CHAPTER 8

BABEL

Mathematical probability as we know it arose somewhat fortuitously. It gradually supplanted a much older and more ambiguous sense of probability that had prevailed for many centuries. This new mathematical probability grew directly out of the mathematics of games of chance, and was thus not possible until, in the mid-1650s, a critical mass of mathematical talent happened to be focused, albeit momentarily, on such games. It was then not until the 1680s that Jacob Bernoulli first conceived of probability as being calibrated on a scale between zero and one, and 1718 before another lone genius, Abraham de Moivre, defined probability as *a fraction of chances.*

This classical definition of probability was still not firmly established until Pierre-Simon Laplace promoted its broader appreciation and application in the late 1700s. By around 1800, this version of probability had become the "obvious" way to think about uncertainty. When an outcome (or conclusion) was uncertain, it was *as if* the result was chosen by chance, as by a lottery or similar random mechanism. That is, we could regard the outcome as being determined by a *metaphorical lottery*. The laws of nature would of course determine the outcome, but our limited knowledge about

Willful Ignorance: The Mismeasure of Uncertainty, First Edition. Herbert I. Weisberg.

the underlying causal processes would place us in the position of a spectator who observes a game of chance. So, Laplace could assert famously that probability is the measure of our ignorance. By this, he meant that probability is the means of quantifying our partial knowledge based on our degree of ignorance about the true causes.

This metaphorical lottery concept eventually became the basis for what is now called *classical probability*. Gradually, however, the explicit analogy to games of chance became superfluous, retained only as a ritualistic observance. Every textbook on probability theory written prior to the twentieth century began with the formal definition of probability as a fraction of chances. However, the subsequent mathematical development would dispense entirely with any reference to the underlying chances. The mathematical laws of probability became perceived as self-evident and were justified by an appeal to common sense. Based on these seemingly universal laws, Laplace and his followers erected an elaborate intellectual edifice with very broad applicability. The theory of probability was rightly regarded as a major achievement of the Age of Reason.

What happened next, however, was unexpected. Shortly after Laplace's death in 1827, the structure of classical probability came tumbling down, much like the Tower of Babel and for a similar reason. The proponents of classical probability were guilty of overreaching, and their hubris was punished by a strong backlash. Over time, probability as a universal language of uncertainty became fractured into many differing interpretations of what probability "really" meant:

> It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel. There must be dozens of different interpretations of probability defended by living authorities, and some authorities hold that several different interpretations may be useful … Considering the confusion about the foundations of statistics, it is surprising, and certainly gratifying, to find that almost everyone is agreed on what the purely mathematical properties of probability are.[1]

This stunning reversal would have astonished Jacob Bernoulli. Probability as a fraction of chances had become *obvious*, while the odds, chances, and expectations from which probability was originally derived had lost much of their intuitive appeal. As a result, no one was quite sure what this word "probability" really meant.

Laplace's classical probability seems to us quaint and outmoded. What are these possible "chances" or equally possible cases, and how do we know they are really equally possible? Is it not circular to define probability as a fraction of chances? But such questions are naïvely predicated on a literal interpretation of the lottery metaphor. Even sophisticated philosophers have fallen prey to this common misconception[2]:

> The fact, Φ, that the rear tires are completely bald, confers 60% probability on having a flat before reaching Massachusetts. It is a mere myth that Φ should break down into a set of cases Φ_1, which entail getting a flat tire and another set Φ_2, entailing the opposite. We certainly know of no such decomposition and have no good reason to think there is one. Yet the myth that every problem in probability can be reduced to a set of favourable and unfavourable cases persisted for centuries.

Hacking here seems to be saying that classical probability referred to a set of actual chances. For instance, there might be 60 future possible eventualities that would entail a flat tire and 40 that would not.

Was this a myth, or a useful metaphor? I am quite confident that Bernoulli, de Moivre, and Laplace did not interpret the concept of equally likely cases literally, except possibly for games of chance. Framing our probabilities as if they are similar to fractions of chances in a metaphorical lottery does not imply we believe there are actual chances. Rather, we have a degree of uncertainty that can be described by such a hypothetical fraction. Recall that Bernoulli had explained that virtually *any* event "can be conceived as divisible with regard to the number of cases in which it can be gained or lost, happen or fail to happen."[3]

Because the event "can be divided in this conceptual way," it is only the *ratio* of the chances for happening to the chances for failing that matter. Bernoulli was the first to realize that if we conceived of things, such as the weather, *as if* such cases did really exist, we could learn about the corresponding ratio of hypothetical cases from empirical data. Although Laplace was maddeningly vague about his interpretation of the equally likely cases, I believe he had something similar in mind. Nowhere except in relation to games of chance or lotteries did he ever spell out what the cases actually were or how they might be counted. Moreover, he certainly understood that probability dealt only with the ratio of chances, never their absolute number: "The preceding notion of probability supposes that, in increasing in the same ratio the number of favorable cases and that of all the cases possible, the probability remains the same."[4]

THE GREAT UNRAVELING

Classical probability eventually reached its zenith during the early nineteenth century. Laplace and his followers had crafted a coherent, and seemingly airtight, system for thinking about uncertainty. Until around 1840, this system remained unassailable, and its scope of application gradually expanded. About that time, the general consensus about the meaning of probability began to break down. This occurred not primarily because classical probability was deemed defective *per se*, but rather because the ambitions of those applying this probability theory were too grand. In particular, the promise of being able to resolve difficult social and political issues through mathematical analysis was deemed unrealistic.

Laplace, initially inspired by Condorcet, had been optimistic about the ability of probability theory to place the "moral sciences" on a more scientific and rational footing. In his *Philosophical Essay*, three chapters were devoted to such applications.[5] These concerned the credibility of testimonies, the decisions of assemblies, and the judgments of tribunals. For example, Laplace was very concerned about the optimal number of jurors in a legal trial and the strength of the majority that should be required to condemn an accused person (e.g., 7 votes to convict out of 12 or 8 out of 10, etc.).

His followers, including most notably Siméon Poisson (1781–1840), drove the mathematical analysis of these issues to great lengths.[6] These efforts reached their culmination in Poisson's famous and controversial monograph (*Research on the Probability of Judgments in Criminal and Civil Matters*), published in 1837. However, the proposal of submitting such delicate considerations to a precise mathematical calculus provoked a strong reaction. In 1843, the philosopher John Stuart Mill famously condemned such applications of probability as the "opprobrium" of mathematics.[7]

Landmarks of Probability

1600

1654 Pascal & Fermat
1657 Huygens

1700

1713 Bernoulli
1718 De Moivre

1764 Bayes & Price
1774 Laplace

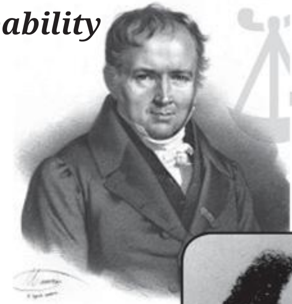
1800

1837 Poisson
1838 De Morgan
1843 Mill
1843 Cournot
1849 Ellis
1854 Boole
1866 Venn

1892 Peirce

1900

1921 Keynes
1921 Knight
1922 Fisher
1926 Ramsey
1926 Neyman & Pearson
1928 Von Mises
1937 De Finetti
1949 Reichenbach
1954 Savage
1959 Popper

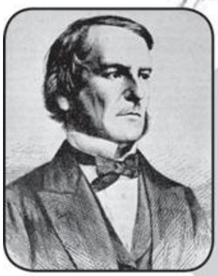


Siméon Denis Poisson

Augustus De Morgan

John Stuart Mill

Robert Leslie Ellis



George Boole

John Venn

Charles Sanders Peirce

A particular target of criticism was the theory of inverse probability, which had been proposed by Bayes and Laplace. Recall that Laplace had advocated the principle of insufficient reason: when we are completely ignorant, we should assign an equal degree of probability to each of the possible causes. Laplace went even further by equating each possible a priori value of the unknown probability with a distinct cause. Applying the principle of insufficient reason to underlying causes, he could supply a priori probabilities that would then be modified by empirical observations, using what we now call Bayes's rule. In this way, an inverse probability distribution for the unknown probability could be obtained.

Laplace's approach led him to derive a general rule for estimating the probability that a particular event will occur after a succession of occurrences and nonoccurrences. For example, suppose that all we know is that in a series of encounters between two top tennis players, Player A has won six matches and Player B has won two. What is the probability that Player A will prevail the next time they meet? Using inverse probability and the principle of insufficient reason, Laplace could calculate the answer to be 7/10. More generally, if an event (say, Player A wins) has occurred n times and failed m times, Laplace's famous Rule of Succession would estimate the probability of the event on the subsequent trial as

$$P(E) = \frac{n+1}{n+m+2}$$

Of particular significance is the special case when all of the previous trials have resulted in the event's occurrence. A frequently mentioned example was the probability that the sun will rise tomorrow, given that it has done so a great many times in succession so far. If n is the number of successive events observed, then the probability of an occurrence on the next possible occasion would be

$$P(E) = \frac{n+1}{n+2}$$

Of course, there are many possible arguments against such a facile application of probability theory, and Laplace himself would undoubtedly have applied his formula rather judiciously. Nonetheless, the idea that our ignorance itself could be harnessed to power the engine of inductive inference seemed too good to be true, rather like the dream of a perpetual motion machine.

Although the mathematical rules and usefulness of probability had been firmly established, the philosophical foundation of this concept began to be questioned. Probability still retained the formal definition of being a fraction of chances, but the intuitive understanding of "chances" so familiar to Huygens and Bernoulli had faded. By Poisson's time, probability was at least as "real" as the metaphysical chances on which it was, in theory, based. But without the intuitive "ratio of chances" conceptualization that had motivated Bernoulli and Huygens, probability lacked a firm foundation. Difficult questions began to be posed about the true meaning of probability.

