They are based on different meanings for probability.

In practice, this consensus is an agreement to disagree. The two camps, the frequentists and the subjectivists, agree on the mathematics of probability, but they also agree that everyone has a right to give whatever interpretation they please to this mathematics. Though they continue to debate the fruitfulness of their different interpretations, they have come to realize that they are talking past each other. They have no common language beyond the mathematics on which they agree so perfectly.

The consensus has become so ingrained in our thinking that it seems natural and unavoidable. All mathematics has been axiomatic since the work of David Hilbert, and any axiomatic system, as Kolmogorov himself pointed out, admits "an unlimited number of concrete interpretations besides those from which it is derived" (Kolmogorov 1950, p. 1). So every branch of pure mathematics can declare its conceptual independence of its applications. As Daston puts it, "For modern mathematicians, the very existence of a discipline of applied mathematics is a continuous miracle—a kind of prearranged harmony between the 'free creations of the mind' which constitute pure mathematics and the external world."

We should remember, however, that not all fields that use mathematics have ceded primacy to pure mathematics drained of meaning to the extent probability has. In physics, for example, axioms are secondary to physical theory, which melds mathematics and meaning in a way that goes beyond any single set of axioms. The physicist is usually interested in a physical theory that can be axiomatized in different and sometimes incompatible ways, not in a single axiomatic theory that can be interpreted in incompatible ways.

I remember vividly a lecture by one of my own physics teachers, in which he derived one physical relation from another and then gave a second derivation that went more or less in the opposite direction. When a student pointed out the near circularity, he launched into a passionate discussion of the difference between the physicist and the mathematician. This blackboard is the world, he said. The mathematician wants to find a single starting place—a particular dot of chalk—from which to derive everything else. The physicist does not see the point of this. The physicist takes whatever starting point is convenient for getting where he or she wants to go. Sometimes the physicist goes from here to there, sometimes from there to here—he drew arrows all over the blackboard. The point is to see how things hang together and to understand parts you did not understand before, not to get everywhere from one place.

My purpose in this chapter is to urge that we once again look at probability the way physicists look at a physical theory. Probability is not a physical theory, but it does have an object. The axioms are about something. This something is an unusual situation—a situation that occasionally occurs naturally, sometimes can be contrived, and often can only be imagined. In this unusual situation, "probability" is not devoid of meaning. It has many meanings, just as "energy" or "work" have many meanings within the situation described by the theory of mechanics. The numerical probabilities in the unusual situation described by the theory of probability are simultaneously fair prices, warranted degrees of belief, and frequencies.

Since the unusual situation the theory of probability describes occurs infrequently and may be imperfect even when it does occur, I will call it "the ideal picture of probability."

Outline of the chapter. The next section, Section 2, describes informally the simplest case of the ideal picture of probability, the case where a fair coin is flipped repeatedly. We see there how the ideal picture ties frequencies, fair prices for gambles, and warranted degrees of belief together in a circle of reasoning, any point of which can be used as a starting point for an axiomatic theory.

Section 3 refines the informal account of Section 2 into a mathematical framework and formulates axioms for fair price and probability that resemble Kolmogorov's axioms yet capture aspects of the ideal picture that are left outside Kolmogorov's framework.

Section 4 relates the ideal picture to the philosophical history of mathematical probability. The ideas that make up the ideal picture had been developed and even unified to some extent by the end of the eighteenth century.

But this unity fell victim to the extreme empiricism of the nineteenth century, which saw frequency as an acceptable basis for a scientific theory but rejected fair price and warranted degree of belief as metaphysical fictions. In the twentieth century, the subjectivists have matched the frequentists' empiricism with a story about personal betting rates that sounds like an empirical description of people's behavior. Both the frequentist and subjectivist foundations for probability have elements of truth, but they become fully cogent only when they are brought back together and seen as alternative descriptions of the same ideal picture.

The mistake that nineteenth-century empiricists made about the mathematical theory of probability was to suppose that it could be used only by fitting it term-by-term to some reality. They believed that using the theory meant finding numbers in the world—frequencies or betting rates—that followed the rules for probabilities. In the late twentieth-century, however, we can take a more flexible view of the relation between theory and application. We can take the view that the mathematical theory of probability is first of all a theory about an ideal picture, and that applying the theory to a problem means relating the ideal picture to the problem in any of several possible ways.

Section 5 discusses some of the ways the ideal picture can be used. Some statistical modeling uses the ideal picture as a model for reality, but much statistical modeling uses it only as a standard for comparison. Another way to use the ideal picture is to draw an analogy between the evidence in a practical problem of judgment and evidence in the ideal picture. We can also use simulations of the ideal picture—sequences of random numbers—to draw samples and assign treatments in experiments, so that probabilities in the ideal picture become indirect evidence for practical judgments.

2 An informal description of the ideal picture

The ideal picture of probability is more subtle than the pictures drawn by most physical theories, because it involves knowledge as well as fact. Probability, in this picture, is *known* long-run frequency. The picture involves both a sequence of questions and a person. The person does not know the answers to the questions but does know the frequencies with which different answers occur. Moreover, the person knows that nothing else she knows can help her guess the answers.

This section briefly describes the ideal picture informally, with an emphasis on its intertwining of fact and knowledge. It deals with the simplest case, the fair coin repeatedly flipped. This simple case is adequate to demonstrate how the ideal picture ties three ideas—knowledge of the long run, fair price, and warranted belief—in a circle of reasoning. We can choose any point in this circle as a starting point for an axiomatic theory, but no single starting point does full justice to the intertwining of the ideas.

The picture of the fair coin generalizes readily to biased coins and experiments with more than two possible outcomes, and to the case where the experiment to be performed may depend on the outcomes of previous experiments. These more general cases are not considered in this section, but they are accommodated by the formal framework of Section 3.

For a more detailed description of the ideal picture, see Shafer (1990a).

2.1 Flipping a fair coin

Imagine a coin that is flipped many times. The successive flips are called trials. Spectators watch the trials and bet on their outcomes. The knowledge of these spectators is peculiarly circumscribed. They know the coin will land heads about half the time, 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. And they know this.

Just before each trial, the spectators have an opportunity to make small even-money bets on heads or on tails. But since they are unable to predict the outcomes, they cannot take advantage of these opportunities with any confidence. Each spectator knows she will lose approximately half the time. A net gain, small relative to the amount of money bet, is possible, but a comparable net loss is also possible. No plan or strategy based on earlier outcomes can assure a net gain. For all these reasons, the spectators consider even-money bets on the individual trials fair.

Since a spectator begins with only a limited stake, she may be bankrupted before she can make as many bets as she wants. She can avoid bankruptcy by making the even-money bets smaller when her reserves dwindle, but this will make it even harder to recover lost ground. Consequently, she can hope only for gains comparable in size to her initial stake. No strategy can give her any reasonable hope of parlaying a small stake into a large fortune. This is another aspect of the fairness of the even-money bets.

The spectators also bet on 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, or on the event that it comes up heads on exactly five hundred of the first thousand trials. They agree on fair odds for all such events. These odds change as the trials involved in the events are performed. They are fair for the same reasons that the even odds for individual trials are fair. A spectator betting at these odds cannot be confident of any gain and has no reasonable hope of parlaying a small stake into a large fortune. Moreover, if she makes many small bets involving different trials, she will approximately break even.

Fairness has both long-run and short-run aspects. The statement about bets involving many different trials is strictly a statement about the long run. But the other statements apply to the short run as well. No way of compounding bets, whether it involves many trials or only a few, can make a spectator certain of gain or give her a reasonable hope of substantially multiplying her stake.

Precise statements about the long run are themselves events to which the spectators assign odds. They give great odds that the coin will land heads on approximately half of any large number of trials. They give 600 to 1 odds, for example, that the number of heads in the first thousand tosses will be between 450 and 550. They also give great odds against any strategy for increasing initial capital by more than a few orders of magnitude. They give at least 1,000 to 1 odds, for example, against any particular strategy for parlaying $20 into $20,000. Thus the knowledge of the long run that helps justify the fairness of the odds is expressed directly by these odds.

Just as very great odds seem to express knowledge,[1] less great but substantial odds seem to express guarded belief. The spectators' degree of belief in an event is measured numerically by the odds they give. Since the odds are warranted by knowledge of the short and long runs, this numerical degree of belief is not a matter of whim; it is a warranted partial belief.

