From "The Drunkard's Walk"

CHAPTER 5

The Dueling Laws of Large and Small Numbers

In their work, Cardano, Galileo, and Pascal assumed that the probabilities relevant to the problems they tackled were known. Galileo, for example, assumed that a die has an equal chance of landing on any of its six faces. But how solid is such "knowledge"? The grand duke's dice were probably designed not to favor any face, but that doesn't mean fairness was actually achieved. Galileo could have tested his assumption by observing a number of tosses and recording how often each face came up. If he had repeated the test several times, however, he would probably have found a slightly different distribution each time, and even small deviations might have mattered, given the tiny differential he was asked to explain. In order to make the early work on randomness applicable to the real world, that issue had to be addressed: What is the connection between underlying probabilities and observed results? What does it mean, from a practical point of view, when we say the chances are 1 in 6 a die will land on 2? If it doesn't mean that in any series of tosses the die will land on the 2 exactly 1 time in 6, then on what do we base our belief that the chances of throwing a 2 really are 1 in 6? And what does it mean when a doctor says that a drug is 70 percent effective or has serious side effects in 1 percent of the cases or when a poll finds that a candidate has support of 36 percent of voters? These are deep questions, related to the very meaning of the concept of randomness, a concept mathematicians still like to debate.

I recently engaged in such a discussion one warm spring day with a statistician visiting from Hebrew University, Moshe, who sat across the lunch table from me at Caltech. Between spoonfuls of nonfat yogurt, Moshe espoused the opinion that truly random numbers do not exist. "There is no such thing," he said. "Oh, they publish charts and write computer programs, but they are just fooling themselves. No one has ever found a method of producing randomness that's any better than throwing a die, and throwing a die just won't do it."

Moshe waved his white plastic spoon at me. He was agitated now. I felt a connection between his feelings about randomness and his religious convictions. Moshe is an Orthodox Jew, and I know that many religious people have problems thinking God can allow randomness to exist. "Suppose you want a string of *N* random numbers between 1 and 6," he told me. "You throw a die *N* times and record the string of *N* numbers that comes up. Is that a random string?"

No, he claimed, because no one can make a perfect die. There will always be some faces that are favored and some that are disfavored. It might take 1,000 throws to notice the difference, or 1 billion, but eventually you will notice it. You'll see more 4s than 6s or maybe fewer. Any artificial device is bound to suffer from that flaw, he said, because human beings do not have access to perfection. That may be, but Nature does, and truly random events do occur on the atomic level. In fact, that is the very basis of quantum theory, and so we spent the rest of our lunch in a discussion of quantum optics.

Today cutting-edge quantum generators produce truly random numbers from the toss of Nature's perfect quantum dice. In the past the perfection necessary for randomness was indeed an elusive goal. One of the most creative approaches came from New York City's Harlem crime syndicates around 1920.[1] Needing a daily supply of five-digit random numbers for an illegal lottery, the racketeers thumbed their noses at the authorities by employing the last five digits of the U.S. Treasury balance. (At this writing the U.S. government is in debt by $8,995,800,515,946.50, or $29,679.02 per person, so today the racketeers could have obtained their five digits from the per capita debt!) Their so-called Treasury lottery ran afoul of not only criminal law, however, but also scientific law, for according to a rule called Benford's law, numbers arising in this cumulative fashion are not random but rather are biased in favor of the lower digits.

Benford's law was discovered not by a fellow named Benford but by the American astronomer Simon Newcomb. Around 1881, Newcomb noticed that the pages of books of logarithms that dealt with numbers beginning with the numeral 1 were dirtier and more frayed than the pages corresponding to numbers beginning with the numeral 2, and so on, down to the numeral 9, whose pages, in comparison, looked clean and new. Assuming that in the long run, wear was proportional to amount of use, Newcomb concluded from his observations that the scientists with whom he shared the book were working with data that reflected that distribution of digits. The law's current name arose after Frank Benford noticed the same thing, in 1938, when scrutinizing the log tables at the General Electric Research Laboratory in Schenectady, New York. But neither man proved the law. That didn't happen until 1995, in work by Ted Hill, a mathematician at the Georgia Institute of Technology.

According to Benford's law, rather than all nine digits' appearing with equal frequency, the number 1 should appear as the first digit in data about 30 percent of the time; the digit 2, about 18 percent of the time; and so on, down to the digit 9, which should appear as the first digit about 5 percent of the time. A similar law, though less pronounced, applies to later digits. Many types of data obey Benford's law, in particular, financial data. In fact, the law seems tailor-made for mining large amounts of financial data in search of fraud.

One famous application involved a young entrepreneur named Kevin Lawrence, who raised $91 million to create a chain of high-tech health clubs.[2] Engorged with cash, Lawrence raced into action, hiring a bevy of executives and spending his investors' money as quickly as he had raised it. That would have been fine except for one detail: he and his cohorts were spending most of the money not on the business but on personal items. And since several homes, twenty personal watercraft, forty-seven cars (including five Hummers, four Ferraris, three Dodge Vipers, two DeTomaso Panteras, and a Lamborghini Diablo), two Rolex watches, a twenty-one-carat diamond bracelet, a $200,000 samurai sword, and a commercial-grade cotton candy machine would have been difficult to explain as necessary business expenditures, Lawrence and his pals tried to cover their tracks by moving investors' money through a complex web of bank accounts and shell companies to give the appearance of a bustling and growing business. Unfortunately for them, a suspicious forensic accountant named Darrell Dorrell compiled a list of over 70,000 numbers representing their various checks and wire transfers and compared the distribution of digits with Benford's law. The numbers failed the test.[3] That, of course, was only the beginning of the investigation, but from there the saga unfolded predictably, ending the day before Thanksgiving 2003, when, flanked by his attorneys and clad in light blue prison garb, Kevin Lawrence was sentenced to twenty years without possibility of parole. The IRS has also studied Benford's law as a way to identify tax cheats. One researcher even applied the law to thirteen years of Bill Clinton's tax returns. They passed the test.[4]

Presumably neither the Harlem syndicate nor its customers noticed these regularities in their lottery numbers. But had people like Newcomb, Benford, or Hill played their lottery, in principle they could have used Benford's law to make favorable bets, earning a nice supplement to their scholar's salary.

In 1947, scientists at the Rand Corporation needed a large table of random digits for a more admirable purpose: to help find approximate solutions to certain mathematical equations employing a technique aptly named the Monte Carlo method. To generate the digits, they employed electronically generated noise, a kind of electronic roulette wheel. Is electronic noise random? That is a question as subtle as the definition of randomness itself.

In 1896 the American philosopher Charles Sanders Peirce wrote that a random sample is one "taken according to a precept or method which, being applied over and over again indefinitely, would in the long run result in the drawing of any one of a set of instances as often as any other set of the same number."[5] That is called the frequency interpretation of randomness. The main alternative to it is called the subjective interpretation. Whereas in the frequency interpretation you judge a sample by the way it turned out, in the subjective interpretation you judge a sample by the way it is produced. According to the subjective interpretation, a number or set of numbers is considered random if we either don't know or cannot predict how the process that produces it will turn out.

The difference between the two interpretations is more nuanced than it may seem. For example, in a perfect world a throw of a die would be random by the first definition but not by the second, since all faces would be equally probable but we could (in a perfect world) employ our exact knowledge of the physical conditions and the laws of physics to determine before each throw exactly how the die will land. In the imperfect real world, however, a throw of a die is random according to the second definition but not the first. That's because, as Moshe pointed out, owing to its imperfections, a die will not land on each face with equal frequency; nevertheless, because of our limitations we have no prior knowledge about any face being favored over any other.

In order to decide whether their table was random, the Rand scientists subjected it to various tests. Upon closer inspection, their system was shown to have biases, just like Moshe's archetypally imperfect dice.[6] The Rand scientists made some refinements to their system but never managed to completely banish the regularities. As Moshe said, complete chaos is ironically a kind of perfection. Still, the Rand numbers proved random enough to be useful, and the company published them in 1955 under the catchy title *A Million Random Digits.*

In their research the Rand scientists ran into a roulette-wheel problem that had been discovered, in some abstract way, almost a century earlier by an Englishman named Joseph Jagger.[7] Jagger was an engineer and a mechanic in a cotton factory in Yorkshire, and so he had an intuitive feel for the capabilities—and the shortcomings—of machinery and one day in 1873 turned his intuition and fertile mind from cotton to cash. How perfectly, he wondered, can the roulette wheels in Monte Carlo really work?