These questions have been discussed and debated ever since. There have been a great many attempts by mathematicians, philosophers, statisticians, and scientists to pin down this elusive concept.[8] For the most part, the various proposals have laid primary emphasis on one or another of four common intuitions about probability.

- Probability pertains to a relative *frequency* of occurrence.
- Probability is a *logical* relationship between evidence and belief.
- Probability measures a *subjective* degree of belief.
- Probability is a measure of the *propensity* to occur.

From the classical (metaphorical lottery) perspective, these are four aspects, or connotations, of probability. Each aspect is consistent with and implied by the fundamental analogy between something uncertain and the outcome of an idealized game of chance. Defining probability in terms of any one of these connotations promises greater clarity but seems in some way inadequate.

PROBABILITY AS A RELATIVE FREQUENCY

One of the earliest proposals to define unambiguously the true meaning of probability was put forward by an English mathematician named Robert Leslie Ellis (1817–1859) in an essay read to the Cambridge Philosophical

Society in 1842.[9] "On the Foundations of the Theory of Probabilities" was published in 1849. In it, Ellis suggested a radically new interpretation of Jacob Bernoulli's law of large numbers. Recall that Bernoulli in Part Four of the *Ars Conjectandi* had, in effect, been the first to prove that the observed frequency of an event in a large number of trials would eventually converge toward the true underlying ratio of chances.

Bernoulli believed his proof provided a rigorous justification for adopting the long-run observed frequency as an approximation to the true underlying degree of belief, or probability. Ellis lived in a time when the concept of hidden cases, or chances, had become rather nebulous. Meanwhile, mathematical probability had become intuitively meaningful. Additionally, there was a lot more actual data available, so that it was easier to imagine frequencies of events in a large number of trials. Ellis believed that these observable frequencies constituted the natural foundation for the concept of probability. He interpreted the vaunted laws of large numbers proved by Bernoulli, de Moivre, Laplace, and Poisson as mere tautologies:

> If the probability of a given event be correctly determined, the event will, on a long run of trials, tend to recur with frequency proportional to this probability. This is generally proved mathematically. It seems to me to be true *a priori*.

His main argument was that this frequency in a long run of independent trials is in fact what we really *mean* by a probability:

> When on a single trial we expect one event rather than another, we necessarily believe that on a series of similar trials the former event will occur more frequently than the latter. The connection between these two things seems to me to be an ultimate fact … the evidence of which must rest upon an appeal to consciousness.… after giving a painful degree of attention to the point, I have been unable to sever the judgment that one event is more likely to happen than another, or that it is to be expected in preference to it, from the belief that on the long run it will occur more frequently.

For Ellis, and for many who followed in his tracks, the notion of a long-run frequency in a potentially observable series of trials was both intuitive and logically defensible. For them, recourse to some vague metaphysical notion of equally likely "chances" was not persuasive. Why not simply dispense with such a fiction and define an event's probability to be its proportion of occurrences in a long series of similar occasions?

Independently, and almost simultaneously, a very similar idea occurred to the French economist and philosopher Antoine Augustin Cournot (1801–1877).[10] In his *Exposition of the Theory of Chances and Probabilities* (1843), Cournot distinguished between an objective version of probability and a subjective one. Objective probabilities, reflected in observable frequencies, "have an objective existence, which gives a measure of the possibility of things." These he contrasts with subjective probability, which are "relative in part to our knowledge, in part to our ignorance, and which vary from one individual to another, according to their capacities and the data provided to them." Thus began a long conversation that has persisted to the present day regarding the subjective and objective aspects of probability.

Two decades later, John Venn (1834–1923) expanded upon the idea of objective probability in his *Logic of Chance*, published in 1866.[11] Venn asserted that the word "probability" is best reserved for a particular kind of use. Probability for him represented an efficient way to describe or summarize a complex empirical reality. John Maynard Keynes a half-century later summarized Venn's theory:

> The two principle tenets, then, of Venn's system are these—that probability is concerned with series or groups of events, and that all the requisite facts must be determined empirically, a statement in probability merely summing up in a convenient way a group of experiences. Aggregate regularity combined with individual difference happens, he says, to be characteristic of many events in the real world.… As our knowledge regarding the class as a whole may give us valuable guidance in dealing with an individual instance, we require a convenient way of saying that an individual belongs to a class in which certain characteristics appear on the average with a known frequency … The importance of probability depends solely upon the actual existence of such groups … and a judgment of probability must necessarily depend for its validity upon our empirical knowledge of them.[12]

In this view, probability is explicitly objective in nature, rather than conceptual, and is a literal, if approximate, statistical summary of empirical reality.

As Keynes noted, this perspective has the virtue of avoiding metaphysical ambiguities but at the cost of severely restricting the meaning and usefulness of probability:

> It is the obvious, as well as the correct, criticism of such a theory, that the identification of probability with statistical frequency is a very grave departure from the established use of words; for it clearly excludes a great number of judgments which are generally believed to deal with probability.... Venn's theory by itself has few practical applications, and if we allow it to hold the field, we must admit that probability is *not* the guide of life, and that in following it we are not acting according to reason.

Venn's hard-headed *frequentist* alternative was widely regarded as a salutary antidote to the excesses of Laplace and Poisson. For example, to assign a probability to the event that a tribunal would reach a correct decision was nonsense, according to Venn's point of view. Where could we obtain data pertaining to the probability that an individual judge would rule correctly?

The spirit of Venn's frequentism has animated most statistical analyses ever since. Most statisticians believe that their subject deals with summarizing and analyzing patterns in large aggregates of observations or measurements. Their concerns are pragmatic, and their implicit philosophy tends to be positivistic. Early in the twentieth century, frequentist probability attracted champions among positivistic philosophers of science, most notably Richard von Mises (1883–1953) and Hans Reichenbach (1891–1953).

Von Mises saw probability as a form of applied mathematics applicable to "mass phenomena" in the same way that differential and integral calculus are necessary to describe the laws of physics.[13] The essence of a mass phenomenon was "unlimited repetition" of a particular kind of uniform occurrence:

> The rational concept of probability, which is the only basis of probability calculus, applies only to problems in which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time. Using the language of physics, we may say that in order to apply the theory of probability we must have a practically unlimited sequence of uniform observations.

Central to his conceptualization was the idea of a *collective*. A collective "denotes a sequence of uniform events or processes which differ by certain attributes."

> All the peas grown by a botanist concerned with the problem of heredity may be considered as a collective, the attributes in which we are interested being the colors of the flowers. All the throws of dice made in the course of a game form a collective, wherein the attribute of the single event is the number of points thrown. Again, all the molecules in a given volume of gas may be considered as a collective, and the attribute of a single molecule might be its velocity. A further example of a collective is the whole class of men and women whose ages at death have been registered by an insurance office.

Von Mises adopted the motto: "First the collective—then the probability." He tried to explain what he meant more precisely:

> We will say that a collective is a mass phenomenon or a repetitive event … for which there are sufficient reasons to believe that the relative frequency of the observed attribute would tend to a fixed limit if the observations were indefinitely continued. This limit will be considered *the probability of the attribute considered within the given collective* … 'The probability of winning a battle', for instance, has no place in our theory of probability, because we cannot think of a collective to which it belongs.

It is evident that for von Mises the "scientific," as opposed to colloquial, meaning of probability pertained to objective facts about the real world. He believed that certain mass phenomena could be identified that corresponded more or less closely to the idealization of a collective. To the extent that a designated collective in the real world satisfied his definition of a collective, the usual rules of mathematical probability would be applicable.

Like his predecessors, such as Ellis and Venn, von Mises took as a given that the sequence would approach a fixed limiting value of the relative frequency in a long run; it was not something that needed to be proved. Once a sequence of observations was properly defined, they thought, the proportion of events in this extended sequence was by definition the probability. However, von Mises recognized something more that needed to be considered to avoid an essential ambiguity. As an example of the problem, he discussed a series of road markers consisting of small stones placed at intervals of exactly 1/10 of a mile and a large stone placed at each mile mark. He noted that, on such a road, the fraction of large stones observed after several miles would be approximately 1/10 and would approach this fraction more and more closely as we proceed. He contrasted this situation to that in a game of chance:

> After having just passed a large stone, we are in no doubt about the size of the next one; there is no chance of its being large. If, however, we cast a double 6 with two dice, this fact in no way affects our chances of getting the same result in the next cast. Similarly, the death of an insured person during his forty-first year does not give the slightest indication of what will be the fate of another who is registered next to him in the books of the insurance company, regardless of how the company's list was prepared.

Accordingly, von Mises posited that a collective must be a sequence of observations that satisfy a condition of complete lawlessness, or randomness. Here, von Mises confronted head-on one of the essential aspects of ambiguity inherent in (additive) mathematical probability by virtue of its metaphorical lottery origins. He made a heroic attempt to eliminate, or at least minimize, this ambiguity by introducing the *principle of the impossibility of a gambling system*. He regarded a sequence to be random if there was no possibility of identifying any subset of observations whose limiting frequency differed from that characterizing the whole sequence.

Identifying such a subset of observations would allow cherry-picking of a favorable subsequence in betting against someone who accepted the odds associated with the entire sequence. This principle of the excluded gambling system had for him a status akin to that of the principle of conservation of energy for the physicist. It was his sole justification for treating the outcomes of mass phenomena like genetic variability or life expectancy as probabilistic:

> The whole financial basis of insurance would be questionable if it were possible to change the relative frequency … by excluding, for example, every tenth one of the insured persons, or by some other selection principle. The principle of the impossibility of a gambling system has the same importance for the insurance companies as the principle of the conservation of energy for the electric power station: it is the rock on which all the calculations rest.