The spectator's numerical degrees of belief express quantitatively 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. The spectators' certainty that long shots, or very ambitious gambling strategies, will fail is a limiting case of their skepticism about all gambling strategies, more ambitious and less ambitious. They give at least k to 1 odds against any strategy for multiplying initial capital by k—two to one odds against doubling initial capital, thousand to one odds against increasing initial capital a thousandfold, and so on. Similarly, their certainty that heads will come up half the time in the long run is a limiting case of their belief that the proportion of heads will not be too far from one-half in the shorter run. The degree of belief and the degree of closeness expected both increase steadily with the number of trials.

---

[1] There is a consensus in philosophy that knowledge is justified true belief. We cannot know something unless it is true. By equating knowledge with mere great odds, I may appear to challenge this consensus. The spectators can know something that might not be true. It is not my intention, however, to enter into a debate about the nature of knowledge. I merely ask leave to use the word in an ordinary sloppy way.

2.2 A circle of reasoning

Our description of the ideal picture traced a circle. We started with knowledge of the long run. Then we talked about the odds warranted by this knowledge. Then we interpreted these odds as a measure of warranted belief—i.e., as a measure of probability. And we noted that the knowledge of the long run with which we began was expressed by certain of these odds.

This circle of description can be refined into a circle of reasoning. The spectators can reason from their knowledge of the long run to the assignment of fair odds to individual trials. They can argue from the odds for individual trials to odds for events involving more than one trial. They can argue that all these odds should be interpreted as degrees of warranted belief (or probabilities). Then they can deduce very high probabilities for events that express the knowledge of the long run with which they began.

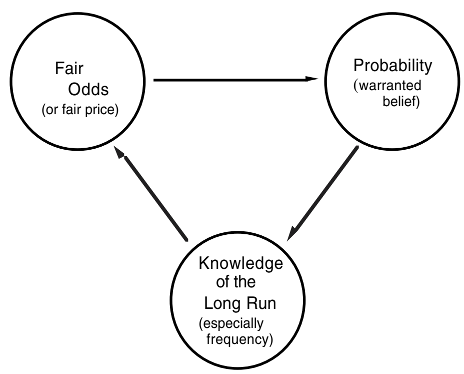

Figure 2. The circle of probability ideas

This circle of reasoning is depicted by Figure 2. The first step is represented by the arrow from "Knowledge of the long run" to "Fair odds." The spectators move along this arrow when they argue that even odds for individual trials are sensible and fair, because these odds take all their relevant knowledge into account, and because someone who makes many bets at these odds will approximately break even, and so on.

The next step can be located inside "Fair odds." This is the step from odds on individual trials to odds on all events. As it turns out, once we agree on 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 events involving more than one trial so that a person cannot make money for certain by compounding bets at these odds.

The next step, represented by the arrow from "Fair odds" to "Probability," is to interpret fair odds as a measure of warranted belief. The spectators 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 measure their beliefs.

Within "Probability," the spectators deduce that their degrees of belief, or probabilities, for complicated events include very high probabilities that the coin will land heads approximately half the time in any particular long run of trials and that any particular scheme for parlaying small sums into large ones will not succeed. This allows them to travel the final arrow, from "Probability" back to "Knowledge of the long run."

2.3 Making the picture into mathematics

The reasoning we have just described is not axiomatic mathematics. Much of it is rhetorical rather than deductive. And it goes in a circle. This is typical of informal mathematical reasoning. When we axiomatize such reasoning, we choose a particular starting point. We then use the rhetorical reasoning to justify definitions, and the deductive reasoning to prove theorems.

In Figure 2, the arrows represent the major rhetorical steps and hence the major potential definitions. The spectators can define odds on the basis of their long-run knowledge, they can define warranted belief in terms of odds, and they can define knowledge as very great warranted belief. The circles joined by the arrows represent the potential starting points. An axiomatic theory can be based on axioms for knowledge of the long run, axioms for fair odds, or axioms for warranted belief.

In deference to the weight of popular opinion in favor of the frequentist interpretation of probability, I began this description of the ideal picture with knowledge of the long run. But, as we will see in Section 3, it is actually easier to begin an axiomatic theory with fair odds or with warranted belief.

The fact that knowledge of the long run, fair odds, and warranted belief can each be used as a starting point for an axiomatic theory should not be taken to mean that any one of these ideas is sufficient for grounding the theory of probability in a conceptual sense. The axioms we need in order to begin with any one of these starting points 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.

Section 4 will support this claim with the historical record. Historically, the three starting possible points are represented by Kolmogorov's axioms (probability), von Mises's random sequences (long-run frequency), and de Finetti's two-sided betting rates (odds or price). Kolmogorov's axioms were always intended as a formal starting point, not a conceptual one; everyone agrees that they must be justified either by a frequency or betting interpretation. Von Mises did want to make long-run frequency a self-sufficient starting point, but his work, together with that of Wald and Ville, leads to the conclusion that knowledge of long-run frequency is only one aspect of the knowledge that justifies calling the odds in the ideal picture fair. De Finetti wanted to make odds or price a self-sufficient starting point, without any appeal to the long-run to justify the fairness of odds or prices, but this too fails to provide a full grounding for the ideal picture.

2.4 Conclusion

The situation described in this section is only one version of the ideal picture of probability. Like the situation described by any physical theory, the ideal picture has many variations, not all of which are strictly compatible with each other. It would be unwise, therefore, to claim too much for the story told here. But the intertwining of knowledge, fair odds, and belief described here occurs, in one way or another, in all the visions that have informed the growth of mathematical probability.

3 A formalization of the ideal picture

The preceding section pointed to several possible axiomatizations of the ideal picture. This section develops a formal mathematical framework in which some of these axiomatizations can be carried out.

The most fundamental feature of any mathematical framework for probability is its way of representing events. Kolmogorov represented events as subsets of an arbitrary set. The framework developed here is slightly less abstract. Events are subsets, but the set of which they are subsets is structured by a *situation tree*, which indicates the different ways events can unfold. This brings into the basic structure of the theory the idea of a sequence of events and hence the possibility of talking about frequencies.

The section begins with an explanation of the idea of a situation tree. It then shows how an axiomatic theory can be developed within this situation tree. We start with axioms for fair price, and we translate them into axioms for probability. We show how knowledge of the long run can be deduced from these axioms. We conclude by briefly comparing the axioms with Kolmogorov's axioms.

To validate completely the claims made in the preceding section, we should also develop axioms for knowledge of the long run. This task was undertaken, in a certain sense, by von Mises, Wald, and especially Kolmogorov, in his work on complexity theory and the algorithmic definition of probability. We will glance at this work in Section 4.3, but it would stretch this chapter too far, in length and mathematical complexity, to review it in detail and relate it to the other ideas in Figure 2.

The framework developed in this section is more general than the story about the fair coin. This framework permits biased coins, as well as experiments with more than two outcomes, and it also permits the choice of the experiment to be performed on a given trial to depend on the outcomes of preceding trials. It does not, however, encompass all versions of the ideal picture. It does not, for example, allow the spectators to choose the sequence in which they see the outcomes of trials.

3.1 The framework for events

Situation trees provide a framework for talking about events, situations, expectations, and strategies.

Situation trees. Figure 3 is one example of a situation tree. It 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. Each circle and each stop sign is a *situation* that can arise in the course of the flips. The circle at the top is 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.

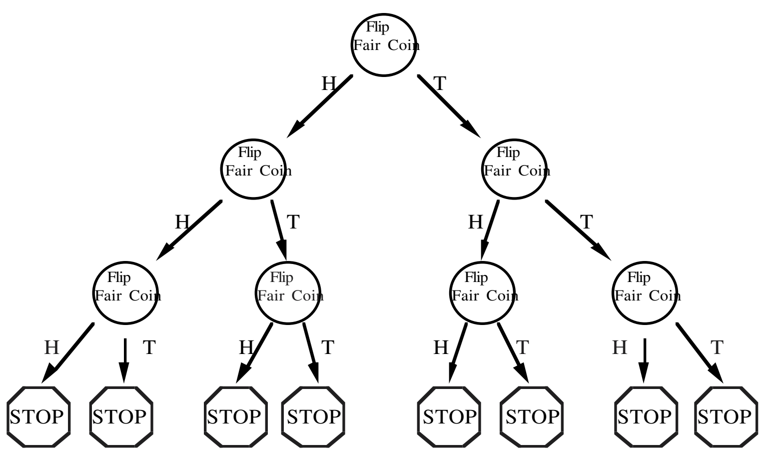

Figure 3. A situation tree for three flips of a fair coin

Figure 4 depicts another situation tree, one that involves several different experiments. The first experiment is a flip of a fair coin. Depending on how it comes out, the second is either another flip of a fair coin or a flip of a coin that is biased 3 to 1 for heads. Later experiments may include flipping another fair coin, flipping a coin 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 odds for each experiment are specified in some way; we specify the bias or lack of bias for each coin, and we say that the die is fair.