The roulette wheel—invented, at least according to legend, by Blaise Pascal as he was tinkering with an idea for a perpetual-motion machine—is basically a large bowl with partitions (called frets) that are shaped like thin slices of pie. When the wheel is spun, a marble first bounces along the rim of the bowl but eventually comes to rest in one of the compartments, which are numbered 1 through 36, plus 0 (and 00 on American roulette wheels). The bettor's job is simple: to guess in which compartment the marble will land. The existence of roulette wheels is pretty good evidence that legitimate psychics don't exist, for in Monte Carlo if you bet $1 on a compartment and the marble lands there, the house pays you $35 (plus your initial dollar). If psychics really existed, you'd see them in places like that, hooting and dancing and pushing wheelbarrows of cash down the street, and not on Web sites calling themselves Zelda Who Knows All and Sees All and offering twenty-four-hour free online love advice in competition with about 1.2 million other Web psychics (according to Google). For me both the future and, increasingly, the past unfortunately appear obscured by a thick fog. But I do know one thing: my chances of losing at European roulette are 36 out of 37; my chances of winning, 1 out of 37. That means that for every $1 I bet, the casino stands to win $(\frac{36}{37} \times \$1) - (\frac{1}{37} \times \$35)$. That comes to $\frac{1}{37}$ of a dollar, or about 2.7¢. Depending on my state of mind, it's either the price I pay for the enjoyment of watching a little marble bounce around a big shiny wheel or else the price I pay for the opportunity of having lightning strike me (in a good way). At least that is how it is supposed to work.

But does it? Only if the roulette wheels are perfectly balanced, thought Jagger, and he had worked with enough machines to share Moshe's point of view. He was willing to bet they weren't. So he gathered his savings, traveled to Monte Carlo, and hired six assistants, one for each of the casino's six roulette wheels. Every day his assistants observed the wheels, writing down every number that came up in the twelve hours the casino was open. Every night, back in his hotel room, Jagger analyzed the numbers. After six days, he had not detected any bias in five of the wheels, but on the sixth wheel nine numbers came up noticeably more often than the others. And so on the seventh day he headed to the casino and started to bet heavily on the nine favored numbers: 7, 8, 9, 17, 18, 19, 22, 28, and 29.

When the casino shut that night, Jagger was up $70,000. His winnings did not go without notice. Other patrons swarmed his table, tossing down their own cash to get in on a good thing. And casino inspectors were all over him, trying to decipher his system or, better, catch him cheating. By the fourth day of betting, Jagger had amassed $300,000, and the casino's managers were desperate to get rid of the mystery guy, or at least thwart his scheme. One imagines this being accomplished by a burly fellow from Brooklyn. Actually the casino employees did something far more clever.

On the fifth day, Jagger began to lose. His losing, like his winning, was not something you could spot immediately. Both before and after the casino's trick, he would win some and lose some, only now he lost more often than he won instead of the other way around. With the casino's small margin, it would take some pretty diligent betting to drain Jagger's funds, but after four days of sucking in casino money, he wasn't about to let up on the straw. By the time his change of luck deterred him, Jagger had lost half his fortune. One may imagine that by then his mood—not to mention the mood of his hangers-on—was sour. How could his scheme have suddenly failed?

Jagger at last made an astute observation. In the dozens of hours he had spent winning, he had come to notice a tiny scratch on the roulette wheel. This scratch was now absent. Had the casino kindly touched it up so that he could drive them to bankruptcy in style? Jagger guessed not and checked the other roulette wheels. One of them had a scratch. The casino managers had correctly guessed that Jagger's days of success were somehow related to the wheel he was playing, and so overnight they had switched wheels. Jagger relocated and again began to win. Soon he had pumped his winnings past where they had been, to almost half a million.

Unfortunately for Jagger, the casino's managers, finally zeroing in on his scheme, found a new way to thwart him. They decided to move the frets each night after closing, turning them along the wheel so that each day the wheel's imbalance would favor different numbers, numbers unknown to Jagger. Jagger started losing again and finally quit. His gambling career over, he left Monte Carlo with $325,000 in hand, about $5 million in today's dollars. Back home, he left his job at the mill and invested his money in real estate.

It may appear that Jagger's scheme had been a sure thing, but it wasn't. For even a perfectly balanced wheel will not come up on 0, 1, 2, 3, and so on, with exactly equal frequencies, as if the numbers in the lead would politely wait for the laggards to catch up. Instead, some numbers are bound to come up more often than average and others less often. And so even after six days of observations, there remained a chance that Jagger was wrong. The higher frequencies he observed for certain numbers may have arisen by chance and may not have reflected higher probabilities. That means that Jagger, too, had to face the question we raised at the start of this chapter: given a set of underlying probabilities, how closely can you expect your observations of a system to conform to those probabilities? Just as Pascal's work was done in the new climate of (the scientific) revolution, so this question would be answered in the midst of a revolution, this one in mathematics—the invention of calculus.

IN 1680 a great comet sailed through our neighborhood of the solar system, close enough that the tiny fraction of sunlight it reflected was sufficient to make it prominent in the night sky of our own planet. It was in that part of earth's orbit called November that the comet was first spotted, and for months afterward it remained an object of intense scrutiny, its path recorded in great detail. In 1687, Isaac Newton would use these data as an example of his inverse square law of gravity at work. And on one clear night in that parcel of land called Basel, Switzerland, another man destined for greatness was also paying attention. He was a young theologian who, gazing at the bright, hazy light of the comet, realized that it was mathematics, not the church, with which he wanted to occupy his life.[8] With that realization sprouted not just Jakob Bernoulli's own career change but also what would become the greatest family tree in the history of mathematics: in the century and a half between Jakob's birth and 1800 the Bernoulli family produced a great many offspring, about half of whom were gifted, including eight noted mathematicians, and three (Jakob, his younger brother Johann, and Johann's son Daniel) who are today counted as among the greatest mathematicians of all times.

Comets at the time were considered by theologians and the general public alike as a sign of divine anger, and God must have seemed pretty pissed off to create this one—it occupied more than half the visible sky. One preacher called it a "heavenly warning of the Allpowerful and Holy God written and placed before the powerless and unholy children of men." It portended, he wrote, "a noteworthy change in spirit or in worldly matters" for their country or town.[9] Jakob Bernoulli had another point of view. In 1681 he published a pamphlet titled *Newly Discovered Method of How the Path of a Comet or Tailed Star Can Be Reduced to Certain Fundamental Laws, and Its Appearance Predicted.*

Bernoulli had scooped Newton on the comet by six years. At least he would have scooped him had his theory been correct. It wasn't, but claiming publicly that comets follow natural law and not God's whim was a gutsy thing to do, especially given that the prior year—almost fifty years after Galileo's condemnation—the professor of mathematics at the University of Basel, Peter Megerlin, had been roundly attacked by theologians for accepting the Copernican system and had been banned from teaching it at the university. A forbidding schism lay between the mathematician-scientists and the theologians in Basel, and Bernoulli was parking himself squarely on the side of the scientists.

Bernoulli's talent soon brought the embrace of the mathematics community, and when Megerlin died, in late 1686, Bernoulli succeeded him as professor of mathematics. By then Bernoulli was working on problems connected with games of chance. One of his major influences was a Dutch mathematician and scientist, Christiaan

Huygens, who in addition to improving the telescope, being the first to understand Saturn's rings, creating the first pendulum clock (based on Galileo's ideas), and helping to develop the wave theory of light, had written a mathematical primer on probability inspired by the ideas of Pascal and Fermat.