Later critics have adduced many problems with a frequentist interpretation along the lines laid out by von Mises and the positivistic philosophy upon which it was based. Three particular issues have received the most attention.

- The meaning of randomness
- The reference class problem
- The problem of the single case

I will briefly consider each of these issues from the metaphorical lottery perspective.

The Meaning of Randomness

Pure randomness can never be defined unambiguously, because absence of any possible pattern cannot be made precise. On the other hand, randomization, as what happens approximately when a well-designed "random" mechanism is employed, is easy enough to grasp conceptually. Modern frequentists, starting with von Mises, have often appealed to practical experience to support the existence of randomness, as *exemplified* by well-constructed games of chance.

The lack of a successful gambling system in any real-world situation, such as insurance, means that the relevant causal process is very similar to that in an idealized lottery. But the precise nature of this similarity cannot be spelled out explicitly. We are left, as always, with the idea of an analogy between some real-world process and a game of chance. All gambling systems fail precisely *because* the process is in certain relevant respects *just like* an idealized game of chance. So, defining randomness in terms of such failure is ultimately circular; it all boils down to an analogy with a game of chance.

On the other hand, it has been argued that a pragmatic definition of randomness might be adequate for practical purposes. Suppose we can specify some observable property of a process or an actual sequence that would allow it to be considered effectively random. Such a "pseudorandom" sequence would certainly be satisfactory for many practical purposes, such as selecting a representative sample from some population or assigning treatments in a clinical trial to compare alternative treatments. In fact, computer algorithms that generate such pseudorandom sequences are employed routinely for such purposes.

Should we perhaps define a collective to be a sequence in which pseudorandomness of some sort prevails? That would solve the philosophical conundrum, as some have suggested, but would be rather contrived. It is not what we intuitively mean about randomness. This problem was recognized by other philosophers, including Hans Reichenbach. His proposal was to define randomness relative to our ability to detect a pattern[14]:

> Random sequences are characterized by the peculiarity that a person who does not know the attributes of the elements is unable to construct a mathematical selection by which he would, on an average, select more hits than would correspond to the frequency of the major sequence .... In this form, the impossibility of making a deviating selection is expressed by a psychological, not a logical, statement; it refers to acts performed by a human being. This may be called a psychological randomness.

For instance, consider a very long sequence that contains half zeroes and half ones. You cannot see this sequence but are allowed to specify any pattern of your own, and will receive a point for each time your value matches that of the unknown sequence. Then the unknown sequence would be deemed "psychologically random" if no human being could concoct a system that would in the long run allow her to obtain more than 50% of hits.

Writing in 1949, Reichenbach could not envision the possibilities that have become open to a human aided by a high-speed computer. Perhaps he would have attempted a definition of "digital randomness" limited by the computational capabilities of digital computers. In the end, I feel confident that no such "rigorous" definition of randomness is possible. Rather, we must accept that pure randomness is merely an idealization of that which can occur in an actual game of chance.

To be random is to lack any pattern or order, but what do we mean by order? So, the idea of randomness is intuitively meaningful but ultimately ambiguous, much like causation. Nevertheless, like the tangent line to a curve, a purely random process seems to "exist" as a kind of limit. We can obviously construct *approximately* random mechanisms (e.g., dice) that are certainly good enough artificial randomizers for most practical purposes. So, it is not difficult for us to imagine a perfect lottery, for example, as an idealization that can only be approached but never completely realized.

The Reference Class Problem

The question of which reference class to adopt has always been awkward for frequentists. On the one hand, they perceive probability as an objective aspect of the real world. On the other, they recognize that there are many possible reference classes to which an individual event might be assigned. The choice of reference class is somewhat arbitrary and subjective, at the discretion of the person making the probability assessment. How can these divergent perspectives possibly be reconciled?

Consider the problem faced by a surgical candidate, Jane, and her physician. There are many factors that might be considered relevant to the outcome of her contemplated operation. Which of these should be taken into account in specifying Jane's "real" probability of a successful outcome? Of these, which can *in fact* be taken into account because there is pertinent data? Von Mises did not have much to say about such situations. He seemed content to restrict "scientific" probability to those few "mass phenomena" in which it was apparent that the outcomes would occur effectively at random. Presumably, the aggregate of all surgical operations like Jane's would not qualify as a collective in his sense.

Reichenbach, like many other frequentists, was not satisfied with such a narrow view of the scope of probability. So, he grappled with the reference-class problem and adopted a more pragmatic position than von Mises. He believed that the reference class should be chosen to admit as many cases "similar" to the one in question as possible but no more. Similarity was defined in terms of "relevant" characteristics of the situation. In other words, we must strive to account for what is truly relevant and to ignore all else. In that way, our probability is relative to a "homogeneous" reference class. This homogeneous class is conceptualized as the broadest class within which there are no relevant differences among its members.

Furthermore, Reichenbach was aware that in practice, our data pertaining to potentially relevant factors is always limited. So, even if we can delineate in theory the relevant factors, we may not have sufficient data to adjust for them. For example, Jane's age and state of health are obviously relevant, but the available statistics might not be broken down by these factors. Even if we possess some relevant data, the number of previous cases in each of the subclasses may be inadequate for precise estimation of probabilities. Reichenbach recommended that we rely on the "smallest" class about which we have "adequate" statistics. This makes sense conceptually perhaps but remains quite ambiguous.

The Problem of the Single Case

Closely related to the reference class problem is the problem of the single case. From a frequentist perspective, probability is a property of the entire reference class. Only secondarily and figuratively can we refer to the probability of a unique event: "First the collective—then the probability." From this perspective, the probability of a particular event is merely a shorthand for the long-run frequency of occurrence that is approached in the limit. It has a sort of "as-if" quality; we speak about the probability of a particular event, but we always have in mind a class of events.

Von Mises was completely untroubled by the inapplicability of probability to unique events. He stated explicitly that probability as applied to an individual event was meaningless. Other frequentists have been less doctrinaire. They accept the basic idea that probability must refer to some kind of sequence of homogeneous occurrences, either actual or conceptual, but seek a bridge to the more colloquial usage of the word.

PROBABILITY AS A LOGICAL RELATIONSHIP

A contemporary of Robert Leslie Ellis, Augustus De Morgan (1806–1871), took a very different tack regarding probability. De Morgan was a professor of mathematics at University College in London who was best known for his pioneering work on symbolic logic. Unlike Ellis, De Morgan basically subscribed to the "classical" interpretation of probability bequeathed by Laplace and Poisson. He shared their belief that probability was a measure of our ignorance about the true state of affairs.

For De Morgan, probability had no necessary connection with real frequencies of repetitive phenomena, but pertained merely to our mental grasp of the real world. In an influential essay published in 1838, he situated probability squarely within the realm of logical analysis: "I throw away objective probability altogether, and consider the word as meaning the state of mind with respect to an assertion, a coming event, or any other matter on which absolute knowledge does not exist."[15,16]

For De Morgan, probability was much broader in scope than it was for Ellis or Venn; it could be applied to any assertion or event. However, De Morgan gave little guidance concerning how he believed the "state of mind" about an assertion or event might be obtained. He regarded probability as based upon a *rational* opinion, such as that which would be entertained by someone (presumably an educated Victorian Englishman) who could correctly interpret all available information. Undoubtedly, this sort of appeal to an ideal standard of rationality was more persuasive in 1838 than it would be for us today. Even so, De Morgan's concept represented only a beginning that would soon be fleshed out by others.

Perhaps the quintessential rational man of the nineteenth century was George Boole (1815–1864). Most famous today as the eponymous inventor of Boolean algebra, he made substantial contributions to several areas of mathematics and logic. Indeed, he is largely responsible for the fact that we consider logic today to be fundamentally mathematical in nature. Unfortunately for him, however, his logical acumen was apparently more theoretical than practical.

One day, while walking the two miles from home to his class at the University of Cork, in Ireland, where he had taught for many years, Prof. Boole was caught in a sudden downpour. Soldiering on, he proceeded to lecture for a full hour while remaining in his soaking garments. Upon returning home, he became seriously ill and was confined to bed. To compound matters, according to lore, his devoted but misguided wife, apparently subscribing to the "hair of the dog" theory of medicine, decided to pour cold water over him periodically. Alas, this regimen proved a dismal failure, and probably contributed to the good professor's untimely demise.

A decade prior to this unfortunate incident, Boole had published his masterpiece: *An Investigation of the Laws of Thought* (1854).[17] Boole created symbolic logic in order to express rigorously how we ought to reason about the relationships among statements that can be true or false. Given relevant, but the available statistics might not be broken down by these factors. Even if we possess some relevant data, the number of previous cases in each of the subclasses may be inadequate for precise estimation of probabilities. Reichenbach recommended that we rely on the "smallest" class about which we have "adequate" statistics. This makes sense conceptually perhaps but remains quite ambiguous.

The Problem of the Single Case

Closely related to the reference class problem is the problem of the single case. From a frequentist perspective, probability is a property of the entire reference class. Only secondarily and figuratively can we refer to the probability of a unique event: "First the collective—then the probability." From this perspective, the probability of a particular event is merely a shorthand for the long-run frequency of occurrence that is approached in the limit. It has a sort of "as-if" quality; we speak about the probability of a particular event, but we always have in mind a class of events.