The ideal picture involves a situation tree like Figure 3 or Figure 4, except that all the paths down the tree are very long. In each situation, we specify an experiment with a finite number of possible outcomes, and we specify in some way the odds for these outcomes.

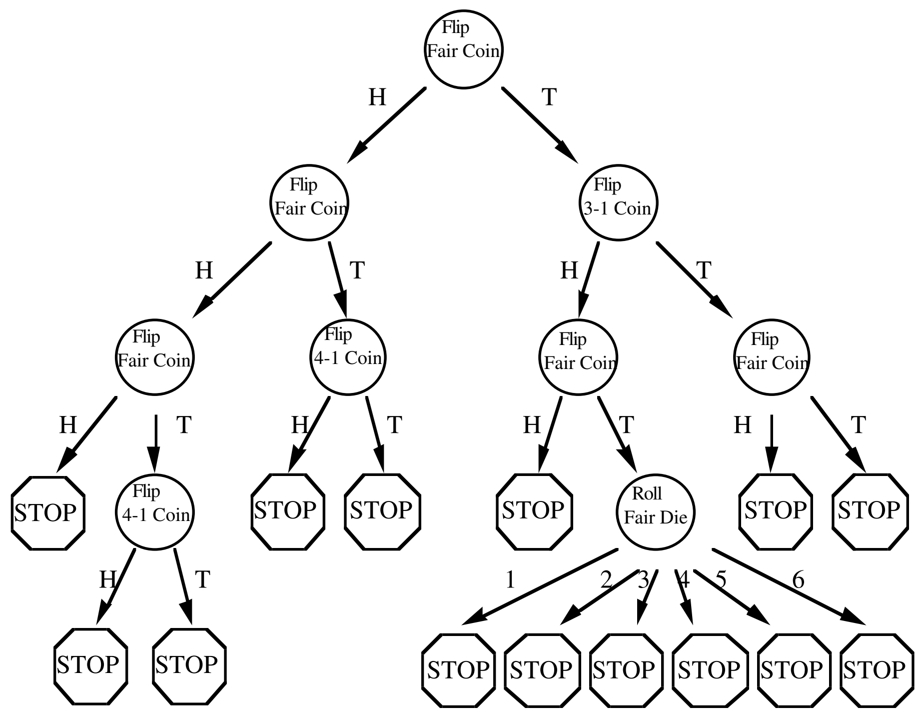

Figure 4. A more complicated situation tree

Events. An event is something that happens or fails as we move down the situation tree. Getting heads on the first flip is an event. Getting exactly two heads in the course of the first three flips is an event. Formally, we can identify an event with a set of stop signs—the set consisting of those stop signs in which the event has happened. We can illustrate this using the lettered stop signs of Figure 5. Here the event that we get heads on the first flip is the set {a,b,c,d} of stop signs. The event that we get exactly two heads is the set {b,c,e}. And so on.

Notice that for each situation there is a corresponding event—the set of stop signs that lie below it. This is the event that we get to the situation. It is often convenient to identify the situation with this event. We identify the situation S in Figure 5, for example, with the event {g,h}. Not all events are situations. The event {b,c,e} in Figure 5, for example, is not a situation.

We say that the event A is *certain* in the situation S if S is contained in A. We say that A is *impossible* in S if the intersection of A and S is empty.

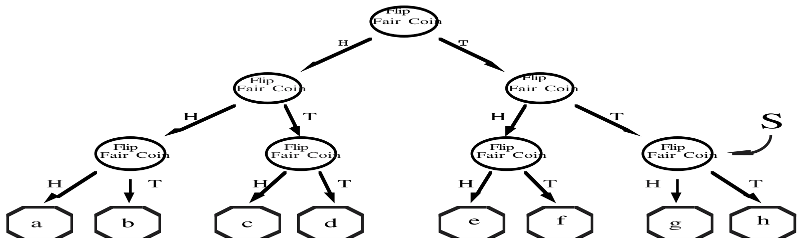

Figure 5. Events as sets of stop signs

Expectations. Let us call a function that assigns a real number—positive, zero, or negative—to every stop sign an *expectation*.[1] We call the numbers assigned by an expectation *payoffs*. A positive payoff is the number of dollars the holder of the expectation will receive in that stop sign; a negative payoff is the number of dollars the holder must pay. Figure 6 shows an expectation that pays the holder a dollar for every head in three flips.

Let us use upper class letters from the end of the alphabet—X, Y, Z, and so on—for expectations, and let us write X(i) for X's payoff in the stop sign i. Expectations can be added; we simply add their payoffs in each stop sign. The expectation X+Y has the payoff X(i)+Y(i) in stop sign i. We can also add constants to expectation. The expectation X+r has the payoff X(i)+r in i.

An $r *ticket* on an event A is an expectation that pays $r if A happens and $0 if A does not happen. Figure 7 shows a $1 ticket on the event {b,c,e}. We write ⟨$r,A⟩ for an $r ticket on A.

Suppose you bet $p on an event at odds p to (1-p), where 0 ≤ p ≤ 1. This means that you pay $p, you will get a total of $1 back if the event happens, and you will get nothing back if the event fails. Thus you have paid $p for a $1 ticket on the event. So stating odds on an event is equivalent to setting a price for a ticket on the event. Saying that p:(1-p) is the fair odds on A is the same as saying that $p is the fair price for a $1 ticket on A.

The sum of two tickets is an expectation. It is not always a ticket; but sometimes it is; for example, ⟨$r,A⟩ + ⟨$s,A⟩ = ⟨$(r+s),A⟩.

[1] This is now usually called a “random variable.” I use the eighteenth-century term, “expectation,” in order to avoid evoking twentieth-century presumptions about the meaning of randomness.

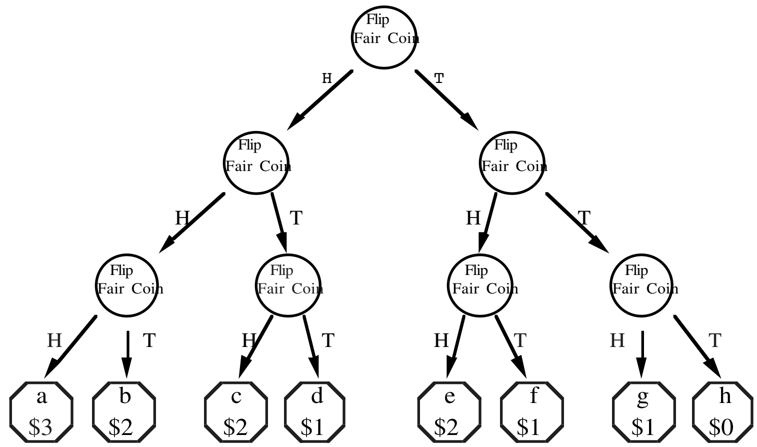

Figure 6. An expectation

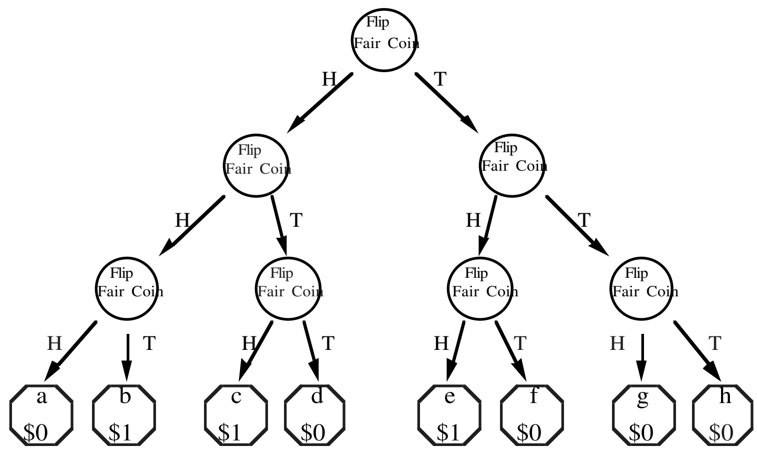

Figure 7. A $1 ticket on {b,c,e}

Every expectation is the sum of tickets, but a given expectation can be obtained as a sum of tickets in more than one way. The expectation in Figure 6, for example, is the sum of a $3 ticket on {a}, a $2 ticket on {b,c,e}, and a $1 ticket on {d,f,g}, but it is also the sum of 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}.