For Bernoulli, Huygens's book was an inspiration. And yet he saw in the theory Huygens presented severe limitations. It might be sufficient for games of chance, but what about aspects of life that are more subjective? How can you assign a definite probability to the credibility of legal testimony? Or to who was the better golfer, Charles I of England or Mary, Queen of Scots? (Both were keen golfers.) Bernoulli believed that for rational decision making to be possible, there must be a reliable and mathematical way to determine probabilities. His view reflected the culture of the times, in which to conduct one's affairs in a manner that was consistent with probabilistic expectation was considered the mark of a reasonable person. But it was not just subjectivity that, in Bernoulli's opinion, limited the old theory of randomness. He also recognized that the theory was not designed for situations of ignorance, in which the probabilities of various outcomes could be defined in principle but in practice were not known. It is the issue I discussed with Moshe and that Jagger had to address: What are the odds that an imperfect die will come up with a 6? What are your chances of contracting the plague? What is the probability that your breastplate can withstand a thrust from your opponent's long sword? In both subjective and uncertain situations, Bernoulli believed it would be "insanity" to expect to have the sort of prior, or a priori, knowledge of probabilities envisioned in Huygens's book.[10]

Bernoulli saw the answer in the same terms that Jagger later would: instead of depending on probabilities being handed to us, we should discern them through observation. Being a mathematician, he sought to make the idea precise. Given that you view a certain number of roulette spins, how closely can you nail down the underlying probabilities, and with what level of confidence? We'll return to those questions in the next chapter, but they are not quite the questions Bernoulli was able to answer. Instead, he answered a closely related question: how well are underlying probabilities reflected in actual results? Bernoulli considered it obvious that we are justified in expecting that as we increase the number of trials, the observed frequencies will reflect—more and more accurately—their underlying probabilities. He certainly wasn't the first to believe that. But he was the first to give the issue a formal treatment, to turn the idea into a proof, and to quantify it, asking how many trials are necessary, and how sure can we be. He was also among the first to appreciate the importance of the new subject of calculus in addressing these issues.

THE YEAR Bernoulli was named professor in Basel proved to be a milestone year in the history of mathematics: it was the year in which Gottfried Leibniz published his revolutionary paper laying out the principles of integral calculus, the complement to his 1684 paper on differential calculus. Newton would publish his own version of the subject in 1687, in his *Philosophiae Naturalis Principia Mathematica*, or *Mathematical Principles of Natural Philosophy*, often referred to simply as *Principia*. These advances would hold the key to Bernoulli's work on randomness.

By the time they published, both Leibniz and Newton had worked on the subject for years, but their almost simultaneous publications begged for controversy over who should be credited for the idea. The great mathematician Karl Pearson (whom we shall encounter again in chapter 8) said that the reputation of mathematicians "stands for posterity largely not on what they did, but on what their contemporaries attributed to them."[11] Perhaps Newton and Leibniz would have agreed with that. In any case neither was above a good fight, and the one that ensued was famously bitter. At the time the outcome was mixed. The Germans and Swiss learned their calculus from Leibniz's work, and the English and many of the French from Newton's. From the modern standpoint there is very little difference between the two, but in the long run Newton's contribution is often emphasized because he appears to have truly had the idea earlier and because in *Principia* he employed his invention in the creation of modern physics, making *Principia* probably the greatest scientific book ever written. Leibniz, though, had developed a better notation, and it is his symbols that are often used in calculus today.

Neither man's publications were easy to follow. In addition to being the greatest book on science, Newton's *Principia* has also been called "one of the most inaccessible books ever written."[12] And Leibniz's work, according to one of Jakob Bernoulli's biographers, was "understood by no one"; it was not only unclear but also full of misprints. Jakob's brother Johann called it "an enigma rather than an explanation."[13] In fact, so incomprehensible were both works that scholars have speculated that both authors might have intentionally made their works difficult to understand to keep amateurs from dabbling. This enigmatic quality was an advantage for Jakob Bernoulli, though, for it did separate the wheat from the chaff, and his intellect fell into the former category. Hence once he had deciphered Leibniz's ideas, he possessed a weapon shared by only a handful of others in the entire world, and with it he could easily solve problems that were exceedingly difficult for others to attempt.

The set of concepts central to both calculus and Bernoulli's work is that of sequence, series, and limit. The term *sequence* means much the same thing to a mathematician as it does to anybody else: an ordered succession of elements, such as points or numbers. A series is simply the sum of a sequence of numbers. And loosely speaking, if the elements of a sequence seem to be heading somewhere—toward a particular endpoint or a particular number—then that is called the limit of the sequence.

Though calculus represents a new sophistication in the understanding of sequences, that idea, like so many others, had already been familiar to the Greeks. In the fifth century B.C., in fact, the Greek philosopher Zeno employed a curious sequence to formulate a paradox that is still debated among college philosophy students today, especially after a few beers. Zeno's paradox goes like this: Suppose a student wishes to step to the door, which is 1 meter away. (We choose a meter here for convenience, but the same argument holds for a mile or any other measure.) Before she arrives there, she first must arrive at the halfway point. But in order to reach the halfway point, she must first arrive halfway to the halfway point—that is, at the one-quarter-way point. And so on, ad infinitum. In other words, in order to reach her destination, she must travel this sequence of distances: ½ meter, ¼ meter, ⅛ meter, ¹⁄₁₆ meter, and so on. Zeno argued that because the sequence goes on forever, she has to traverse an *infinite* number of *finite* distances. That, Zeno said, must take an infinite amount of time. Zeno's conclusion: you can never get anywhere.

Over the centuries, philosophers from Aristotle to Kant have debated this quandary. Diogenes the Cynic took the empirical approach: he simply walked a few steps, then pointed out that things in fact do move. To those of us who aren't students of philosophy, that probably sounds like a pretty good answer. But it wouldn't have impressed Zeno. Zeno was aware of the clash between his logical proof and the evidence of his senses; it's just that, unlike Diogenes, what Zeno trusted was logic. And Zeno wasn't just spinning his wheels. Even Diogenes would have had to admit that his response leaves us facing a puzzling (and, it turns out, deep) question: if our sensory evidence is correct, then what is wrong with Zeno's logic?

Consider the sequence of distances in Zeno's paradox: ½ meter, ¼ meter, ⅛ meter, ¹⁄₁₆ meter, and so on (the increments growing ever smaller). This sequence has an infinite number of terms, so we cannot compute its sum by simply adding them all up. But we can notice that although the number of terms is infinite, those terms get successively smaller. Might there be a finite balance between the endless stream of terms and their endlessly diminishing size? That is precisely the kind of question we can address by employing the concepts of sequence, series, and limit. To see how it works, instead of trying to calculate how far the student went after the entire infinity of Zeno's intervals, let's take one interval at a time. Here are the student's distances after the first few intervals:

After the first interval: ½ meter
After the second interval: ½ meter + ¼ meter = ¾ meter
After the third interval: ½ meter + ¼ meter + ⅛ meter =
⅞ meter
After the fourth interval: ½ meter + ¼ meter + ⅛ meter +
1⁄16 meter = 15⁄16 meter

There is a pattern in these numbers: ½ meter, ¾ meter, ⅞ meter, 15⁄16 meter . . . The denominator is a power of two, and the numerator is one less than the denominator. We might guess from this pattern that after 10 intervals the student would have traveled 1,023⁄1,024 meter; after 20 intervals, 1,048,575⁄1,048,576 meter; and so on. The pattern makes it clear that Zeno is correct that the more intervals we include, the greater the sum of distances we obtain. But Zeno is not correct when he says that the sum is headed for infinity. Instead, the numbers seem to be approaching 1; or as a mathematician would say, 1 meter is the limit of this sequence of distances. That makes sense, because although Zeno chopped her trip into an infinite number of intervals, she had, after all, set out to travel just 1 meter.

Zeno's paradox concerns the amount of time it takes to make the journey, not the distance covered. If the student were forced to take individual steps to cover each of Zeno's intervals, she would indeed be in some time trouble (not to mention her having to overcome the difficulty of taking submillimeter steps)! But if she is allowed to move at constant speed without pausing at Zeno's imaginary checkpoints—and why not?—then the time it takes to travel each of Zeno's intervals is proportional to the distance covered in that interval, and so since the total distance is finite, as is the total time—and fortunately for all of us—motion is possible after all.