Von Mises was completely untroubled by the inapplicability of probability to unique events. He stated explicitly that probability as applied to an individual event was meaningless. Other frequentists have been less doctrinaire. They accept the basic idea that probability must refer to some kind of sequence of homogeneous occurrences, either actual or conceptual, but seek a bridge to the more colloquial usage of the word.

PROBABILITY AS A LOGICAL RELATIONSHIP

A contemporary of Robert Leslie Ellis, Augustus De Morgan (1806–1871), took a very different tack regarding probability. De Morgan was a professor of mathematics at University College in London who was best known for his pioneering work on symbolic logic. Unlike Ellis, De Morgan basically subscribed to the "classical" interpretation of probability bequeathed by Laplace and Poisson. He shared their belief that probability was a measure of our ignorance about the true state of affairs.

For De Morgan, probability had no necessary connection with real frequencies of repetitive phenomena, but pertained merely to our mental grasp of the real world. In an influential essay published in 1838, he situated probability squarely within the realm of logical analysis: "I throw away objective probability altogether, and consider the word as meaning the state of mind with respect to an assertion, a coming event, or any other matter on which absolute knowledge does not exist."[15,16]

For De Morgan, probability was much broader in scope than it was for Ellis or Venn; it could be applied to any assertion or event. However, De Morgan gave little guidance concerning how he believed the "state of mind" about an assertion or event might be obtained. He regarded probability as based upon a *rational* opinion, such as that which would be entertained by someone (presumably an educated Victorian Englishman) who could correctly interpret all available information. Undoubtedly, this sort of appeal to an ideal standard of rationality was more persuasive in 1838 than it would be for us today. Even so, De Morgan's concept represented only a beginning that would soon be fleshed out by others.

Perhaps the quintessential rational man of the nineteenth century was George Boole (1815–1864). Most famous today as the eponymous inventor of Boolean algebra, he made substantial contributions to several areas of mathematics and logic. Indeed, he is largely responsible for the fact that we consider logic today to be fundamentally mathematical in nature. Unfortunately for him, however, his logical acumen was apparently more theoretical than practical.

One day, while walking the two miles from home to his class at the University of Cork, in Ireland, where he had taught for many years, Prof. Boole was caught in a sudden downpour. Soldiering on, he proceeded to lecture for a full hour while remaining in his soaking garments. Upon returning home, he became seriously ill and was confined to bed. To compound matters, according to lore, his devoted but misguided wife, apparently subscribing to the "hair of the dog" theory of medicine, decided to pour cold water over him periodically. Alas, this regimen proved a dismal failure, and probably contributed to the good professor's untimely demise.

A decade prior to this unfortunate incident, Boole had published his masterpiece: *An Investigation of the Laws of Thought* (1854).[17] Boole created symbolic logic in order to express rigorously how we ought to reason about the relationships among statements that can be true or false. Given a specified set of premises, how should we logically infer that a particular conclusion is true or false? Probability theory, on the other hand, concerned how we should think rationally when our knowledge was incomplete: "Probability is expectation founded upon partial knowledge."

Like De Morgan, Boole asserted that probability pertained to a rational state of mind about some event or proposition. However, just as logic concerned relationships among propositions about events, not actual events themselves, so probability also dealt with propositions:

> Although the immediate business of the theory of probability is with the frequency of the occurrence of events … the theory of probabilities must bear some definite relation to logic. The events of which it takes account are expressed by propositions; their relations are involved in the relations of propositions. Regarded in this light, the object of the theory of probabilities may be thus stated: Given the separate probabilities of any propositions to find the probability of another proposition.[18]

Boole was emphasizing here the epistemic aspect of probability, somewhat echoing Jacob Bernoulli's ideas. Indeed, like Bernoulli, Boole explicitly stated that probability statements must rest on the evidence available:

> *Probability*, in its mathematical acceptation has reference to the state of our knowledge of the circumstances under which an event may happen or fail. With the degree of information which we possess concerning the circumstances of an event, the reason that we have to think that it will occur, or, to use a single term, our *expectation* of it, will vary.[19]

In this passage, he seems to have in mind some way of combining "arguments" along the lines suggested in Part Four of *Ars Conjectandi*. However, there is a critical difference. Writing 200 years after the seminal correspondence between Pascal and Fermat in 1654, Boole can only mean by probability "the ratio of the number of cases favorable to that event, to the total number of cases favorable or contrary, and all equally possible." He construes the mathematical laws of this classical probability to be in fact the "laws of thought" that should govern our rational deliberations. In this respect he is closer in spirit to Laplace, for whom probability embodied "common sense reduced to calculus."

To grasp Boole's logic we must understand what he considered to be the scope of probability and the kind of evidence on which it can depend.

In order to be the basis of rational expectations regarding events, this evidence must either be "deduced from a knowledge of the peculiar constitution of things" or "derived from the long-continued observation of past series of their occurrences and failures." By the former, he meant such considerations of symmetry as prevail in games of chance; by the latter, he meant statistical evidence based on empirical data.

In contrast with frequentists like Ellis and Venn, however, Boole was not *defining* the probability as an actual, or even hypothetical, frequency. For him, probability was essentially a measure of the rational grounds for belief. Thus, knowledge of symmetry or of statistical frequencies could sometimes provide the evidential basis, or at least some substantial part of it. Moreover, *only* when our evidence has this character, according to Boole's thinking, can we truly be said to have a *rational* expectation.

By adopting this stance, Boole narrowed the scope of probability just as surely as the frequentists had. For example, the probability that France would invade Holland would have had no meaning for him, unless some series of situations could be invoked as a logical reference class. Moreover, purely personal opinions about an event that were not grounded in symmetry considerations or statistical frequencies were effectively barred from his theory. In essence, Boole was completely in accord with Jacob Bernoulli when it came to situations in which any relevant information about individual circumstances (such as our surgical candidate Jane's state of health) was lacking. However, the broader concept of epistemic probability familiar to Bernoulli had by his time lost all meaning.

An important corollary of Boole's position was his strong rejection of Laplace's principle of insufficient reason and its implications. Recall that for Laplace, the starting point for many analyses was an a priori distribution of ignorance. All possible values of an unknown probability could be deemed equally probable if we had no "sufficient reason" to believe otherwise. This a priori supposition could then be combined with the likelihood provided by observed data using Bayes's Theorem to derive the inverse probabilities (probabilities of causes).

As mentioned previously, by 1854 the bloom had already faded from this rose, thanks to skeptics like John Stuart Mill who felt that Laplace and Poisson had gone much too far. Laplace's famous Rule of Succession was subjected to particular derision. Boole held that a firm basis for probability judgments, related to symmetry or statistical frequencies, was necessary.

So, he could not tolerate ignorance *per se* as such a basis. Commenting on the Rule of Succession, he noted:

> It has been said that the principle involved … is that of the equal distribution of our knowledge, or rather of our ignorance—the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability. I apprehend, however, that this is an arbitrary method of procedure.[20]

Remember that Boole had defined probability as a fraction of favorable cases among a total number of cases, *all equally possible*. Laplace might have argued that absent knowledge to the contrary, all cases can be regarded as equally possible. This was very convenient, but not good enough for Boole, who demanded some positive grounds upon which to base an assumption of equally possible.

Boole's perspective on probability was highly influential and the first of many attempts over the ensuing 150 years to cast probability as a logical relationship. Some advocates of the logical interpretation have been closer in spirit to the frequentists, emphasizing the importance of a long series of similar events as the primary source of evidence in practical applications. These have tended to be unsympathetic to the principle of insufficient reason. Others have been more open to the principle and "Bayesian" analysis more generally.

Indeed, Boole's most immediate successor in the logical school of probability held quite a different view of Laplace's controversial proposal.[21] William Stanley Jevons (1835–1882) began as a professor of logic and moral philosophy and ended as a professor of political economy. Like Boole, Jevons believed that probability belonged "wholly to the mind" and dealt with our "quantity of knowledge" about something. He even rejected "belief" altogether as a word much too nebulous to be scientifically useful. Probability "does not measure what the belief is, but what it ought to be."

Had he managed to survive his ill-fated drenching episode to read these central tenets in Jevons's *The Principles of Science* (1873), Boole would certainly have agreed. However, Jevons parted company with Boole when it came to the logical basis for probability assessments.[22] Jevons swallowed wholeheartedly Laplace's theory of inverse probability. To apply the inverse method in practice, Laplace's principle of insufficient reason was, according to Jevons, a useful expedient. He admitted it to be somewhat arbitrary, as alleged by Boole and others, but defended its utility on pragmatic grounds.

The idea of probability as a branch of logic has continued to attract advocates down to the present day. Many different theories along these lines have been propounded, varying widely in their details.[23] However, nearly all of these doctrines have had one critical aspect in common: they assume the usual mathematical version of probability.[24] With one major exception, proponents of logical probability have adhered to the restriction that probability must satisfy the Kolmogorov axioms. The one great thinker who adopted a much broader view of the subject is all but ignored today, although his economic theories remain very influential. Indeed, it is in the context of economics that his aphorisms about uncertainty are often quoted.

Keynesian Probability

John Maynard Keynes (1883–1946) is perhaps the most well-known economist of the twentieth century. His early intellectual development and career, starting with his membership in a genius cluster that developed around Cambridge University in the early 1900s is well chronicled. This extraordinary group included such luminaries as Bertrand Russell (1872–1970), Alfred North Whitehead (1861–1947), Lytton Strachey (1880–1932), G.E. Moore (1873–1958), G.H. Hardy (1877–1947), and E.M. Forster (1879–1970). During the years 1906–1911, Keynes was developing his theory of probability as a branch of logic that could deal with situations in which the evidence was not fully conclusive, but rather entailed some degree of uncertainty.