In general, gambling means buying and selling expectations. We can think of this in several ways. On the one hand, we can think of it in terms of tickets on events. Since all expectations are sums of tickets, gambling boils down to buying and selling tickets. On the other hand, we can think in terms of the total expectation we acquire by all our buying and selling. If we buy a collection Φ_1 of tickets for \$r, and we sell a collection Φ_2 of tickets for \$s, then the net result is that we have added the expectation

$$\sum_{X \varepsilon \Phi_1} X - \sum_{X \varepsilon \Phi_2} X - \$r + \$s$$

to whatever expectation we already had.

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

A *strategy* is a plan for how to gamble as the experiments proceed. To specify a strategy, we specify what expectations to buy and sell in each situation, subject to the restrictions just stated. In Section 2, we said that a strategy could take the outcomes of preceding experiments into account. This is explicit in the framework of a situation tree. A situation is defined by the outcomes so far, so when a spectator specifies what expectations she will buy and sell in a situation, she is specifying what she will do if these are the outcomes.

A strategy boils down, in the end, to an expectation. The spectator's initial capital, say \$r, and her strategy, say $\mathcal{S}$, together determine, for each stop sign i, the capital, say $X_{r,\mathcal{S}}(i)$, that the spectator will have in i. So the strategy amounts to trading the \$r for the expectation $X_{r,\mathcal{S}}$.

The strategy $\mathcal{S}$ is *permissible* in the situation S for a spectator with capital \$r in S (and no other expectations or obligations) only if $X_{r,\mathcal{S}}(i)$ is non-negative for all i in S. Unlike businesspeople in real life, a spectator in the ideal picture is not allowed to undertake obligations that she may not be able to meet.

3.2 Axioms for fair price

Now let us use the framework provided by the situation tree to develop some of the possibilities for axiomatization mentioned in Section 2. It is convenient to begin with fair price. We will formulate axioms for fair price and relate these axioms to the circle of probability ideas in the way suggested by Figure 2. In other words, we will informally justify the axioms by the knowledge we claim of the long run (this is the arrow from knowledge of the long run to fair price), we will use the axioms to derive rules for probability (this is the arrow from fair price to probability), and then we will deduce the knowledge of the long run that motivated the axioms (this is the arrow from probability to knowledge of the long run).

There are a number of ways to formulate axioms for fair odds or fair prices. For this brief exposition, it is convenient to emphasize fair prices for tickets on events.

Let us write $V_S(X)$ for the fair price of the ticket X in the situation S. We will omit the parentheses when we use the bracket notation for a ticket; in other words, we will write $V_S\langle\$r,A\rangle$ instead of $V_S(\langle\$r,A\rangle)$.

Here are our axioms for the ticket prices $V_S(X)$:

Axiom T1. If A is certain in S, then $V_S\langle\$1,A\rangle = 1$.

Axiom T2. If A is impossible in S, then $V_S\langle\$1,A\rangle = 0$.

Axiom T3. If A is possible but not certain in S, then $0 < V_S\langle\$1,A\rangle < 1$.

Axiom T4. If $0 \le r \le t$, then $V_S\langle\$r,A\rangle \le V_S\langle\$t,A\rangle$.

Axiom T5. If the sum of the tickets X and Y is also a ticket, then $V_S(X+Y) = V_S(X) + V_S(Y)$.

Axiom T6. If X and Y are tickets, S precedes T, and $V_T(X) = V_T(Y)$, then $V_S\langle X,T\rangle = V_S\langle Y,T\rangle$.

Axiom T6 extends our notation by using a ticket as a prize in another ticket. The idea is that $\langle X,T\rangle$ is the ticket that pays X if T happens and nothing otherwise. This does not really extend what we mean by a ticket, because the compounded ticket $\langle X,T\rangle$ still boils down to a ticket that pays a certain amount of money if a certain event happens and nothing otherwise. If $X = \langle\$r,A\rangle$, for example, then $\langle X,T\rangle = \langle\langle\$r,A\rangle,T\rangle = \langle\$r,A\cap T\rangle$.

The derivation of these axioms from knowledge of the long run begins with an argument for the existence of fair prices for all tickets. This knowledge of the long run explicitly includes knowledge of fair odds for outcomes of each individual experiment, odds that do not change until that experiment is performed. So we can call $\$r\cdot p$ the fair price in situation S of a \$r ticket on an outcome of an experiment that is to be performed in S or later and for which the fair odds are p to (1-p). (If the experiment is to be performed before S, or only in situations incompatible with S, then the fair price is either \$0 or \$r.) Fairness means that a person breaks even in the long run by betting on these events at these odds, and that no one can compound bets, over the short run or the long run, to make money for certain. By buying tickets on various outcomes in various situations (this may involve buying a ticket in one situation to provide funds to buy a ticket in another situation), we can put together a ticket on any event, so we conclude that there are fair prices for all tickets.

The axioms then follow from the idea that one cannot make money for certain by compounding tickets. Axiom T1, for example, is justified because otherwise one could make money for certain in S merely by buying or selling the ticket $\langle\$1,A\rangle$. Axiom T5 holds because otherwise one could make money for certain in S by buying X and Y separately and selling X+Y, or vice versa. Axiom T6 holds because otherwise one could make money for certain in S by buying ⟨X,T⟩ and selling ⟨Y,T⟩ in S and then, if one arrives in T, selling X and buying Y (or vice versa).

Axiom T3 requires special comment. Strictly speaking, only the weaker statement that $0 \leq V_S\langle\$1,A\rangle \leq 1$ is justified, but the strict inequalities are convenient. Allowing equality would mean, in effect, allowing events to have zero probability even though they are possible. Since our framework is finite—there are a finite number of experiments each with finite number of outcomes—there is no need for this.

Axioms T1-T6 are only about tickets. But all expectations are sums of tickets, and the assumption of fairness implies that all ways of compounding an expectation from tickets yield the same total price. So every expectation has a fair price. As it turns out, we can deduce this from Axioms T1-T6 alone, without appealing to the background knowledge about fairness that justifies these axioms. More precisely, we can deduce from these axioms the existence of prices $E_S(X)$ for all situations S and all expectations X such that $E_S(X) = V_S(X)$ when X is a ticket. We can deduce that these prices add:

$$\text{If } Y = \sum_{X \in \Phi} X \text{, then } E_S(Y) = \sum_{X \in \Phi} E_S(X) .$$

We can also deduce that

$$\min_{i \varepsilon S} X(i) \leq E_S(X) \leq \max_{i \varepsilon S} X(i), \qquad (1)$$

and more generally that if Π is a partition of S into situations, then

$$\min_{T \varepsilon \Pi} E_T(X) \leq E_S(X) \leq \max_{T \varepsilon \Pi} E_T(X). \qquad (2)$$

Formula (1) says that you cannot make money for sure by buying X in S and collecting on it when you get to a stop sign (or by selling X in S and paying it off when you get to a stop sign), and formula (2) says that you cannot make money for sure by buying X in S and selling it when you get to a situation in Π (or by selling X in S and buying it back when you get to a situation in Π).

We can also deduce that strategies are to no avail. More precisely, we can deduce that if the strategy s is permissible in S for a spectator with capital \$r in S, then $E_S(X_{r,s}) = r$. Thus a strategy accomplishes nothing that we cannot accomplish directly by paying the fair price for an expectation.

3.3 Axioms for probability

We have completed our work inside the circle labeled “Fair odds” in Figure 2. Now we move along the arrow from fair odds to probability by using the fair odds on an event as a measure of warranted belief in the event.

Actually, we will not exactly use the odds p:(1-p) on A as the measure of our belief in A. Since we are accustomed to a scale from zero to one for belief, we will use instead the price p. We will write

$$P_S(A) = V_S\langle \$1,A\rangle, \tag{3}$$

and we call $P_S(A)$ the probability of A in S.

The following axioms for probabilities follow from Axioms T1-T6 for fair prices.

Axiom P1. $0 \leq P_S(A) \leq 1$.

Axiom P2. $P_S(A) = 0$ if and only if A is impossible in S.

Axiom P3. $P_S(A) = 1$ if and only if A is certain in S.

Axiom P4. If A and B are incompatible in S, then

$$P_S(A \approx B) = P_S(A) + P_S(B).$$

Axiom P5. If T follows S, and U follows T, then

$$P_S(U) = P_S(T) \cdot P_T(U).$$

Axioms P1-P5 are essentially equivalent to Axioms T1-T6. If we start with Axioms P1-P5 and define ticket prices by

$$V_S\langle \$r,A\rangle = r \cdot P_S(A),$$

then we can derive Axioms T1-T6. It then turns out that

$$E_S(X) = \sum_{i \in S} X(i) \cdot P_S(\{i\} \tag{4}$$

for every expectation X.