Though the modern concept of limits wasn't worked out until long after Zeno's life, and even Bernoulli's—it came in the nineteenth century[14]—it is this concept that informs the spirit of calculus, and it is in this spirit that Jakob Bernoulli attacked the relationship between probabilities and observation. In particular, Bernoulli investigated what happens in the limit of an arbitrarily large number of repeated observations. Toss a (balanced) coin 10 times and you might observe 7 heads, but toss it 1 zillion times and you'll most likely get very near 50 percent. In the 1940s a South African mathematician named John Kerrich decided to test this out in a practical experiment, tossing a coin what must have seemed like 1 zillion times—actually it was 10,000—and recording the results of each toss.[15] You might think Kerrich would have had better things to do, but he was a war prisoner at the time, having had the bad luck of being a visitor in Copenhagen when the Germans invaded Denmark in April 1940. According to Kerrich's data, after 100 throws he had only 44 percent heads, but by the time he reached 10,000, the number was much closer to half: 50.67 percent. How do you quantify this phenomenon? The answer to that question was Bernoulli's accomplishment.

According to the historian and philosopher of science Ian Hacking, Bernoulli's work "came before the public with a brilliant portent of all the things we know about it now; its mathematical profundity, its unbounded practical applications, its squirming duality and its constant invitation for philosophizing. Probability had fully emerged." In Bernoulli's more modest words, his study proved to be one of "novelty, as well as . . . high utility." It was also an effort, Bernoulli wrote, of "grave difficulty."[16] He worked on it for twenty years.

JAKOB BERNOULLI called the high point of his twenty-year effort his "golden theorem." Modern versions of it that differ in their technical nuance go by various names: Bernoulli's theorem, the law of large numbers, and the weak law of large numbers. The phrase *law of large numbers* is employed because, as we've said, Bernoulli's theorem concerns the way results reflect underlying probabilities when we make a large number of observations. But we'll stick with Bernoulli's terminology and call his theorem the golden theorem because we will be discussing it in its original form.[17]

Although Bernoulli's interest lay in real-world applications, some of his favorite examples involved an item not found in most house-holds: an urn filled with colored pebbles. In one scenario, he envisioned the urn holding 3,000 white pebbles and 2,000 black ones, a ratio of 60 percent white to 40 percent black. In this example you conduct a series of blind drawings from the urn "with replacement"—that is, replacing each pebble before drawing the next in order not to alter the 3:2 ratio. The a priori chances of drawing a white pebble are then 3 out of 5, or 60 percent, and so in this example Bernoulli's central question becomes, how strictly should you expect the proportion of white pebbles drawn to hew to the 60 percent ratio, and with what probability?

The urn example is a good one because the same mathematics that describes drawing pebbles from an urn can be employed to describe any series of trials in which each trial has two possible outcomes, as long as those outcomes are random and the trials are independent of each other. Today such trials are called Bernoulli trials, and a series of Bernoulli trials is a Bernoulli process. When a random trial has two possible outcomes, one is often arbitrarily labeled "success" and the other "failure." The labeling is not meant to be literal and sometimes has nothing to do with the everyday meaning of the words—say, in the sense that if you can't wait to read on, this book is a success, and if you are using this book to keep yourself and your sweetheart warm after the logs burned down, it is a failure. Flipping a coin, deciding to vote for candidate A or candidate B, giving birth to a boy or girl, buying or not buying a product, being cured or not being cured, even dying or living are examples of Bernoulli trials. Actions that have multiple outcomes can also be modeled as Bernoulli trials if the question you are asking can be phrased in a way that has a yes or no answer, such as "Did the die land on the number 4?" or "Is there any ice left on the North Pole?" And so, although Bernoulli wrote about pebbles and urns, all his examples apply equally to these and many other analogous situations.

With that understanding we return to the urn, 60 percent of whose pebbles are white. If you draw 100 pebbles from the urn (with replacement), you might find that exactly 60 of them are white, but you might also draw just 50 white pebbles or 59. What are the chances that you will draw between 58 percent and 62 percent white pebbles? What are the chances you'll draw between 59 percent and 61 percent? How much more confident can you be if instead of 100, you draw 1,000 pebbles or 1 million? You can never be 100 percent certain, but can you draw enough pebbles to make the chances 99.9999 percent certain that you will draw, say, between 59.9 percent and 60.1 percent white pebbles? Bernoulli's golden theorem addresses questions such as these.

In order to apply the golden theorem, you must make two choices. First, you must specify your tolerance of error. How near to the underlying proportion of 60 percent are you demanding that your series of trials come? You must choose an interval, such as plus or minus 1 percent or 2 percent or 0.00001 percent. Second, you must specify your tolerance of uncertainty. You can never be 100 percent sure a trial will yield the result you are aiming for, but you can ensure that you will get a satisfactory result 99 times out of 100 or 999 out of 1,000.

The golden theorem tells you that it is always possible to draw enough pebbles to be almost certain that the percentage of white pebbles you draw will be near 60 percent no matter how demanding you want to be in your personal definition of *almost certain* and *near.* It also gives a numerical formula for calculating the number of trials that are "enough," given those definitions.

The first part of the law was a conceptual triumph, and it is the only part that survives in modern versions of the theorem. Concerning the second part—Bernoulli's formula—it is important to understand that although the golden theorem specifies a number of trials that is sufficient to meet your goals of confidence and accuracy, it does not say you can't accomplish those goals with fewer trials. That doesn't affect the first part of the theorem, for which it is enough to know simply that the number of trials specified is finite. But Bernoulli also intended the number given by his formula to be of practical use. Unfortunately, in most practical applications it isn't. For instance, here is a numerical example Bernoulli worked out himself, although I have changed the context: Suppose 60 percent of the voters in Basel support the mayor. How many people must you poll for the chances to be 99.9 percent that you will find the mayor's support to be between 58 percent and 62 percent—that is, for the result to be accurate within plus or minus 2 percent? (Assume, in order to be consistent with Bernoulli, that the people polled are chosen at random, but with replacement. In other words, it is possible that you poll a person more than once.) The answer is 25,550, which in Bernoulli's time was roughly the entire population of Basel. That this number was impractical wasn't lost on Bernoulli. He also knew that accomplished gamblers can intuitively guess their chances of success at a new game based on a sample of far fewer than thousands of trial games.

One reason Bernoulli's numerical estimate was so far from optimal was that his proof was based on many approximations. Another reason was that he chose 99.9 percent as his standard of certainty—that is, he required that he get the wrong answer (an answer that differed more than 2 percent from the true one) less than 1 time in 1,000. That is a very demanding standard. Bernoulli called it moral certainty, meaning the degree of certainty he thought a reasonable person would require in order to make a rational decision. It is perhaps a measure of how much the times have changed that today we've abandoned the notion of moral certainty in favor of the one we encountered in the last chapter, statistical significance, meaning that your answer will be wrong less than 1 time in 20.

With today's mathematical methods, statisticians have shown that in a poll like the one I described, you can achieve a statistically significant result with an accuracy of plus or minus 5 percent by polling only 370 subjects. And if you poll 1,000, you can achieve a 90 percent chance of coming within 2 percent of the true result (60 percent approval of Basel's mayor). But despite its limitations, Bernoulli's golden theorem was a milestone because it showed, at least in principle, that a large enough sample will almost certainly reflect the underlying makeup of the population being sampled.

. . .

In real life we don't often get to observe anyone's or anything's performance over thousands of trials. And so if Bernoulli required an overly strict standard of certainty, in real-life situations we often make the opposite error: we assume that a sample or a series of trials is representative of the underlying situation when it is actually far too small to be reliable. For instance, if you polled exactly 5 residents of Basel in Bernoulli's day, a calculation like the ones we discussed in chapter 4 shows that the chances are only about 1 in 3 that you will find that 60 percent of the sample (3 people) supported the mayor.