Keynes was a rising star in the intellectual firmament, and his book was much anticipated. However, owing to the Great War and his civic responsibilities, it was not until 1921 that *A Treatise on Probability* was finally published. The philosopher C.D. Broad (1887–1971) captured in his review the initial enthusiasm with which the *Treatise* was greeted: "I can only conclude by congratulating Mr. Keynes on finding time, amidst so many public duties, to complete this book, and the philosophical public on getting the best work on Probability that they are likely to see in this generation."[25] However, the mathematical development and practical applications of probability soon took center stage. Keynes's ambitious and profound philosophical contribution is almost entirely unknown to practicing statisticians and scientists today.

A substantial portion of the *Treatise* is taken up with a kind of symbolic logic somewhat similar to that developed by Russell and Whitehead in their famous *Principia Mathematica*, which was being composed around the same time. Of much greater practical importance, however, was his extensive discussion of the philosophical underpinnings of inductive logic. Keynes took a very broad view of probability, which he conceived as pertaining to "the various degrees of rational belief about a proposition which different amounts of knowledge authorize us to entertain."

Keynes regarded ordinary mathematical probability as an important aspect of the logical relations between uncertain propositions, but only applicable under certain limited conditions. In this respect, his notion of probability harked back to the kind of probability envisioned by Jacob Bernoulli, whom he considered "the real founder of the classical school of mathematical probability." Like Bernoulli, Keynes emphasized what he termed Bernoulli's second axiom, which he interpreted as meaning "that in reckoning a probability we must take everything into account."

Keynes followed Boole in that his concept of probability was thoroughly epistemic; probability pertained to propositions, not to events *per se*. A fundamental idea for Keynes was that the probability of any *hypothesis* must be conditioned on the available *evidence*, whatever that evidence might be. Given the evidence, our knowledge may assume a variety of different forms, and may authorize a merely qualitative or sometimes a quantitative probability assessment. For some propositions, our relevant information may be so limited that a probability cannot even be said to exist at all.

For example, suppose that someone is asked about the probability that Brazil will win the next soccer World Cup. Both the magnitude of such a probability and its precision would vary widely across individuals. At one extreme, a person might have only the vaguest of conceptions, not even amenable to quantification. At the other extreme, a soccer aficionado might be aware of many relevant factors, including recent performance by Brazil's team, injuries to key players, and the odds currently being offered in Rio de Janeiro.

The distinction between a hypothesis whose probability is unknown and one whose probability is nonexistent, or indeterminate, was critical for Keynes.

> I am here dealing with a probability in its widest sense, and am averse to confining its scope to a limited type of argument. If the opinion that not all probabilities can be measured seems paradoxical, it may be due to this divergence from a usage which the reader may expect … I maintain, then, in what follows, that there are some pairs of probabilities between the members of which no comparison of magnitude is possible; that we may say, nevertheless, … that the one is greater and the other less, although it is not possible to measure the difference between them; and that in a very special type of case … a meaning can be given to a numerical comparison of magnitude.[26]

From this perspective, an assessment of probability can be expressed as a number between zero and one in the usual manner *only* under special conditions that Keynes attempted to spell out.

One such fundamental condition is that the base of evidence assumed relevant to a proposition and its contrary must be the same. Otherwise, comparing the probability of a statement and its contrary would not be meaningful. Recall that Jacob Bernoulli discussed, in Part Four of his *Ars Conjectandi*, a situation in which the probability of an event was 3/4 and its contrary 2/3. Keynes recognized that to avoid such a possibility would require the tacit assumption that both probabilities were based on the same body of evidence:

> Some probabilities are not comparable in respect of more or less, because there exists more than one path, so to speak, between proof and disproof, between certainty and impossibility; and neither of two probabilities, which lie on different paths, bears to the other and to certainty the relation of 'between' which is necessary for quantitative comparison.

In other words, if there is ambiguity about the situation under consideration, we cannot determine uniquely how "far" from certainty to place a particular outcome. Only if we can eliminate such ambiguity can our degrees of uncertainty be ranged on the single dimension of doubtfulness, so that various degrees of evidentiary support can be compared directly.

Keynes recognized that a *numerical* probability was conceptually based on some notion of a fraction of equally possible cases. But to be deemed equally possible, the cases must first of all be comparable. So, in principle, we would need to establish somehow that all the cases can be arrayed along the same evidentiary "path." Laplace's principle of insufficient reason held that we could always assume equal probabilities absent any sufficient reason to the contrary. Keynes found this unsatisfactory, especially in the important context of unknown prior probabilities. Keynes pointed out that applying this principle when we are truly ignorant about prior probabilities required some positive justification.

Beyond this general philosophical objection, he discussed a number of logical paradoxes that could arise when attempting to apply the principle of insufficient reason. For example, recall the confusion of Jean le Rond D'Alembert, Laplace's mentor, regarding the probability of obtaining heads at least once when allowed two tosses of a coin that can land heads (H) or tails (T). He observed that there are three possible outcomes: H on the first toss, T on the first and H on the second, and T on both. Therefore, D'Alembert at first believed that the probability was 2/3 of winning. He was later embarrassed to admit his error, after he realized that these three outcomes were not equally likely, so the true probability must be 3/4. This example was mentioned by Laplace in his discussion of the "delicate" considerations that can be entailed in probability theory.

What Laplace had no way to fully appreciate was the degree of delicacy that was truly involved. Classical probability essentially defines probability as a fraction of imaginary chances in a metaphorical lottery. However, the precise properties of this probability can depend on exactly how we conceive of this metaphorical lottery, and there may be different ways to do this. For example, Bayes imagined a probability as the fraction of the distance traveled by a ball on a smooth table relative to the table's length. Laplace, on the other hand, envisioned a lottery in which an unknown proportion of tickets were winners. Both "models" of reality are conceivable, but their consequences are not always identical!

Keynes considered the case in which we select a ball from an urn in which there are a given number of balls, some black and some white, in an unknown proportion.[27] Suppose we are interested in the proportion of white balls in the urn. The usual assumption is that all the possible (i.e., conceivable) *proportions* are equally likely. Keynes discussed the situation in which there are four balls in the urn. Suppose first we assume that there are five equally possible values for the number of white balls: 0, 1,

TABLE 8.1 Keynes's Urn: Possible *Constitutions* vs. Possible *Numbers* of White Balls

Constitution	Ball 1	Ball 2	Ball 3	Ball 4	Number White
1	W	W	W	W	**4**
2	W	W	W	B	**3**
3	W	W	B	W	**3**
4	W	B	W	W	**3**
5	B	W	W	W	**3**
6	W	W	B	B	**2**
7	W	B	W	B	**2**
8	W	B	B	W	**2**
9	B	W	B	W	**2**
10	B	B	W	W	**2**
11	B	W	W	B	**2**
12	W	B	B	B	**1**
13	B	W	B	B	**1**
14	B	B	W	B	**1**
15	B	B	B	W	**1**
16	B	B	B	B	**0**

2, 3, and 4. This assumption corresponds to a method that first selects a number between zero and four at random, and then places that number of white balls and the rest black in the urn. However, there is another equally logical way to express our ignorance.

Each of the four balls in the urn can be either black or white. Imagine that the balls are selected and placed in the urn one at a time, and the color determined by tossing a fair coin. Then, there would be 16 possible *constitutions* of the urn, as shown in Table 8.1. Suppose we consider all 16 constitutions equally likely. Then, the probabilities of the different *proportions* would not be equal. For example, the probability of getting zero white balls would be 1/16, but the probability of getting two would be 3/8. Conversely, if we chose to regard each proportion as equally probable, then the constitutions would not be equally likely.

Which assumption is correct? If we mean which is the "real" probability, there is no unique answer. In an actual situation in which a real urn had been filled with real balls, there would in fact be a definitive answer; the probabilities would depend on exactly how the balls were selected and placed in the urn. But in the abstract scenario, as usually described, we are not given these details. We may not even know the number of balls, let alone the procedure for filling the urn. So, there is an inherent ambiguity in the determination of these probabilities.

Nonetheless, Keynes drew a distinction that he believed could often help to avoid such ambiguity. He suggested that the principle of insufficient reason, which he preferred to term the *principle of indifference*, can only generate a paradox when the presumed equally likely events are not *elementary units*:

> The examples in which the Principle of Indifference broke down had a great deal in common. We broke up the field of possibility, as we may term it, into a number of disjunctive judgments.… The paradoxes and contradictions arose, in each case, when the alternatives, which the Principle of Indifference treated as equivalent, actually contained or might contain a different or an indefinite number of more elementary units.

Applying this insight to the four-ball urn example, Keynes concluded that the approach based on constitutions was potentially acceptable, while the one based on the proportions was not.

Keynes's notion of "elementary units" contains the germ of an important idea, although difficult to apply. Mathematical probability is always rooted conceptually in the analogy to a metaphorical lottery in which one among some number of possible chances, or eventualities, is being chosen at random. Thus, applying this conceptual model to an actual situation in which the various eventualities are actually specified may not make sense if some of these eventualities can be regarded as *compound* in nature.

In D'Alembert's coin-tossing problem, the compound nature of one eventuality (heads on the first trial) seems evident to us. Fermat was the first to suggest that it could usefully be viewed as two outcomes (HT, HH) among the four that could possibly occur (HT, HH, TH, and TT). But when we enter the domain of real-world problems, the precise meanings of "elementary" and "compound" are quite ambiguous. Keynes went on to discuss how, once a set of alternatives has been established, we ought to judge whether the principle of indifference is warranted:

> The principle states that 'There must be no known reason for preferring one of a set of alternatives to any other.' What does this mean? What are 'reasons,' and how are we to know whether they do or do not justify us in preferring one alternative to another? I do not know any discussion of Probability in which this question has been so much as asked.