The fact that we can begin with Axioms P1-P5 does not, of course, make probability autonomous of the other ideas in the circle of reasoning. Like each of the other ideas, probability is caught in the circle of reasoning. It can serve as a formal starting point, but when it does, it uses axioms whose motivation derives from the other starting points. The only apparent justification for Axioms P1-P5 lies in the long-run and short-run fairness of odds that we used to justify Axioms T1-T6.

Axioms P1-P5 are quite similar to Kolmogorov's axioms. We will return to this point in Section 3.5. First, let us travel one more step in our circle, from probability to knowledge of the long run.

3.4 Implications for the short and long runs

The axioms we have just formulated capture the essential properties of fair price and probability in the ideal picture, and from them we can derive the spectators' knowledge of the long run. The details cannot be crowded into this chapter, but we can state the most basic results.

One aspect of the spectators' knowledge of the long run is their knowledge that no strategy, short-run or long-run, can assure a net gain. Since, as we have already seen, a strategy always boils down to buying an expectation, it suffices to show that buying an expectation cannot assure a net gain. And this is easy. It follows from (3) that for any expectation X and any situation S,

$$\text{if } P_S\{X > E_S(X)\} > 0, \text{ then } P_S\{X < E_S(X)\} > 0.$$

If X can pay more than its price, then it can also pay less.

Another aspect of the spectators' knowledge of the long run is that no strategy can give a reasonable hope of parlaying a small stake into a large fortune. Since following a strategy in S boils down to using one's entire capital in S to buy a non-negative expectation X, it suffices to show that the probability of a non-negative expectation paying many times its price is very small. And this again is easy. It is easy to show that

$$P_S\{X \geq k \cdot E_S(X)\} \leq \frac{1}{k},$$

when X is non-negative. The odds against a strategy for multiplying one's capital by k are at least k to one.

Finally, consider the frequency aspect of the spectators' knowledge of the long run. In the case of the fair coin, the spectators know that the proportion of heads is one-half in the long run. They know something similar in the general case. In order to derive this knowledge from our axioms, we need to formulate the idea of a spectator's successive net gains from holding an expectation.

Let Ω denote the initial situation in a situation tree, and suppose that a spectator acquires an expectation X in Ω. She holds this expectation until she comes to a stop sign, but every time she moves down from one situation to the next, she takes note of X's change in value. She calls this change her net gain. Her first net gain is

$$G_1 = E_{S_1}(X) - E_\Omega(X),$$

where S_1 is the situation at which she arrives immediately after Ω. Her second net gain is

$$G_2 = E_{S_2}(X) - E_{S_1}(X),$$

where S_2 is the situation at which she arrives immediately after S_1. And so on. The net gains $G_1, G_2, \ldots$ depend on the path she takes down the tree (because $S_1, S_2, \ldots$ depend on the path she takes down the tree). In other words, they are expectations. And we can prove the following theorem about them.

> *Theorem.* Suppose the net gains G_j are uniformly bounded. In other words, there exists a constant κ such that $\left|G_j(i)\right| \leq \kappa$ for every j and every stop sign i. And suppose ε and δ are positive numbers. Then there exists an integer N such that

$$P_\Omega\left(\left|\frac{\sum_{j=1}^{n} G_j}{n}\right| \leq \varepsilon\right) \geq 1-\delta$$

whenever $n \geq N$.

In other words, the average net gain in n trials is almost certainly (with probability 1-δ) approximately (within ε of) zero. This theorem is one version of the law of large numbers, first proven by James Bernoulli. For a proof of this version, see Shafer (1985).

To see what this theorem means in the case of the fair coin, we can suppose the spectator chooses a number n and bets $1 on heads for each of the first n trials. Altogether she must pay $n, and she will get back 2Y, where Y is the total number of heads in the first n trials. So her net expectation is

$$X = 2Y - n.$$

We have $E_\Omega(Y) = \frac{n}{2}$ and $E_\Omega(X) = 0$. The jth net gain from X, G_j, is $1 if the jth trial comes up heads and -$1 if it comes up tails. And

$$X = \sum_{j=1}^{n} G_j \ .$$

Hence

$$\left|\frac{\sum_{j=1}^{n} G_j}{n}\right| \leq \varepsilon$$

is equivalent to

$$\left|\frac{X}{n}\right| \leq \varepsilon$$

or

$$\left|\frac{Y}{n} - \frac{1}{2}\right| \leq \frac{\varepsilon}{2}$$

So the theorem says that $\frac{Y}{n}$, the frequency of heads, is almost certainly close to $\frac{1}{2}$.

The frequency aspect of the long run in a general situation tree is only a little more complicated. To derive it from the theorem, we assume that the spectator bets $1 in each situation on the outcome of the experiment to be performed in that situation. If we also assume that the probabilities of the events on which she bets never fall below a certain minimum, so that the possible gains for $1 bets are bounded, then the theorem applies, and it says that the frequency with which the spectator wins is almost certainly close to the average of the probabilities for the events on which she bets.

3.5 The role of Kolmogorov's axioms

Kolmogorov's axioms are similar to Axioms P1-P5, but simpler. The simplicity is appropriate, because these axioms serve as a mathematical rather than as a conceptual foundation for probability.

Kolmogorov begins not with a situation tree, but simply with a set Ω of possible outcomes of an experiment. Events are subsets of Ω. We may assume, in order to make Kolmogorov's axioms look as much as possible like Axioms P1-P5, that Ω is finite. In this case, Kolmogorov assumes that every event A has a probability P(A), and his axioms can be formulated as follows:

Axiom K1. $0 \leq P(A) \leq 1$.

Axiom K2. P(A) = 0 if A is impossible.

Axiom K3. P(A) = 1 if A is certain.

Axiom K4. If A and B are incompatible, then P(A≈B) = P(A)+P(B).

Here "A is impossible" means that A = ∅, "A is certain" means that A = Ω, and "A and B are incompatible" means that $A \cap B \neq \emptyset$. In addition to the axioms, we have several definitions. We call

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{5}$$

the *conditional probability* of A given B, and we say that A and B are *independent* if P(A|B) = P(A). We call a real-valued function X on Ω a *random variable*, we set

$$E(X) = \sum_{i \in \Omega} X(i) \cdot P(\{i\}) , \tag{6}$$

and we call E(X) the *expected value* of X.

Axioms K1-K4 are basically the same as Axioms P1-P4. Definition (5) corresponds to Axiom P5, and definition (6) is similar to (4). But the comparison brings out the sense in which Kolmogorov's axioms do not provide a conceptual foundation for probability. Kolmogorov himself was a frequentist, and yet the axioms do not involve any structure of repetition. This is something that must be added, through the construction of product probability spaces.

Kolmogorov's axioms are justly celebrated in their role as a mathematical foundation for probability. They are useful even in understanding situation trees, for probability spaces, sets with probability measures in Kolmogorov's sense, are needed to provide probabilities for the individual experiments in a situation tree. We should not try, however, to use these axioms as a guide to the meaning of probability. Doing so only produces conundrums. It makes us puzzle over the “probability of a unique event.” It makes independence seem like a mysterious extra ingredient added to the basic idea of probability. It makes conditional probability equally mysterious, by making it seem completely general—a definition that applies to any two events. Independence and conditional probability have a role in the ideal picture of probability, and this role can give us guidance about their use, but all such guidance lies outside Kolmogorov's axiomatic framework.

The framework provided by situation trees and Axioms P1-P5 does not create these mysteries and confusions. This framework makes it clear that events do not have probabilities until they are placed in some structure of repetition. Independence has a role in this structure; events involved in successive trials are independent if the experiment performed in each situation does not depend on earlier outcomes. But we can relax this condition, for successive net gains are uncorrelated even when the experiment performed in each situation does depend on earlier outcomes. And we do not talk arbitrarily about the conditional probability of one event given another; we talk instead about the probability of an event in a situation.

4. Historical perspective

This section relates the ideal picture described in Sections 2 and 3 to the historical development of probability theory.

Remarkably, the early development of mathematical probability in the seventeenth and eighteenth centuries followed a path similar to the path we have followed through Figure 2, except that it began with fair price as a self-evident idea, not one that had to be justified by an appeal to knowledge of the long run. By the end of the eighteenth century, the different elements of the ideal picture were relatively unified, but this unity was not well articulated, and it was broken up by the empiricism of the nineteenth century. This break-up persists in today's stand-off between frequentists and subjectivists. I argue, however, that the competing philosophical foundations for probability that these two groups have advanced are fully coherent only when they are reunified within the ideal picture.