Only 1 in 3? Shouldn't the true percentage of the mayor's supporters be the *most probable* outcome when you poll a sample of voters? In fact, 1 in 3 *is* the most probable outcome: the odds of finding 0, 1, 2, 4, or 5 supporters are lower than the odds of finding 3. Nevertheless, finding 3 supporters is not likely: because there are so many of those nonrepresentative possibilities, their combined odds add up to twice the odds that your poll accurately reflects the population. And so in a poll of 5 voters, 2 times out of 3 you will observe the "wrong" percentage. In fact, about 1 in 10 times you'll find that all the voters you polled agree on whether they like or dislike the mayor. And so if you paid any attention to a sample of 5, you'd probably severely over- or underestimate the mayor's true popularity.

The misconception—or the mistaken intuition—that a small sample accurately reflects underlying probabilities is so widespread that Kahneman and Tversky gave it a name: the law of small numbers.[18] The law of small numbers is not really a law. It is a sarcastic name describing the misguided attempt to apply the law of large numbers when the numbers aren't large.

If people applied the (untrue) law of small numbers only to urns, there wouldn't be much impact, but as we've said, many events in life are Bernoulli processes, and so our intuition often leads us to misinterpret what we observe. That is why, as I described in chapter 1, when people observe the handful of more successful or less successful years achieved by the Sherry Lansings and Mark Cantons of the world, they assume that their past performance accurately predicts their future performance.

Let's apply these ideas to an example I mentioned briefly in chapter 4: the situation in which two companies compete head-to-head or two employees within a company compete. Think now of the CEOs of the Fortune 500 companies. Let's assume that, based on their knowledge and abilities, each CEO has a certain probability of success each year (however his or her company may define that). And to make things simple, let's assume that for these CEOs successful years occur with the same frequency as the white pebbles or the mayor's supporters: 60 percent. (Whether the true number is a little higher or a little lower doesn't affect the thrust of this argument.) Does that mean we should expect, in a given five-year period, that a CEO will have precisely three good years?

No. As the earlier analysis showed, even if the CEOs all have a nice cut-and-dried 60 percent success rate, the chances that in a given five-year period a particular CEO's performance will reflect that underlying rate are only 1 in 3! Translated to the Fortune 500, that means that over the past five years about 333 of the CEOs would have exhibited performance that did not reflect their true ability. Moreover, we should expect, by chance alone, about 1 in 10 of the CEOs to have five winning or losing years in a row. What does this tell us? It is more reliable to judge people by analyzing their abilities than by glancing at the scoreboard. Or as Bernoulli put it, "One should not appraise human action on the basis of its results."[19]

Going against the law of small numbers requires character. For while anyone can sit back and point to the bottom line as justification, assessing instead a person's actual knowledge and actual ability takes confidence, thought, good judgment, and, well, guts. You can't just stand up in a meeting with your colleagues and yell, "Don't fire her. She was just on the wrong end of a Bernoulli series." Nor is it likely to win you friends if you stand up and say of the gloating fellow who just sold more Toyota Camrys than anyone else in the history of the dealership, "It was just a random fluctuation." And so it rarely happens. Executives' winning years are attributed to their brilliance, explained retroactively through incisive hindsight. And when people don't succeed, we often assume the failure accurately reflects the proportion with which their talents and their abilities fill the urn.

Another mistaken notion connected with the law of large numbers is the idea that an event is more or less likely to occur because it has or has not happened recently. The idea that the odds of an event with a fixed probability increase or decrease depending on recent occurrences of the event is called the gambler's fallacy. For example, if Kerrich landed, say, 44 heads in the first 100 tosses, the coin would not develop a bias toward tails in order to catch up! That's what is at the root of such ideas as "her luck has run out" and "He is due." That does not happen. For what it's worth, a good streak doesn't jinx you, and a bad one, unfortunately, does not mean better luck is in store. Still, the gambler's fallacy affects more people than you might think, if not on a conscious level then on an unconscious one. People expect good luck to follow bad luck, or they worry that bad will follow good.

I remember, on a cruise a few years back, watching an intense pudgy man sweating as he frantically fed dollars into a slot machine as fast as it would take them. His companion, seeing me eye them, remarked simply, "He is due." Although tempted to point out that, *no, he isn't due,* I instead walked on. After several steps I halted my progress owing to a sudden flashing of lights, ringing of bells, not a little hooting on the couple's part, and the sound of, for what seemed like minutes, a fast stream of dollar coins flying out of the machine's chute. Now I know that a modern slot machine is computerized, its payoffs driven by a random-number generator, which by both law and regulation must truly generate, as advertised, random numbers, making each pull of the handle completely independent of the history of previous pulls. And yet . . . Well, let's just say the gambler's fallacy is a powerful illusion.

THE MANUSCRIPT in which Bernoulli presented his golden theorem ends abruptly even though he promises earlier in the work that he will provide applications to various issues in civic affairs and economics. It is as if "Bernoulli literally quit when he saw the number 25,550," wrote the historian of statistics Stephen Stigler.[20] In fact, Bernoulli was in the process of publishing his manuscript when he died "of a slow fever" in August 1705, at the age of fifty. His publishers asked Johann Bernoulli to complete it, but Johann refused, saying he was too busy. That may appear odd, but the Bernoullis were an odd family. If you were asked to choose the most unpleasant mathematician who ever lived, you wouldn't be too far off if you fingered Johann Bernoulli. He has been variously described in historical texts as jealous, vain, thin-skinned, stubborn, bilious, boastful, dishonest, and a consummate liar. He accomplished much in mathematics, but he is also known for having his son Daniel tossed out of the Académie des Sciences after Daniel won a prize for which Johann himself had competed, for attempting to steal both his brother's and Leibniz's ideas, and for plagiarizing Daniel's book on hydrodynamics and then faking the publication date so that his book would appear to have been published first.

When he was asked to complete his late brother's manuscript, he had recently relocated to Basel from the University of Groningen, in the Netherlands, obtaining a post not in mathematics but as a professor of Greek. Jakob had found this career change suspicious, especially since in his estimation Johann did not know Greek. What Jakob suspected, he wrote Leibniz, was that Johann had come to Basel to usurp Jakob's position. And, indeed, upon Jakob's death, Johann did obtain it.

Johann and Jakob had not gotten along for most of their adult lives. They would regularly trade insults in mathematics publications and in letters that, one mathematician wrote, "bristle with strong language that is usually reserved for horse thieves."[21] And so when the need arose to edit Jakob's posthumous manuscript, the task fell further down the food chain, to Jakob's nephew Nikolaus, the son of one of Jakob's other brothers, also named Nikolaus. The younger Nikolaus was only eighteen at the time, but he had been one of Jakob's pupils. Unfortunately he didn't feel up to the task, possibly in part because he was aware of Leibniz's opposition to his uncle's ideas about applications of the theory. And so the manuscript lay dormant for eight years. The book was finally published in 1713 under the title *Ars conjectandi*, or *The Art of Conjecture.* Like Pascal's *Pensées,* it is still in print.

Jakob Bernoulli had shown that through mathematical analysis one could learn how the inner hidden probabilities that underlie natural systems are reflected in the data those systems produce. As for the question that Bernoulli did not answer—the question of how to infer, from the data produced, the underlying probability of events—the answer would not come for several decades more.

CHAPTER 6

False Positives and Positive Fallacies

IN THE 1970S a psychology professor at Harvard had an odd-looking middle-aged student in his class. After the first few class meetings the student approached the professor to explain why he had enrolled in the class.[1] In my experience teaching, though I have had some polite students come up to me to explain why they were dropping my course, I have never had a student feel the need to explain why he was taking it. That's probably why I can get away with happily assuming that if asked, such a student would respond, "Because I am fascinated by the subject, and you are a fine lecturer." But this student had other reasons. He said he needed help because strange things were happening to him: his wife spoke the words he was thinking before he could say them, and now she was divorcing him; a co-worker casually mentioned layoffs over drinks, and two days later the student lost his job. Over time, he reported, he had experienced dozens of misfortunes and what he considered to be disturbing coincidences.