He proceeded to explain what he believed to be the essential consideration that must be entailed in a rational judgment of indifference.

Any relevant information pertaining to the alternatives must be equally balanced. For example, in the case of two possible alternatives:

> There must be no *relevant* evidence relating to one alternative, unless there is *corresponding* evidence relating to the other; our relevant evidence, that is to say, must be symmetrical with regard to the alternatives, and must be applicable to each in the same manner. This is the rule at which the Principle of Indifference somewhat obscurely aims.

Determining the relevance of our evidence requires "a series of judgments of relevance not easily reduced to rule."

In other words, the formulation of a numerical probability via the principle of indifference entails *judgment*. But the grounds for such judgment must depend on some understanding of what the possible events are. If the nature of these possibilities is completely unknown, then our uncertainty contains too much ambiguity to support any numerical probabilities. In particular, if all we know is that *any* numerical probability is *conceivable*, we really know nothing at all upon which to form our judgments. So, Keynes effectively rejected Laplace's principle of insufficient reason in the case of complete ignorance about prior probabilities, but endorsed it in the case that there is actual evidence about a set of possible probabilities that is equally balanced.

PROBABILITY AS A SUBJECTIVE ASSESSMENT

Both frequentists and proponents of logical probability were concerned primarily with what we *ought* to believe about an uncertain event or proposition. How could our measure of belief be concordant with the evidence that was available? In this sense, probability was impersonal, based upon some actual or perceived aspect of the real world (observed frequency, evidence, etc.) and a rational inference regarding the implications of this aspect. However, in the 1920s this fundamental presumption began to be called into question.

During the "Roaring Twenties" much received wisdom was being questioned. In the wake of unprecedented chaos occasioned by the Great War, many traditional assumptions and values, along with the social institutions they supported, no longer seemed so secure. Relativity was in the air, from physics to morality. So, it was perhaps inevitable that objective and rational foundations for probability would come under fire.

Another Cambridge Prodigy

In a response to Keynes's *Treatise*, his younger compatriot Frank Plumpton Ramsey (1903–1930) offered a radically new perspective on probability. Ramsey was a multifaceted genius whose life was tragically cut short by complications of surgery. In addition to his influential papers on probability, Ramsey derived an important result in combinatorial mathematics known as Ramsey's theorem. He also made major contributions to mathematical economics, including the Ramsey model and the theory of Ramsey pricing. In philosophy, he authored several original publications and translated the famous *Tractatus Logico Philosophicus* of Ludwig Wittgenstein (1891–1951) from German to English.

In 1926, Ramsey produced an essay on "Truth and Probability." He began by attacking the notion that there could exist the kind of logical relation that Keynes had posited between evidence and belief: "There do not seem to be any such things as the probability relations he describes."[28] Keynes was in fact a colleague, supporter, and admirer of Ramsey; he was apparently not insulted, and wrote a glowing eulogy after Ramsey's shocking demise.

Ramsey's attack was actually more narrowly targeted than is often believed. Ramsey was not disputing all of Keynes's insights regarding probability as it pertained to inductive inference in scientific analysis. However, he took Keynes to task for attempting to locate his theory within the framework of pure logic. Ramsey argued that the *logic* of partial belief ought to be much more limited in scope. Ordinary logic, he observed, dealt with consistency between assumed premises and the conclusions that were entailed by them. Similarly, a logic of partial entailment should delineate only what conclusions were not *inconsistent* with the premises that one held to be true, that is, one's beliefs. From this narrower perspective, any other principles that constrained rational belief, such as standards for scientific inference, would fall outside the province of logic *per se*.

Ramsey also objected to Keynes's notion that some probabilities were unknown, or only partially known. He believed that probability, as a numerical measure of belief, always existed for any individual, but that "beliefs do differ in measurability." He regarded the problem of elucidating these subjective probabilities as a psychological problem. Armed with such an operationally defined measure of the degree of belief, we might be able to decide whether any particular set of beliefs is rational (consistent).

In his essay, Ramsey sets about establishing the desired method by starting from the principle that someone's beliefs can be inferred by observing his actions. He then goes on to develop a sophisticated theory that allows him to define probability in terms of betting behavior, but to evade problems related to the diminishing marginal utility of money. In essence, he posits the existence of some "ultimate goods" as the currency for his proposed bets. Ramsey defines probability in terms of an individual's certainty equivalents for various proposed bets.

For example, suppose I am offered an award of 50 ultimate goods units (UGU). Alternatively, I can receive an amount of 150 UGU if it rains tomorrow or zero if it does not. Then if I am indifferent between these alternatives, it means that my degree of belief that it will rain must be 1/3. Ramsey proves that in order to be internally consistent, my degrees of belief elicited in this manner must satisfy certain rules, and these turn out to be none other than the usual laws of mathematical probability!

> If anyone's mental condition violated these laws, … He could have a book made against him by a cunning better and would then stand to lose in any event.… the laws of probability are laws of consistency, an extension to partial beliefs of formal logic, the logic of consistency.

Here we have a complete separation between probability and the real world. For Ramsey, probability can refer to any logically consistent set of beliefs, however misguided. The truth of these beliefs and their justification on the basis of evidence are irrelevant.

The only restriction imposed is that one's probabilities must not permit others to hedge their bets in such a way as to guarantee them a profit. For example, suppose we consider the situation discussed by Jacob Bernoulli in which the argument for a certain event implied a probability of 2/3, while the argument for the contrary implied a probability of 3/4. Then, if

TABLE 8.2 Example of a "Dutch Book"

	E Happens	E Fails
Bet $400 on E @ 2:1 odds	Win $800	Lose $400
Bet $300 against E @ 3:1 odds	Lose $900	Win $300
Net payoff:	**Lose $100**	**Lose $100**

we interpret these as subjective mathematical probabilities, a Dutch Book could be made against a person holding these beliefs.

Such a person would presumably be willing to *offer* odds of 3:1 that the event *will not* occur (Bet 1), because that would be a fair bet for someone who had a subjective probability of 3/4. Similarly, he would be willing to *accept* odds of 2:1 that the same event *will* occur (Bet 2). Then this discrepancy allows for the possibility of arbitrage (see Table 8.2). Suppose that I wager $400 that the event will occur and accept the 2:1 odds. Then I will gain $800 if the event occurs, and will lose $400 if it does not. However, suppose that I also bet $900 that the event will *not* occur, and offer 3:1 odds. Then I will lose my $900 if the event occurs, but will win $300 if it does not occur. Regardless of the actual outcome, I stand to lose $100.

Ramsey is often cited as the originator of the idea that probability is *always* subjective, or *personal*, in nature. But Ramsey did not reject the notion of a more objective interpretation of probability outside of formal logic. He stated that there is a "lesser logic" of consistency and the "larger logic" of discovery. Within this larger inductive logic, he believed that empirical evidence, and especially observed frequencies, had an important role to play, although he regarded Keynes's attempts to elucidate this role as inadequate.

Ramsey's understanding of probability in the realm of the larger logic seems to recognize the classical concept of a metaphorical lottery: "partial belief involves reference to a hypothetical or ideal frequency ... belief of degree $\frac{m}{n}$ is ... the kind of belief most appropriate to a number of hypothetical occasions otherwise identical in a proportion $\frac{m}{n}$ of which the proposition in question is true." The frequency and subjective interpretations of probability he regards as the "objective and subjective aspects of the same inner meaning."

Ramsey thus appears to recognize that the essential core of probability is the analogy with a lottery. Each occasion is like a single drawing from a collection of identical hypothetical occasions. Once this conception of a metaphorical lottery is accepted, it is understood that observed frequencies as potentially relevant evidence and personal degrees of belief are both "aspects of the same inner meaning."

Subjectivity Italian Style

As I suggested above, during the 1920s the idea of relativity was *au courant*. So we should not be surprised that almost simultaneously with Ramsey's writing about probability, another young genius was independently formulating a remarkably similar point of view. Like Ramsey, Bruno de Finetti (1906–1985) believed that probability prescribed how an individual should act *relative to* his or her own belief structure. But, he went even further. Unlike Ramsey, he did not feel the need to distinguish the lesser logic of consistency from the larger logic of induction.

De Finetti was born in Innsbruck, then part of the Austro-Hungarian Empire, but lived most of his life in Italy. His father was an engineer, and young Bruno was destined and educated to become one as well. However, after this formal training, his interests turned to mathematics, and especially to probability and statistics, in which he eventfully achieved worldwide renown, coupled with considerable controversy. Beginning in the late 1920s, de Finetti published several papers that dealt with the mathematical and philosophical aspects of probability. This work culminated in his masterful 1937 essay on "Foresight, Its Logical Laws, Its Subjective Sources."[29]

Like Ramsey, de Finetti held that the laws of mathematical probability were standards of consistency, or *coherence*. Like Ramsey, he believed that potential betting behavior could provide an operational basis for measuring subjective probabilities. But de Finetti went further. He asserted that the *only* valid way to interpret probability was as a purely subjective degree of belief. He asserted that there exist:

> rather profound psychological reasons which make the exact or approximate agreement that is observed between the opinions of different individuals very natural, but that there are no reasons, rational, positive, or metaphysical, that can give this fact any meaning beyond that of subjective opinions.