4.1 The original development of the ideal picture

We can use Figure 2, starting with fair odds, as an outline of the development of probability in the seventeenth and eighteenth centuries. The theory of fair price in games of chance was first developed by Pascal, Fermat, and Huygens in the 1650s. The step from fair price to probability was taken during the next fifty years, most decisively by James Bernoulli. Bernoulli also was the first to use ideas of probability to prove the law of large numbers, the central feature of our knowledge of the long run. The final step, from knowledge of the long run back to fair price, was apparently first taken only by Condorcet in the 1780s. I will only sketch these developments here. For more information, see Hacking (1975) and Hald (1990).

The origins of probability theory are usually traced to the theory of fair price developed in the correspondence between Pierre Fermat and Blaise Pascal in 1654 and publicized in a tract published by Christian Huygens in 1657. The word probability did not appear in this work. It is an ancient word, with equivalents in all the languages these scholars spoke. A probability is an opinion, possibility, or option for which there is good proof, reason, evidence, or authority. But these authors were not talking about probability. They were talking about fair price. Essentially, they reasoned along the lines that we have retraced in Section 3.2 to deduce fair prices for some expectations from fair prices for others. They did not, however, use knowledge of the long run frequency to justify the existence of fair price. For them, it was self-evident that an expectation should have a fair price.

In fact, there was remarkably little talk about the long run in the seventeenth-century work. Many of the authors played and observed the games they studied, and we must assume that they all had the practical gambler's sense that fair bets would allow one to break even over the long run. But the connection between fairness and the long run was not used in the theory. On the other hand, the theory did involve repetition. There was always some sequence of play in prospect, and this ordering of events was used to relate fair prices to each other, just as we used it in Section 3.2. Situation trees like Figures 3 and 4 were implicit in the thinking of Pascal, and they were drawn explicitly by Huygens (see Edwards 1987, p. 146).

From the very beginning of the theory of games of chance, people did want to use the theory in other domains. Pascal himself may have been the first to do so, in his famous argument for betting on the existence of God. Even when writing on this topic, Pascal did not use the word probability, but his friends Antoine Arnauld and Pierre Nicole used it in 1662 in their Port Royal *Logic*. In one justly famous passage, they explained how people who are overly afraid of thunder should apportion their fear to the probability of the danger. From this it is only a short step to probability as a number between zero and one, and a number of people took this step. In 1665, at the age of 19, Leibniz proposed using numbers to represent degrees of probability (Hacking 1975, p. 85). The English cleric George Hooper, writing in 1689, used such numbers without hesitation, sometimes calling them probabilities and sometimes calling them credibilities (Shafer 1986).

It was James[1] Bernoulli who made probability an integral part of the mathematical theory. In his book *Ars Conjectandi*, which was published in 1713, five years after his death, Bernoulli defined probability as degree of certainty. The theory of fair price could be applied to probability, Bernoulli explained, because conjecturing is like throwing dice, except that the stakes are certainty. Just as the rounds you win and lose in a game entitle you to a definite portion of the stakes, the arguments you find for and against an opinion entitle you to a definite portion of certainty. This portion is the opinion's probability.

By relating probability to the theory of fair price, Bernoulli set the stage for deriving properties of probability from properties of fair price. Since fair price is additive, probability must also be additive. Since fair prices change as events unfold, probabilities also change as events unfold. This is the substance of the arrow from fair price to probability in Figure 2. Bernoulli did not work out these new properties of mathematical probability,[2] but this was quickly done by Abraham de Moivre. In *The Doctrine of Chances,* which first appeared in 1718, de Moivre established many of the ideas used in probability theory today. He talked about probabilities of events (rather than about probabilities of things, as Bernoulli had), he formulated the idea that the probability of one event may change when another event happens, and he formulated versions of the rules of additivity and compound probability, our Axioms P4 and P5. The arguments that he gave for these rules in the third edition of his book (pp. 5-9) were essentially the same as the argument we used in Section 3 to derive Axioms T5 and T6. Similar arguments were given by Bayes (Shafer 1982, 1985).

Bernoulli's second great contribution was his law of large numbers, a version of which we proved in Section 3.4. Bernoulli saw this theorem as a way to justify the use of observed frequencies as probabilities. This steps out of the ideal picture that we have been studying, for within that ideal picture, the spectators know the probabilities, and hence do not need to use frequencies to estimate them.

Within the ideal picture, the law of large numbers is part of the knowledge of the long run that can be used to justify the existence and derive the properties of fair prices. The fair price is the price that will break even in the long run. This use of the law of large numbers within the ideal picture did emerge in the

---

1 His English contemporaries referred to Bernoulli as James, but it is now more common to use the German Jakob, because he grew up in German-speaking Basel. But he usually wrote in Latin, where his name is Jacob, or French, where his name is Jacques. We still call Bernoulli's country Switzerland in English, instead of choosing among Schweiz, Suisse, Svizzera, and Helvetia. In the same spirit, I call him James.

2 It is not even quite fair to say that these properties are consequences of Bernoulli's work, for he made a looser connection between probability and fair price, one that permitted non-additive probabilities such as those that appear in the theory of belief functions (Shafer 1978, 1986).

eighteenth century, but only at the end. Apparently it was first formulated by Condorcet, in the 1780s (Todhunter 1865, pp. 392-393).

4.2 The disintegration of the ideal picture in the nineteenth century

The eighteenth-century elements of the ideal picture that we have just reviewed were synthesized at the beginning of the nineteenth century in the work of the famous French mathematician Laplace, but this synthesis broke up in the course of the nineteenth century. The break-up can be attributed to the applied ambitions of the theory, together with the empiricism of the philosophy of science of the times. The mathematicians who studied the theory wanted to use it very widely, not just in games of chance, but the empiricism of the times demanded that the terms of a scientific theory have direct empirical reference. It is easy to imagine such a reference for frequency, but not for fair price or warranted belief.

Laplace worked on probability from the early 1770s until 1820, when the third edition of his famous treatise, *Théorie analytique des probabilités,* was published. His most important contributions to probability were mathematical advances, such as the central limit theorem, that facilitated the use of probability in the analysis of data. These contributions can be regarded as the beginning of mathematical statistics (Stigler 1986). We are more concerned here, however, with Laplace's approach to the foundations of the subject. Here he emphasized probability itself, as if there were always warranted numerical degrees of belief that had the properties that De Moivre and Bayes had derived from the properties of fair price. Traces of those derivations remain in Laplace's work, but on the whole, he wrote as if probability were a self-sufficient starting point. One aspect of this de-emphasis of fair price was that the ordering of events, which was so prominent from Pascal to Bayes, was underplayed. Thus for nineteenth century readers, who took Laplace as their authority, this ordering was not an important feature of the ideal picture.

Though he took warranted belief as basic, Laplace integrated it thoroughly with frequency. He had no qualms about Bernoulli's law of large numbers, which purported to prove that the frequency of an event will equal its probability. But as Porter (1986) and Hacking (1990) explain, many later nineteenth-century writers found the direction of this reasoning troublesome. Most were willing to accept it in games of chance, where rational beliefs are justified by the same symmetries that imply equal frequencies. But in other domains, where the probabilists now wanted to ply their craft, frequency itself seemed to be the only empirical basis for probability. Beginning in the 1840s, philosophers and philosophically minded mathematicians of an empiricist bent, especially Cournot, Ellis, Fries, Mill and Venn, advanced the view that probability should be defined as frequency. For many of them, proving that frequency will equal probability was unnecessary and even silly.

The thesis of this chapter is that a synthesis of frequency and rational belief is once again possible and desirable. This is because we can afford to go back to the idea that the theory of probability only applies, in the first instance, to settings such as games of chance. We can afford to do so because our empiricism is more flexible than the empiricism of the nineteenth century in the way it relates theory and application. We can now take the view that applying the theory of probability means relating the ideal picture of probability to reality, and as we will see in Section 5, there are many ways that this can be done.

4.3 The foundations of frequentism

Most probabilists resisted frequentism during the nineteenth century, if only because it seemed to take aim at the most interesting mathematics in their theory. By the end of the century, as positivism became dominant throughout science, probabilists had accepted the idea that probability would ultimately find its foundation in frequency, but they still struggled to reconcile this with the structure of the mathematical theory.