At first the happenings confused him. Then, as most of us would, he formed a mental model to reconcile the events with the way he believed the world behaves. The theory he came up with, however, was unlike anything most of us would devise: he was the subject of an elaborate secret scientific experiment. He believed the experiment was staged by a large group of conspirators led by the famous psychologist B. F. Skinner. He also believed that when it was over, he would become famous and perhaps be elected to a high public office. That, he said, was why he was taking the course. He wanted to learn how to test his hypothesis in light of the many instances of evidence he had accumulated.

A few months after the course had run its course, the student again called on the professor. The experiment was still in progress, he reported, and now he was suing his former employer, who had produced a psychiatrist willing to testify that he suffered from paranoia.

One of the paranoid delusions the former employer's psychiatrist pointed to was the student's alleged invention of a fictitious eighteenth-century minister. In particular, the psychiatrist scoffed at the student's claim that this minister was an amateur mathematician who had created in his spare moments a bizarre theory of probability. The minister's name, according to the student, was Thomas Bayes. His theory, the student asserted, described how to assess the chances that some event would occur if some other event also occurred. What are the chances that a particular student would be the subject of a vast secret conspiracy of experimental psychologists? Admittedly not huge. But what if one's wife speaks one's thoughts before one can utter them *and* co-workers foretell your professional fate over drinks in casual conversation? The student claimed that Bayes's theory showed how you should alter your initial estimation in light of that new evidence. And he presented the court with a mumbo jumbo of formulas and calculations regarding his hypothesis, concluding that the additional evidence meant that the probability was 999,999 in 1 million that he was right about the conspiracy. The enemy psychiatrist claimed that this mathematician-minister and his theory were figments of the student's schizophrenic imagination.

The student asked the professor to help him refute that claim. The professor agreed. He had good reason, for Thomas Bayes, born in London in 1701, really was a minister, with a parish at Tunbridge Wells. He died in 1761 and was buried in a park in London called Bunhill Fields, in the same grave as his father, Joshua, also a minister. And he indeed did invent a theory of "conditional probability" to show how the theory of probability can be extended from independent events to events whose outcomes are connected. For example, the probability that a randomly chosen person is mentally ill and the probability that a randomly chosen person believes his spouse can read his mind are both very low, but the probability that a person is mentally ill *if* he believes his spouse can read his mind is much higher, as is the probability that a person believes his spouse can read his mind *if* he is mentally ill. How are all these probabilities related? That question is the subject of conditional probability.

The professor supplied a deposition explaining Bayes's existence and his theory, though not supporting the specific and dubious calculations that his former student claimed proved his sanity. The sad part of this story is not just the middle-aged schizophrenic himself, but the medical and legal team on the other side. It is unfortunate that some people suffer from schizophrenia, but even though drugs can help to mediate the illness, they cannot battle ignorance. And ignorance of the ideas of Thomas Bayes, as we shall see, resides at the heart of many serious mistakes in both medical diagnosis and legal judgment. It is an ignorance that is rarely addressed during a doctor's or a lawyer's professional training.

We also make Bayesian judgments in our daily lives. A film tells the story of an attorney who has a great job, a charming wife, and a wonderful family. He loves his wife and daughter, but still he feels that something is missing in his life. One night as he returns home on the train he spots a beautiful woman gazing with a pensive expression out the window of a dance studio. He looks for her again the next night, and the night after that. Each night as his train passes her studio, he falls further under her spell. Finally one evening he impulsively rushes off the train and signs up for dance lessons, hoping to meet the woman. He finds that her haunting attraction withers once his gaze from afar gives way to face-to-face encounters. He does fall in love, however, not with her but with dancing.

He keeps his new obsession from his family and colleagues, making excuses for spending more and more evenings away from home.

His wife eventually discovers that he is not working late as often as he says he is. She figures the chances of his lying about his after-work activities are far greater if he is having an affair than if he isn't, and so she concludes that he is. But the wife was mistaken not just in her conclusion but in her reasoning: she confused the probability that her husband would sneak around *if* he were having an affair with the probability that he was having an affair *if* he was sneaking around.

It's a common mistake. Say your boss has been taking longer than usual to respond to your e-mails. Many people would take that as a sign that their star is falling because *if* your star is falling, the chances are high that your boss will respond to your e-mails more slowly than before. But your boss might be slower in responding because she is unusually busy or her mother is ill. And so the chances that your star is falling *if* she is taking longer to respond are much lower than the chances that your boss will respond more slowly *if* your star is falling. The appeal of many conspiracy theories depends on the misunderstanding of this logic. That is, it depends on confusing the probability that a series of events would happen *if* it were the product of a huge conspiracy with the probability that a huge conspiracy exists *if* a series of events occurs.

The effect on the probability that an event will occur *if* or *given that* other events occur is what Bayes's theory is all about. To see in detail how it works, we'll turn to another problem, one that is related to the two-daughter problem we encountered in chapter 3. Let us now suppose that a distant cousin has two children. Recall that in the two-daughter problem you know that one or both are girls, and you are trying to remember which it is—one or both? In a family with two children, what are the chances, if one of the children is a girl, that both children are girls? We didn't discuss the question in those terms in chapter 3, but the *if* makes this a problem in conditional probability. If that *if* clause were not present, the chances that both children were girls would be 1 in 4, the 4 possible birth orders being (boy, boy), (boy, girl), (girl, boy), and (girl, girl). But given the additional information that the family has a girl, the chances are 1 in 3. That is because if one of the children is a girl, there are just 3 possible scenarios for this family—(boy, girl), (girl, boy), and (girl, girl)—and exactly 1 of the 3 corresponds to the outcome that both children are girls. That's probably the simplest way to look at Bayes's ideas—they are just a matter of accounting. First write down the sample space—that is, the list of all the possibilities—along with their probabilities if they are not all equal (that is actually a good idea in analyzing any confusing probability issue). Next, cross off the possibilities that the condition (in this case, "at least one girl") eliminates. What is left are the remaining possibilities and their relative probabilities.

That might all seem obvious. Feeling cocky, you may think you could have figured it out without the help of dear Reverend Bayes and vow to grab a different book to read the next time you step into the bathtub. So before we proceed, let's try a slight variant on the two-daughter problem, one whose resolution may be a bit more shocking.[2]

The variant is this: in a family with two children, what are the chances, if one of the children is a girl named Florida, that both children are girls? Yes, I said a girl named Florida. The name might sound random, but it is not, for in addition to being the name of a state known for Cuban immigrants, oranges, and old people who traded their large homes up north for the joys of palm trees and organized bingo, it is a real name. In fact, it was in the top 1,000 female American names for the first thirty or so years of the last century. I picked it rather carefully, because part of the riddle is the question, what, if anything, about the name Florida affects the odds? But I am getting ahead of myself. Before we move on, please consider this question: in the girl-named-Florida problem, are the chances of two girls still 1 in 3 (as they are in the two-daughter problem)?

I will shortly show that the answer is no. The fact that one of the girls is named Florida changes the chances to 1 in 2: Don't worry if that is difficult to imagine. The key to understanding randomness and all of mathematics is not being able to intuit the answer to every problem immediately but merely having the tools to figure out the answer.

. . .

THOSE WHO DOUBTED Bayes's existence were right about one thing: he never published a single scientific paper. We know little of his life, but he probably pursued his work for his own pleasure and did not feel much need to communicate it. In that and other respects he and Jakob Bernoulli were opposites. For Bernoulli resisted the study of theology, whereas Bayes embraced it. And Bernoulli sought fame, whereas Bayes showed no interest in it. Finally, Bernoulli's theorem concerns how many heads to expect if, say, you plan to conduct many tosses of a balanced coin, whereas Bayes investigated Bernoulli's original goal, the issue of how certain you can be that a coin is balanced if you observe a certain number of heads.