De Finetti offered a series of arguments to buttress his contention that all meaningful probability statements are at base subjective. Consider, for example, the widespread convergence of belief regarding the probability that a die will land on 6. We all accept the value of 1/6 because of the symmetrical construction of the die and, perhaps, our experience playing games that employ dice. Like David Hume, his philosophical hero, de Finetti did not concede that this probability is demonstrably connected to the causal structure of reality. He rejected both the frequentists' objective relative frequencies and the logicists' necessary relationship between evidence and rational belief.

Within de Finetti's dogmatically "relativistic" framework, the importance of inverse probability, as originally proposed by Bayes and Laplace, emerged in a new guise. He offered a sophisticated rationale for utilizing Bayes's Theorem as the basis for all practical applications of probability theory, including statistical inference. Recall that Bayes's Theorem allows us, in principle, to make (posterior) probability statements about some quantity such as an unknown probability or the unknown mean value of some variable in a certain population. However, we need two pieces of information: (a) the conditional probability distribution of the data, given each value of the unknown quantity and (b) the prior distribution of the unknown quantity.

De Finetti asserted that *all* probabilities were properly regarded as subjective degrees of belief, so that a prior probability was just a measure of someone's belief. Furthermore, he turned the tables on traditional views by "proving" that *objective* probabilities were, in a sense, illusory. We could, he believed, justify the use of Bayes's Theorem as a statement about the relationship between prior (he called them *initial*) subjective probabilities and posterior (*final*) subjective probabilities without alluding to any objective probabilities at all!

Consider a coin with an unknown probability of Heads (H) that is tossed three times, resulting in HTH. In this classical application of Bayes's Theorem, the unknown probability of H is understood to be a real property of the coin toss. In de Finetti's view, this "objective" probability is a metaphysical abstraction that has no meaning. All that is needed to obtain a valid posterior probability is an assumption that he termed *exchangeability*. Exchangeability is a property not defined in relation to some unobservable equally possible chances, but is directly related to potentially observable events or quantities (random variables).

In our example, there are three possible outcomes in which H occurs exactly twice (HHT, HTH, THH). In this simple case, the various coin tosses would be considered exchangeable (for me) if my subjective probability depended *only* on how many times H and T occur. For example, the three possible outcomes (HHT, HTH, THH) would have (for me) the same probability. In other words, I make no distinction between the particular coin tosses; the only characteristic I consider relevant to the probability of a sequence is how many times H and T each occur. It is completely irrelevant which specific trials result in H and which result in T.

De Finetti showed that under this exchangeability assumption, Bayes's Theorem must be used to obtain final subjective probabilities, given the initial probabilities and observed data. Furthermore, the laws of large numbers derived by Bernoulli, Laplace, de Moivre, and Poisson could essentially be reinterpreted as properties of subjective probabilities. No other definition of probability is necessary to establish these time-honored results. For instance, the approach of the relative frequency in a very long sequence of trials to the "true" probability is meaningless in de Finetti's scheme. It is quite natural to talk *as if* such a true probability exists, but this usage is just a manner of speaking.

De Finetti's logic may seem compelling, but it rests on a critical premise regarding human psychology. Like Ramsey, de Finetti assumed that degrees of belief are well ordered enough to be elicited by observing betting behavior. Keynes might have argued that in most cases our actual beliefs are quite vague and that forcing them into the mold of coherent mathematical probabilities is artificial. De Finetti had faith in our ability to frame subjective numerical probabilities upon which we would be willing to bet. Where would such probabilities come from? How precisely do they measure our belief structures? Are they any less metaphysical than the "objective" probabilities he denigrates?

De Finetti's brilliant concept of exchangeability was intended to exclude metaphysics by dealing only with observable outcomes. However, it seems to contain within it a hidden assumption. A series of exchangeable observations, such as three consecutive tosses of a coin, is characterized by the property that the observations are indistinguishable: the order in which they are observed is irrelevant. This is in fact the assumption of willful ignorance. That is, there is no basis on which to differentiate among the various observations, just as if they were selected from a metaphorical lottery.

Subjectivity and Statistics

For about 20 years, de Finetti's views were not widely known and exerted little practical influence. However, in the 1950s, a brilliant mathematical statistician named Leonard J. (Jimmie) Savage (1917–1971) rediscovered and expanded upon de Finetti's pioneering work. Savage believed that a subjective interpretation of probability was the most suitable foundation upon which to base the theory of statistical inference.

Savage's views were summarized in *The Foundations of Statistics* (1954).[30] This work is often credited with initiating a revival of interest in Bayesian methods within the field of statistics. At the time of the book's publication, Bayes's Theorem was well established as a mathematical truism and widely used for certain types of problems. However, these applications were almost exclusively those in which the prior probability distribution had an obvious frequentist interpretation. For example, in a manufacturing process, the percentage of defective items expected in a given batch might have a probability distribution that was roughly known from historical experience.

On the other hand, in most applications of statistical inference, there was little or no basis upon which such an "empirical" prior distribution could be based. From the dominant frequentist viewpoint, the use of subjective prior distributions was completely anathema. In the late 1960s, when I entered graduate school, Bayesian statistics was regarded almost as a kind of quackery. This attitude has certainly changed over time, but even today the use of Bayesian methods remains somewhat controversial in many scientific applications.

Savage, like de Finetti, contended that all probabilities were essentially subjective, although he preferred the term *personal*. So, the subjective Bayesian approach of starting with a personal prior distribution and then conditioning on observed data was, for him, quite natural. To deny the inevitable subjectivity inherent in statistical analyses, Savage believed, was simply unrealistic. In holding this opinion, however, Savage differed slightly from de Finetti in one important respect. Savage placed great emphasis on the premise that data analysis was a method of reaching optimal *decisions*. Since these decisions were ultimately intended to achieve the "personal" aims of individuals or institutions, he regarded statistical analysis as a branch of decision theory.

In the 1950s, the centrality of decision-making in statistical theory was becoming widely accepted among statisticians of all stripes, especially in America. The potential role of "statistical decision theory" in certain business and engineering applications began to gain traction. However, reliance on subjective probabilities in general, especially in scientific work, was quite limited. Moreover, the extreme mathematical complexity of implementing the necessary Bayesian analysis posed a major obstacle that was not overcome until the advent of modern high-speed computing capabilities.

PROBABILITY AS A PROPENSITY

By the early 1920s, the importance of probability as an essential aspect of our scientific descriptions of physical reality had become increasingly evident. In physics, statistical mechanics, which describes the observable properties (e.g., heat and energy) of enormous aggregates of molecules was well established, and Brownian motion had recently been explained by Einstein in statistical terms. In biology, the pioneering studies of Gregor Mendel (1822–1884) came to light and led to a probabilistic theory of genetic inheritance. Then, studies of radioactive decay showed that this phenomenon followed probabilistic laws very precisely.

Finally, quantum theory suggested that physical reality at the subatomic level seemed to be *intrinsically* probabilistic. Despite the best efforts of Einstein and others to prove that God does not play dice with the universe, randomness could not be explained away. Laplace's fiction of a hypothetical all-knowing intelligence that could predict everything became passé.

These dramatic developments in the sciences prompted philosophers to ponder anew just what probability really means. In what precise sense does an atom of carbon-14 have a 50% probability of decaying within 5730 years? What does it mean to say that the position of an atom can only be described by a probability distribution? The frequency, logical, and subjective interpretations of probability all seemed somehow inadequate to fully express the nature of these phenomena. So, the notion began to emerge that probability, at least in these scientific contexts, might have another aspect, that of a *propensity*.

An Unorthodox Thinker

The concept of probability as a propensity was foreshadowed by the highly original and somewhat quirky American scientist and philosopher, Charles Sanders Peirce (pronounced like “purse”). C.S. Peirce (1839–1914) grew up in Cambridge, Massachusetts. His father, Benjamin Peirce (1809–1880), was an eminent scientist and mathematician, who taught at Harvard University for nearly 50 years. In addition, he played a leading role in the United States Coast Survey, a predecessor of today’s National Oceanic and Atmospheric Administration. In this role, Peirce senior made many practical mathematical and statistical contributions, including an important paper in 1852 on how to deal with anomalous observations, which we would call outliers today.

Benjamin Peirce and his wife, Sarah Mills Peirce, had four sons and a daughter. Charles, who was a child prodigy, may have been the most precocious, but was the black sheep of the family.[31] Unlike his brothers, all of whom followed conventional and highly successful career paths, C.S. Peirce marched to the beat of his own drum. He made many significant contributions to several fields, including mathematics, logic, psychology, biology, and geodesy. However, he may be best known today as a founder of the philosophy of pragmatism, along with William James (1842–1910) and John Dewey (1859–1952). His broad interests and highly original thinking made Peirce a fascinating figure, but his free-spirited antics caused him much personal grief.

For example, a promising academic career was derailed in 1884, when it became known that he had been living with a woman 20 years his junior while still legally married (albeit separated from his wife for 7 years). Johns Hopkins University, where he had been teaching logic and mathematics since 1879, refused to renew his contract and effectively blacklisted him. For much of his later career, he worked for the Coast Survey, thanks initially to his father’s influence. However, it was not always smooth sailing, and Peirce’s independence got him into hot water. In the end, his refusal to comply with a request to revise a major report on gravitational fields led to his dismissal.

The following description by a contemporary suggests how Peirce's unique combination of creativity and eccentricity could both delight and exasperate those around him:

> His dramatic manner, his reckless disregard of accuracy in what he termed 'unimportant details,' his clever newspaper articles … interested and amused us all.… He was always hard up, living partly on what he could borrow from friends, and partly on what he got from odd jobs such as writing book reviews … He was equally brilliant, whether under the influence of liquor or otherwise, and his company was prized by the various organizations to which he belonged; and he was never dropped from any of them even though he was unable to pay his dues. He infuriated Charlotte Angas Scott by contributing to the New York Evening Post an unsigned obituary of Arthur Cayley in which he stated upon no grounds, except that Cayley's father had for a time resided in Russia, that Cayley had inherited his genius from a Russian whom his father had married in St. Petersburg.[32]

Here was another kindred spirit of Gerolamo Cardano.