A solution of sorts was achieved in the early twentieth century by Kolmogorov's axioms. This solution simply put a wall between the mathematical theory itself, which was to be treated axiomatically, and the interpretation and application of the theory. One could equate probability with frequency by definition in applications, while still deriving frequency from probability within the theory.

Few have been satisfied with this, however. There has been a continuing quest for a deeper frequentist foundation for probability. The most important milestones in this quest have been the work of von Mises on random sequences, the critique of his work by Ville and Wald, and the work by Kolmogorov on algorithmic complexity. (For an overview, see Martin-Löf 1969. For related recent work, see Uspenskii, Semenov, and Shen' 1990.)

Von Mises, who began writing on probability in the 1920s, hoped to buttress the frequency interpretation by establishing the existence of infinite sequences of heads and tails, say, in which exactly half the entries are heads in the limit, both in the sequence as a whole and in subsequences. He proposed deriving the whole theory of probability from the properties of these "random" sequences.

Ville demonstrated that frequency is not a sufficient foundation for probability, even in von Mises's framework. The existence of limiting frequencies is not enough to rule out successful gambling schemes. We can construct an infinite sequence of heads and tails in which half the entries are heads in the limit (the limiting frequency of heads is one-half both in the whole sequence and in subsequences selected on the basis of preceding outcomes) and yet in which an observer can make money by betting on heads, because the number of heads is always slightly greater than the number of tails in any finite initial portion of the sequence.

Wald proved the existence of infinite sequences that (1) have a limiting frequency of heads equal to one-half for the whole sequence and for many subsequences, and (2) rule out many gambling schemes of the type suggested by Ville. In fact, given a countable number of subsequences and a countable number of other gambling schemes, there exists a sequence that cannot be beaten by a bettor that uses any of these subsequences or gambling schemes.

In the 1960s, Kolmogorov and others advanced a definition of randomness that applies to finite rather than infinite sequences. According to this definition, a sequence is random to the extent that it is complex, where complexity is measured by the length of the shortest computer program that will generate the sequence. Kolmogorov presented his complexity definition, just as he had presented his axioms many years earlier, as a foundation for an objective, frequency interpretation of probability. As he and others have now shown, the complexity definition does indeed imply the knowledge of the long run that is claimed in the ideal picture. It implies both the stability of frequency and the futility of gambling schemes.

Did Wald and Kolmogorov succeed in providing foundations for a purely objective conception of probability? The claim to pure objectivity is shaky, for there are obvious subjective elements in their results. In the case of Wald, the subjectivity lies in the choice of the countable number of properties that are demanded of the sequence. A countably infinite set of properties is surely all a person would want, but they do not make a sequence random from the perspective of someone else who chooses a property not in the set. In the case of Kolmogorov, the subjectivity lies in the choice of the computer. Some sequences can be generated by a short program on one computer but only by a very long program on another.

The viewpoint of this chapter suggests that rather than minimize these obvious subjective elements, we should acknowledge them. They too represent aspects of the ideal picture. They represent the relation between knowledge and fact in the ideal picture. The countable set of properties (for Wald) or the computer (for Kolmogorov) represent the spectators. The fact that someone else may know more than these spectators is beside the point. The point is that these spectators' knowledge is limited to knowledge of long-run frequencies and knowledge of their own inability to devise successful betting schemes. Randomness is a property of the relation between fact and observer.

4.4 The foundations of subjectivism

Belief plays such an important role in the ideal picture that frequentism has never been universally persuasive, even among those who share the frequentists' empiricism. The most vigorous opponents of the frequentists in the twentieth century have been the subjectivists, such as Frank P. Ramsey, Bruno de Finetti, and Leonard J. Savage, who have sought an empirical foundation for probability in the ideas of personal (rather than fair) betting rates and personal (rather than warranted) degrees of belief. They have argued that a person's personal belief in an event can be measured by the amount that the person is willing to pay for a $1 ticket on the event.

From the viewpoint of this chapter, this represents an attempt to simplify the ideal picture. The simplification involves dropping the idea of fairness as well as the ordering of events that provides the link with frequency. Is this simplification successful? Can we establish the properties of probability—in the form of Kolmogorov's axioms and definitions, say—from the idea of personal betting rates alone? The subjectivists have argued that we can, but when we compare their arguments with the arguments in Section 3, the holes are apparent.

First, consider the rule of additivity, represented in Section 3 by Axioms T5, P4, and K5. Since they did not want to appeal to fairness, Ramsey, de Finetti, and Savage had to use a very implausible assumption to derive this rule. They had to assume that the greatest price a person is willing to pay for a $1 ticket on a given event is the same as the least price at which she is willing to sell such a ticket. This symmetry is inherent in the notion of fair price; a price cannot be fair unless it is fair to both the buyer and the seller. But it is not inherent in the idea of personal price. It is perfectly rational for a person to refuse both sides of some bets.

The possibility of one-sided betting rates and hence one-sided numerical degrees of belief opens a space for alternative theories of subjective probability, in which degrees of belief do not satisfy the usual axioms of probability. One such theory is the theory of belief functions (Shafer 1990b). The followers of Ramsey, de Finetti, and Savage, have not neglected to denounce this and other one-sided theories as irrational, but there is no argument behind these denunciations.

Second, consider the rule for changes in probability, represented in Section 3 by Axiom T6, Axiom P5, and formula (5). Since they do not assume any ordering of events, the subjectivists do not explain this rule in terms of personal betting rates in successive situations. Instead, they talk about called-off bets; the conditional probability $P(A|B)$, according to de Finetti, is a personal betting rate for a bet on A that will be called off if B does not happen. This makes it possible to derive (5), but it leaves unanswered the question of what rates for called-off bets have to do with changes in probability.

Here, as in the case of frequentism, the foundation advanced by the subjectivists is basically sound. But to make complete sense of it, we need to put it back into the ideal picture.

5 The diversity of application

In the preceding pages, I have repeatedly asserted that applying probability theory to a problem involves relating the ideal picture to that problem. This concluding section briefly reviews some of the ways this can be done.

5.1 Probability models

In the frequentist view, the straightforward way to use probability is to make a probability model of a real repeatable experiment, a model that gives probabilities for the outcomes of that experiment. The model can then be tested by its fit with data from repetitions of the experiment, perhaps after using this data to estimate some of the probabilities in the model.

Such probability modeling is one instance of the thesis that applying probability to a problem means relating the ideal picture to the problem. We are using the ideal picture itself as a model. The question that must be debated is the extent to which the relation between knowledge and fact that is central to the ideal picture carries over to the reality being modeled. Do we really have a probability model, or merely a frequency model? The answer varies. In practice, we never have as much effective knowledge (knowledge of as many probabilities, for example) as the spectators in the ideal picture have, and in some cases we have relatively little. Often the statements we make in probability modeling relate to the knowledge of some idealized observer, not to our own knowledge. It is the negative aspects of the spectators' knowledge in the ideal picture—their inability to take advantage of bets at odds given by the frequencies—that seems to carry over most often to the practical problem.

It should be remembered that even within the ideal picture, not all probabilities are interpretable as frequencies. In a situation tree in which the same experiment is always repeated, we can interpret the probability for an outcome of that experiment as a frequency, and in general situation trees, we can interpret certain averages of probabilities as frequencies. But probabilities for events involving many trials, though they derive their status as fair prices from probabilities interpretable as frequencies, may not themselves be interpretable as frequencies. When we elaborate probability models mathematically by taking limits in the large (infinite numbers of trials) or in the small (continuous models), we tend to create probabilities that do not correspond to frequencies in the reality being modeled even if they do seem interpretable as frequencies in the ideal picture. Matheron (1989) points out that this gap between model and reality grows even wider when we insist on interpreting time-series and geostatistical models in terms of imaginary independent repetitions (see also Stein 1990). By emphasizing that our model is the ideal picture rather the numerical probabilities, we can avoid this insistence.

5.2 Probability as a standard of comparison

We often use the ideal picture of probability as a standard of comparison. This is explicit in some cases, as when we test the performance of experts by comparing their success with random choice, or when we evaluate or compare judges who make probabilistic predictions (Bloch 1990). It is less explicit in other cases, as when we use standard statistical tests to assess whether additional independent variables should be included in least-squares fits to non-observational data.

Several authors, especially Beaton (1981) and Freedman and Lane (1983), have advanced "non-stochastic" interpretations for the standard F-tests in the case of non-observational data. Their arguments involve the deliberate creation, through permutation of residuals, of populations of data to which the actual data can be compared. Without discussing these arguments in detail, I would like to suggest that their rhetoric can be simplified and strengthened if we think in terms of a comparison between the predictive accomplishment in the data of the variables being tested and the predictive accomplishment in the ideal picture of variables that are irrelevant.