The theory for which Bayes is known today came to light on December 23, 1763, when another chaplain and mathematician, Richard Price, read a paper to the Royal Society, Britain's national academy of science. The paper, by Bayes, was titled "An Essay toward Solving a Problem in the Doctrine of Chances" and was published in the Royal Society's *Philosophical Transactions* in 1764. Bayes had left Price the article in his will, along with £100. Referring to Price as "I suppose a preacher at Newington Green," Bayes died four months after writing his will.[3]

Despite Bayes's casual reference, Richard Price was not just another obscure preacher. He was a well-known advocate of freedom of religion, a friend of Benjamin Franklin's, a man entrusted by Adam Smith to critique parts of a draft of *The Wealth of Nations*, and a well-known mathematician. He is also credited with founding actuary science, a field he developed when, in 1765, three men from an insurance company, the Equitable Society, requested his assistance. Six years after that encounter he published his work in a book titled *Observations on Reversionary Payments*. Though the book served as a bible for actuaries well into the nineteenth century, because of some poor data and estimation methods, he appears to have underestimated life expectancies. The resulting inflated life insurance premiums enriched his pals at the Equitable Society. The hapless British government, on the other hand, based annuity payments on Price's tables and took a bath when the pensioners did not proceed to keel over at the predicted rate.

As I mentioned, Bayes developed conditional probability in an attempt to answer the same question that inspired Bernoulli: how can we infer underlying probability from observation? *If* a drug just cured 45 out of 60 patients in a clinical trial, what does that tell you about the chances the drug will work on the next patient? *If* it worked for 600,000 out of 1 million patients, the odds are obviously good that its chances of working are close to 60 percent. But what can you conclude from a smaller trial? Bayes also asked another question: if, before the trial, you had reason to believe that the drug was only 50 percent effective, how much weight should the new data carry in your future assessments? Most of our life experiences are like that: we observe a relatively small sample of outcomes, from which we infer information and make judgments about the qualities that produced those outcomes. How should we make those inferences?

Bayes approached the problem via a metaphor.[4] Imagine we are supplied with a square table and two balls. We roll the first ball onto the table in a manner that makes it equally probable that the ball will come to rest at any point. Our job is to determine, without looking, where along the left-right axis the ball stopped. Our tool in this is the second ball, which we may repeatedly roll onto the table in the same manner as the first. With each roll a collaborator notes whether that ball comes to rest to the right or the left of the place where the first ball landed. At the end he informs us of the total number of times the second ball landed in each of the two general locations. The first ball represents the unknown that we wish to gain information about, and the second ball represents the evidence we manage to obtain. If the second ball lands consistently to the right of the first, we can be pretty confident that the first ball rests toward the far left side of the table. If it lands less consistently to the right, we might be less confident of that conclusion, or we might guess that the first ball is situated farther to the right. Bayes showed how to determine, based on the data of the second ball, the precise probability that the first ball is at any given point on the left-right axis. And he showed how, given additional data, one should revise one's initial estimate. In Bayesian terminology the initial estimates are called prior probabilities and the new guesses, posterior probabilities.

Bayes concocted this game because it models many of the decisions we make in life. In the drug-trial example the position of the first ball represents the drug's true effectiveness, and the reports regarding the second ball represent the patient data. The position of the first ball could also represent a film's appeal, product quality, driving skill, hard work, stubbornness, talent, ability, or whatever it is that determines the success or failure of a certain endeavor. The reports on the second ball would then represent our observations or the data we collect. Bayes's theory shows how to make assessments and then adjust them in the face of new data.

Today Bayesian analysis is widely employed throughout science and industry. For instance, models employed to determine car insurance rates include a mathematical function describing, per unit of driving time, your personal probability of having zero, one, or more accidents. Consider, for our purposes, a simplified model that places everyone in one of two categories: high risk, which includes drivers who average at least one accident each year, and low risk, which includes drivers who average less than one. If, when you apply for insurance, you have a driving record that stretches back twenty years without an accident or one that goes back twenty years with thirty-seven accidents, the insurance company can be pretty sure which category to place you in. But if you are a new driver, should you be classified as low risk (a kid who obeys the speed limit and volunteers to be the designated driver) or high risk (a kid who races down Main Street swigging from a half-empty $2 bottle of Boone's Farm apple wine)? Since the company has no data on you—no idea of the "position of the first ball"—it might assign you an equal prior probability of being in either group, or it might use what it knows about the general population of new drivers and start you off by guessing that the chances you are a high risk are, say, 1 in 3. In that case the company would model you as a hybrid—one-third high risk and two-thirds low risk—and charge you one-third the price it charges high-risk drivers plus two-thirds the price it charges low-risk drivers. Then, after a year of observation—that is, after one of Bayes's second balls has been thrown—the company can employ the new datum to reevaluate its model, adjust the one-third and two-third proportions it previously assigned, and recalculate what it ought to charge. If you have had no accidents, the proportion of low risk and low price it assigns you will increase; if you have had two accidents, it will decrease. The precise size of the adjustment is given by Bayes's theory. In the same manner the insurance company can periodically adjust its assessments in later years to reflect the fact that you were accident-free or that you twice had an accident while driving the wrong way down a one-way street, holding a cell phone with your left hand and a doughnut with your right. That is why insurance companies can give out "good driver" discounts: the absence of accidents elevates the posterior probability that a driver belongs in a low-risk group.

Obviously many of the details of Bayes's theory are rather complex. But as I mentioned when I analyzed the two-daughter problem, the key to his approach is to use new information to prune the sample space and adjust probabilities accordingly. In the two-daughter problem the sample space was initially (boy, boy), (boy, girl), (girl, boy), and (girl, girl) but reduces to (boy, girl), (girl, boy), and (girl, girl) *if* you learn that one of the children is a girl, making the chances of a two-girl family 1 in 3. Let's apply that same simple strategy to see what happens if you learn that one of the children is a girl named Florida.

In the girl-named-Florida problem our information concerns not just the gender of the children, but also, for the girls, the name. Since our original sample space should be a list of all the possibilities, in this case it is a list of both gender and name. Denoting "girl-named-Florida" by girl-F and "girl-not-named-Florida" by girl-NF, we write the sample space this way: (boy, boy), (boy, girl-F), (boy, girl-NF), (girl-F, boy), (girl-NF, boy), (girl-NF, girl-F), (girl-F, girl-NF), (girl-NF, girl-NF), and (girl-F, girl-F).

Now, the pruning. Since we know that one of the children is a girl named Florida, we can reduce the sample space to (boy, girl-F), (girl-F, boy), (girl-NF, girl-F), (girl-F, girl-NF), and (girl-F, girl-F). That brings us to another way in which this problem differs from the two-daughter problem. Here, because it is not equally probable that a girl's name is or is not Florida, not all the elements of the sample space are equally probable.

In 1935, the last year for which the Social Security Administration provided statistics on the name, about 1 in 30,000 girls were christened Florida.[5] Since the name has been dying out, for the sake of argument let's say that today the probability of a girl's being named Florida is 1 in 1 million. That means that if we learn that a particular girl's name is not Florida, it's no big deal, but if we learn that a particular girl's name is Florida, in a sense we've hit the jackpot. The chances of both girls' being named Florida (even if we ignore the fact that parents tend to shy away from giving their children identical names) are therefore so small we are justified in ignoring that possibility. That leaves us with just (boy, girl-F), (girl-F, boy), (girl-NF, girl-F), and (girl-F, girl-NF), which are, to a very good approximation, equally likely.

Since 2 of the 4, or half, of the elements in the sample space are families with two girls, the answer is not 1 in 3—as it was in the two-daughter problem—but 1 in 2. The added information—your knowledge of the girl's name—makes a difference.

One way to understand this, if it still seems puzzling, is to imagine that we gather into a very large room 75 million families that have two children, at least one of whom is a girl. As the two-daughter problem taught us, there will be about 25 million two-girl families in that room and 50 million one-girl families (25 million in which the girl is the older child and an equal number in which she is the younger). Next comes the pruning: we ask that only the families that include a girl named Florida remain. Since Florida is a 1 in 1 million name, about 50 of the 50 million one-girl families will remain. And of the 25 million two-girl families, 50 of them will also get to stay, 25 because their firstborn is named Florida and another 25 because their younger girl has that name. It's as if the girls are lottery tickets and the girls named Florida are the winning tickets. Although there are twice as many one-girl families as two-girl families, the two-girl families each have two tickets, so the one-girl families and the two-girl families will be about equally represented among the winners.