Peirce was apparently the first to suggest that probability was not just a matter of our ignorance regarding underlying causes. Rather, there was an inherently unpredictable, or random, aspect to reality. He coined the term *tychism* to describe this aspect of his philosophy, which was complex and continually evolving, but always entailed a belief in something akin to free will. At any rate, he clearly rejected the idea that all physical, biological, and psychological phenomena could ultimately be reduced to rigid mechanical laws.

His most famous discussion of probability describes the notion of a predisposition, rather like a personal habit, that exists as a kind of potentiality. As an example he explains what it means for a thrown die to have a probability of 1/3 to result in a 3 or 6:

> The statement means that the die has a certain 'would-be'; and to say that a die has a 'would-be' is to say that it has a property quite analogous to any *habit* that a man might have.… and just as it would be necessary, in order to define a man's habit, to describe how it would lead him to behave and upon what sort of occasion— albeit this statement would by no means imply that the habit *consists* in that action—so to define the die's 'would-be,' it is necessary to say how it would lead the die to behave on an occasion that would bring out the full consequence of the 'would-be'; and this statement would not of itself imply that the 'would-be' of the die consists of this behavior.[33]

There is a lot to ponder here.

A World of Propensities

Peirce's idea of probability lay fallow for half-a-century. Then, in the 1950s it was reincarnated as the *propensity* interpretation of probability by Karl Popper (1902–1994), a giant of twentieth-century philosophy of science. Popper became most famous for his theory of *falsifiability*.[34] He held, essentially, that scientific theories gained credibility only by withstanding challenges that might have proved them false. Thus, the method of building support for a scientific hypothesis was one of "conjecture and refutation."

The stiffer the challenge, the more "corroboration" there would be if the hypothesis could not be dislodged. Thus, Newton's theory of universal gravitation explained a great deal and withstood all challenges for over 200 years. It remained true for most practical purposes even after being dethroned by Einstein's general relativity in 1916. Einstein's theory, in turn, has been the reigning champion for nearly a century and seems likely to endure for a while longer.

Early in his career, Popper subscribed to the frequentist interpretation of probabilities, developing his own particular spin on it. Then, in the early 1950s he became disenchanted with his earlier views, mainly because they could not deal with "singular events." He found the frequency approach unsuitable for expressing the kind of uncertainty about physical events that quantum theory posited to be irreducible, even in principle. To cope with such phenomena, Popper proposed to introduce a modification he termed the "propensity interpretation."

Popper's modification transfers attention from a sequence of observations to a set of generating conditions for such a sequence:

> The frequency interpretation always takes probability as relative to a sequence which is assumed as given; and it works on the assumption that a probability is *a property of some given sequence*. But with our modification, the sequence in its turn is defined by its set of *generating conditions*; and in such a way that probability may now be said to be *a property of the generating conditions*.[35]

Popper's basic idea of probability as propensity continued to evolve within his general philosophical framework and has influenced many others. There are currently many prominent philosophers who include a propensity aspect in their thinking. On the other hand, most practitioners, such as statisticians, engineers, and social scientists, are almost entirely oblivious to these developments. For the most part, only philosophers of science and some thoughtful physicists worry about what the probabilities that play such a prominent role in quantum theory "really" mean.

Popper himself, in his later years, found the notion of propensities increasingly central to his world view. In *A World of Propensities* (1990), he gravitated toward an almost mystical understanding of probability. He saw propensities as physical realities, "as real as forces, or fields of forces." Just as a force field defines a potential for movement of an object, a propensity can frame the possible outcomes of virtually anything.

He saw propensity as proscribing the limits within which free action could occur, and actions in turn changing the propensities:

> In our world, the situation and, with it, the possibilities, and thus the propensities, change all the time. They certainly may change if we, or any other organism, *prefer* one possibility to another; or if we *discover* a new possibility where we have not seen one before. Our very understanding of the world changes the conditions of the changing world; and so do our wishes, our preferences, our motivations, our hopes, our dreams, our phantasies, our hypotheses, our theories. ... All this amounts to the fact that *determinism is simply mistaken*; ... This view of propensities allows us to see in a new light the processes that constitute our world, the world process. The world is no longer a causal machine—it can now be seen as a world of propensities, as an unfolding process of realizing possibilities and of unfolding new possibilities.[36]

The propensity view of probability derives from a philosophy that focuses on the dynamic unfolding of phenomena. Potentialities, predispositions to *become* (Peirce's would-be's), are taken to be fundamental, and fixed things or objects are derivative and secondary.

Like the frequency theory, the propensity interpretation of probability intends to describe an objectively existing reality. For example, purely epistemic probabilities, such as my assessment of who will win a future election or football game, are not considered; the theory is not meant to encompass all possible applications of probability. In particular, if we wish to consider the probability that a past event actually occurred, it is hard to see how the propensity theory would apply.

The primary motivation for the propensity theory was to reconcile the notions of *causation* and *uncertainty*. This need became acute with the advent of quantum physics. No longer could probability be explained away as just a function of ignorance about true underlying deterministic causes. The traditional view of causes as deterministic just does not work in quantum physics, which provides an amazingly precise description of reality. So, if probabilistic uncertainty exists in a pure form at the quantum level, why should intrinsic uncertainty not also exist in some form as an aspect of macro phenomena as well? Peirce would presumably have applauded these developments.

Regardless of the philosophical value of propensities as a conceptual framework, I believe this perspective has much practical value. In the biological, economic, social, and psychological sciences, to say nothing of everyday life, it may be fruitful to think of causation on a continuum. At one extreme, there are universal regularities, causal "laws" in the conventional sense. At the other, there is complete lack of predictability, or apparent randomness. Various observable processes can be regarded as somehow intermediate between these extremes.

The propensity interpretation captures an aspect of the metaphorical lottery model that is left out by the subjective, frequency, and logical theories. These other perspectives all fail to suggest a clear sense of how a singular potential event is somehow being impelled into being under certain conditions. Propensity theory reminds us that this "would-be" may be real, but in a way that challenges traditional views of reality.

CHAPTER 9

PROBABILITY AND REALITY

Uncertainty entails two different aspects, or dimensions. We can be uncertain either because we harbor positive doubt about something or because we interpret the situation to be ambiguous. Probability in the broadest sense relates to our uncertainty generally. However, in the face of substantial ambiguity, quantifying our full uncertainty may be difficult or impossible. Since Bernoulli's day, there have been just a few serious attempts to develop a calculus of uncertainty that accounts for potential ambiguity.

Several noted economists, starting with Keynes and Frank Knight, have called attention to the limitations of standard probabilistic models to deal with the uncertainties of economic systems. In the 1960s, Daniel Ellsberg (yes, *that* Daniel Ellsberg, for those of us old enough to remember the Pentagon Papers affair) raised a stir by describing a "paradox" that challenged standard economic theory. The accepted wisdom among economists was that individuals and companies were rational in the sense of attempting to optimize expected values, as defined by mathematical probability theory. Ellsberg argued that the ambiguity pertaining to a situation ought to be taken into account in decision-making.[1]

The latest in this tradition of highlighting the limits of probabilistic models is Nassim Nicholas Taleb, whose jeremiads have indeed proved

Willful Ignorance: The Mismeasure of Uncertainty, First Edition. Herbert I. Weisberg.

prophetic in light of the financial collapse that began in 2007.[2] Taleb credits the iconoclastic genius Benoit Mandelbrot (1924–2010) as his inspiration. Mandelbrot developed a typology of varieties of randomness and argued that real-world economic phenomena could not be well described by the ordinary limited form of randomness implicit in our usual mathematical models based on the theory of probability.[3]

A few statisticians and mathematicians also began to explore some alternative mathematical versions of probability beginning in the 1960s. They have attempted to account for some types of ambiguity by developing various types of "nonadditive" probability. Art Dempster and Glenn Shafer initiated a theory of *belief functions*, often called Dempster–Shafer theory.[4] These efforts, in which as a graduate student I played a small supporting role, can be considered a revival of the grand designs of Jacob Bernoulli, Johann Lambert, and others over two centuries earlier.[5] In addition, computer scientists have tried to incorporate some consideration of ambiguity by proposing *possibility theory*, an offshoot of "fuzzy logic," introduced by Lofti Zadeh.[6]

So far, none of these unconventional mathematical systems has really caught on. The usual version of mathematical probability has become so entrenched that it has made it difficult to imagine any alternatives. Other proposals, whatever their theoretical merits might be, face an enormous barrier to entry. We have come to equate uncertainty with mathematical probability and ceased to take seriously questions regarding ambiguity that Keynes and others addressed a century ago. As a result, discussions about the relation of probability to the real world have been drastically narrowed. The two main "schools" of probability interpretation today essentially ignore this problem entirely.

The frequency theory forecloses ambiguity by defining probability directly in terms of potentially observable sequences of actual occurrences. The subjective approach sidesteps the issue by assuming that a rational individual is capable of resolving any ambiguity in the process of framing a probability. How he manages to accomplish this feat is a personal matter. Frequentists thus assume that the relationship of probabilities to reality is limited, but direct and obvious; subjectivists assume that this relationship does not matter. Both perspectives fail to grasp the essentially metaphorical nature of mathematical probability.