5.3 Randomization

We are all accustomed to computer-generated random numbers. Since the programs that generate such numbers are as deterministic as any other computer programs, the numbers are often characterized as only "pseudo-random." From the viewpoint of this chapter, however, this derogation is inappropriate. As we saw in Section 4.3, randomness is not a property of a sequence of numbers in itself; it is a property of the relation between these numbers and an observer. So the question is not whether a given sequence of numbers is truly random or not; it cannot be random in and of itself. The question is what observer we are talking about. A sequence of numbers generated by a certain program is not random relative to an observer who is able to use the program to reproduce them. It may be more or less random relative to an observer to whom we deny (or who denies herself) this ability.

From the viewpoint of this chapter, we are deliberately creating an instance of the ideal picture when we generate a sequence of random numbers. When we use the sequence to choose a sample from a population or to assign treatments in an experiment, we are deliberately entangling the ideal picture with data bearing on a practical question, so that probabilities in the ideal picture come to bear, indirectly, on that question.

We cannot enter here into the debates about the validity of arguments contrived in this way. I would like to suggest, however, that by explicitly bringing the ideal picture into the story, we can make these arguments more persuasive. The ideal picture itself requires a certain form of ignorance, and creating it typically requires some enforcement of that ignorance (not using the program to regenerate the random numbers, not looking at them, etc.), and hence it is no additional paradox that entangling this ideal picture with a practical problem should involve ignoring certain information (which labels were sampled or assigned to which treatments).

5.4 Argument by analogy to the ideal picture

The ideal picture, as described in this chapter, always involves some element of repetition, even if this repetition is not exact. There is some sequence of events that allows us to bring a frequency or long-run element into the discussion. How, then, can we apply probability to a problem for which we do not have a sequence of similar problems?

One answer to this question is that using probability requires creating some such sequence—some reference class. This is a fair answer, and it has the virtue of making the deliberate nature of subjective probability judgment clear. When the reference class is almost completely imaginary, however, it may be more instructive to say that we are drawing an analogy between the question that interests us and a question in the ideal picture. We are saying that our knowledge (and ignorance) about this question is similar to our knowledge about a certain question in a certain version of the ideal picture. Our evidence about the question is similar in strength, and perhaps in structure, to knowing the probabilities in that version of the ideal picture.

Shafer and Tversky (1985) discuss how both Bayesian and belief-function probability arguments can be seen in this way.

5.5 Conclusion

To use probability; we must relate the ideal picture to a problem. It is obvious from the examples that we have just discussed that the success we have in doing this in any particular instance will always be debatable. The ideal picture may or may not be good enough a model; it may or may not be relevant as a standard of comparison; it may or may not provide a convincing analogy. This is life. But the debate in each particular case need not be an empty debate. We can formulate criteria for judging the excellence of each of these kinds of probability argument.

As Meier, Sacks, and Zabell (1984, pp. 161-164) have pointed out, the real debate in applied statistics is not between the formulas of the frequentists and the formulas of the Bayesians. The real debate is between "strict constructionists," who would limit the use of probability to those situations where frequentist assumptions are fully satisfied, and "Benthamites," who find the mathematical precision of probability useful no matter how little evidence they have at hand. The framework of this chapter is designed to focus this debate on examples and make it more productive. It provides a common language in which to criticize and praise both Bayesian and frequentist analyses. To use this language, the frequentist must go beyond saying that assumptions are or are not satisfied; she must draw an analogy between her relation with her data the spectator's relation with the outcomes in the ideal picture. The Bayesian must go beyond saying that certain numbers represent her beliefs; she must defend the analogy by which these numbers are produced. This puts both the frequentist and the Bayesian in the position of discussing the quality of their analyses, not the ideology that underlies them.

The philosophy of probability advanced in this chapter unifies the frequentist and subjectivist approaches at a level deeper than the level of axioms. It allows us to bring together in one framework the unified eighteenth-century understanding of probability, the frequentist foundations of von Mises and Kolmogorov, and the subjectivist foundations of de Finetti. It also allows us to spell out explicitly the different ways we construct probability arguments. It merits the name given it in Figure 8: the constructive philosophy of probability.

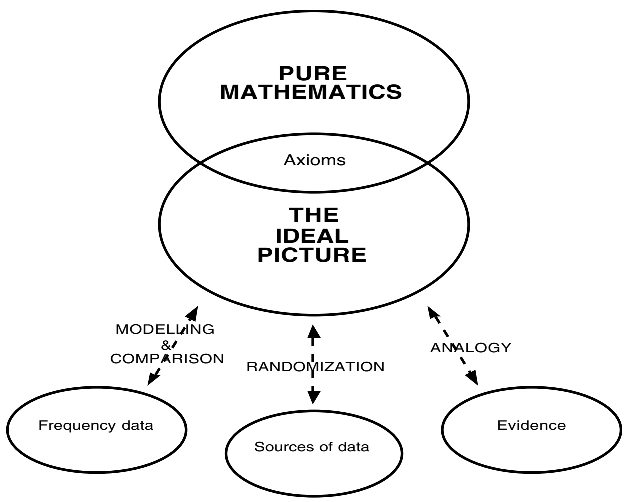

Figure 8. The constructive philosophy of probability

References

Arnauld, Antoine, and Pierre Nicole (1662). *l'Art de penser*. Paris. Widely used as a textbook, with many editions and translations, this book is often referred to as the Port Royal *Logic*.

Beaton, Albert E. (1981). Interpreting least squares without sampling assumptions. Research Report. Educational Testing Service. Princeton, New Jersey.

Bloch, Daniel A. (1990). Evaluating predictions of events with binary outcomes: An appraisal of the Brier score and some of its close relatives. Technical Report No. 135. Division of Biostatistics. Stanford University. Stanford, California.

Daston, Lorraine (1988). *Classical Probability in the Enlightenment.* Princeton University Press, Princeton, New Jersey.

Edwards, A.W.F. (1987). *Pascal's Arithmetic Triangle.* Oxford University Press, New York.
Freedman, David, and David Lane (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics* **1** 292-298.
Hacking, Ian (1975). *The Emergence of Probability.* Cambridge University Press, Cambridge, England.
Hacking, Ian (1990). *The Taming of Chance*. Cambridge University Press, Cambridge, England.
Hald, Anders (1990). *A History of Probability and Statistics and their Applications before 1750.* Wiley, New York.
Kolmogorov, Andrei (1950). *Foundations of the Theory of Probability.* Translated from the Russian by Nathan Morrison. Chelsea, New York.
Martin-Löf, Per (1969). The literature on von Mises' Kollektivs revisited. *Theoria* **35:1** 12-37.
Matheron, Georges (1989). *Estimating and Choosing, An Essay on Probability in Practice*. Translated from the French by A.M. Hasofer. Springer-Verlag, Berlin.
Meier, Paul, Jerome Sacks, and Sandy L. Zabell (1984). What happened in *Hazelwood:* statistics, employment discrimination, and the 80% rule. *American Bar Foundation Research Journal*. Volume 1984, pp. 139-186.
Porter, Theodore M. (1986). *The Rise of Statistical Thinking, 1820-1900*. Princeton University Press. Princeton, New Jersey.
Shafer, Glenn (1978). Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences* **19** 309-370.
Shafer, Glenn (1982). Bayes's two arguments for the rule of conditioning. *Annals of Statistics* **10** 1075-1089.
Shafer, Glenn (1985). Conditional probability (with discussion). *International Statistical Review* **53** 261-277.
Shafer, Glenn (1986). The combination of evidence. *International Journal of Intelligent Systems* **1** 127-135.
Shafer, Glenn (1990a). The unity of probability. Pp. 95-126 of *Acting Under Uncertainty: Multidisciplinary Conceptions*, edited by George von Furstenberg, Kluwer.
Shafer, Glenn (1990b). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning* **4** 323-362.
Shafer, Glenn, and Amos Tversky (1985). Languages and designs for probability judgement. *Cognitive Science* **9** 309-339.
Stein, Michael (1990). Review of Matheron (1989). *Technometrics* **32** 358-359.
Stigler, Stephen M. (1986). The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press. Cambridge, Massachusetts.
Todhunter, Isaac (1865). *A History of the Mathematical Theory of Probability*. Macmillan, London.
Uspenskii, V.A., A.L. Semenov, and A.Kh. Shen' (1990). Can an individual sequence of zeros and ones be random? *Russian Mathematical Surveys* **45** 121-189.