I have described the girl-named-Florida problem in potentially annoying detail, the kind of detail that sometimes lands me on the do-not-invite list for my neighbors' parties. I did this not because I expect you to run into this situation. I did it because the context is simple, and the same kind of reasoning will bring clarity to many situations that really are encountered in life. Now let's talk about a few of those.

MY MOST MEMORABLE ENCOUNTER with the Reverend Bayes came one Friday afternoon in 1989, when my doctor told me by telephone that the chances were 999 out of 1,000 that I'd be dead within a decade. He added, "I'm *really* sorry," as if he had some patients to whom he would say he is sorry but not mean it. Then he answered a few questions about the course of the disease and hung up, presumably to offer another patient his or her Friday-afternoon news flash. It is hard to describe or even remember exactly how the weekend went for me, but let's just say I did not go to Disneyland. Given my death sentence, why am I still here, able to write about it?

The adventure started when my wife and I applied for life insurance. The application procedure involved a blood test. A week or two later we were turned down. The ever economical insurance company sent the news in two brief letters that were identical, except for a single additional word in the letter to my wife. My letter stated that the company was denying me insurance because of the "results of your blood test." My wife's letter stated that the company was turning her down because of the "results of your husband's blood test." When the added word *husband's* proved to be the extent of the clues the kindhearted insurance company was willing to provide about our uninsurability, I went to my doctor on a hunch and took an HIV test. It came back positive. Though I was too shocked initially to quiz him about the odds he quoted, I later learned that he had derived my 1-in-1,000 chance of being healthy from the following statistic: the HIV test produced a positive result when the blood was not infected with the AIDS virus in only 1 in 1,000 blood samples. That might sound like the same message he passed on, but it wasn't. My doctor had confused the chances that I would test positive *if* I was not HIV-positive with the chances that I would not be HIV-positive *if* I tested positive.

To understand my doctor's error, let's employ Bayes's method. The first step is to define the sample space. We could include everyone who has ever taken an HIV test, but we'll get a more accurate result if we employ a bit of additional relevant information about me and consider only heterosexual non-IV-drug-abusing white male Americans who have taken the test. (We'll see later what kind of difference this makes.)

Now that we know whom to include in the sample space, let's classify the members of the space. Instead of boy and girl, here the relevant classes are those who tested positive and are HIV-positive (true positives), those who tested positive but are not positive (false positives), those who tested negative and are HIV-negative (true negatives), and those who tested negative but are HIV-positive (false negatives).

Finally, we ask, how many people are there in each of these classes? Suppose we consider an initial population of 10,000. We can estimate, employing statistics from the Centers for Disease Control and Prevention, that in 1989 about 1 in those 10,000 heterosexual non-IV-drug-abusing white male Americans who got tested were infected with HIV.[6] Assuming that the false-negative rate is near 0, that means that about 1 person out of every 10,000 will test positive due to the presence of the infection. In addition, since the rate of false positives is, as my doctor had quoted, 1 in 1,000, there will be about 10 others who are not infected with HIV but will test positive anyway. The other 9,989 of the 10,000 men in the sample space will test negative.

Now let's prune the sample space to include only those who tested positive. We end up with 10 people who are false positives and 1 true positive. In other words, only 1 in 11 people who test positive are really infected with HIV. My doctor told me that the probability that the test was wrong—and I was in fact healthy—was 1 in 1,000. He should have said, "Don't worry, the chances are better than 10 out of 11 that you are not infected." In my case the screening test was apparently fooled by certain markers that were present in my blood even though the virus this test was screening for was not present.

It is important to know the false positive rate when assessing any diagnostic test. For example, a test that identifies 99 percent of all malignant tumors sounds very impressive, but I can easily devise a test that identifies 100 percent of all tumors. All I have to do is report that everyone I examine has a tumor. The key statistic that differentiates my test from a useful one is that my test would produce a high rate of false positives. But the above incident illustrates that knowledge of the false positive rate is not sufficient to determine the usefulness of a test—you must also know how the false-positive rate compares with the true prevalence of the disease. If the disease is rare, even a low false-positive rate does not mean that a positive test implies you have the disease. If a disease is common, a positive result is much more likely to be meaningful. To see how the true prevalence affects the implications of a positive test, let's suppose now that I had been homosexual and tested positive. Assume that in the male gay community the chance of infection among those being tested in 1989 was about 1 percent. That means that in the results of 10,000 tests, we would find not 1 (as before), but 100 true positives to go with the 10 false positives. So in this case the chances that a positive test meant I was infected would have been 10 out of 11. That's why, when assessing test results, it is good to know whether you are in a high-risk group.

. . .

BAYES'S THEORY shows that the probability that A will occur if B occurs will generally differ from the probability that B will occur if A occurs.[7] To not account for this is a common mistake in the medical profession. For instance, in studies in Germany and the United States, researchers asked physicians to estimate the probability that an asymptomatic woman between the ages of 40 and 50 who has a positive mammogram actually has breast cancer if 7 percent of mammograms show cancer when there is none.[8] In addition, the doctors were told that the actual incidence was about 0.8 percent and that the false-negative rate about 10 percent. Putting that all together, one can use Bayes's methods to determine that a positive mammogram is due to cancer in only about 9 percent of the cases. In the German group, however, one-third of the physicians concluded that the probability was about 90 percent, and the median estimate was 70 percent. In the American group, 95 out of 100 physicians estimated the probability to be around 75 percent.

Similar issues arise in drug testing in athletes. Here again, the oft-quoted but not directly relevant number is the false positive rate. This gives a distorted view of the probability that an athlete is guilty. For example, Mary Decker Slaney, a world-class runner and 1983 world champion in the 1,500 and 3,000 meter race, was trying to make a comeback when, at the U.S. Olympic Trials in Atlanta in 1996, she was accused of doping violations consistent with testosterone use. After various deliberations, the IAAF (known officially since 2001 as the International Association of Athletics Federations) ruled that Slaney "was guilty of a doping offense," effectively ending her career. According to some of the testimony in the Slaney case the false-positive rate for the test to which her urine was subjected could have been as high as 1 percent. This probably made many people comfortable that her chance of guilt was 99 percent, but as we have seen that is not true. Suppose, for example, 1,000 athletes were tested, 1 in 10 was guilty, and the test, when given to a guilty athlete, had a 50 percent chance of revealing the doping violation. Then for every thousand athletes tested, 100 would have been guilty and the test would have fingered 50 of those. Meanwhile, of the 900 athletes who are innocent, the test would have fingered 9. So what a positive-doping test really meant was not that the probability she was guilty was 99 percent, but rather $^{50}/_{59}$ = 84.7 percent. Put another way, you should have about as much confidence that Slaney was guilty based on that evidence as you would that the number 1 won't turn up when she tossed a die. That certainly leaves room for reasonable doubt, and, more important, indicates that to perform mass testing (90,000 athletes have their urine tested annually) and make judgments based on such a procedure means to condemn a large number of innocent people.[9]

In legal circles the mistake of inversion is sometimes called the prosecutor's fallacy because prosecutors often employ that type of fallacious argument to lead juries to convicting suspects on thin evidence. Consider, for example, the case in Britain of Sally Clark.[10] Clark's first child died at 11 weeks. The death was reported as due to sudden infant death syndrome, or SIDS, a diagnosis that is made when the death of a baby is unexpected and a postmortem does not reveal a cause of death. Clark conceived again, and this time her baby died at 8 weeks, again reportedly of SIDS. When that happened, she was arrested and accused of smothering both children. At the trial the prosecution called in an expert pediatrician, Sir Roy Meadow, to testify that based on the rarity of SIDS, the odds of both children's dying from it was 73 million to 1. The prosecution offered no other substantive evidence against her. Should that have been enough to convict? The jury thought so, and in November 1999, Mrs. Clark was sent to prison.

Sir Meadow had estimated that the odds that a child will die of SIDS are 1 in 8,543. He calculated his estimate of 73 million to 1 by multiplying two such factors, one for each child. But this calculation assumes that the deaths are independent—that is, that no environmental or genetic effects play a role that might increase a second child's risk once an older sibling has died of SIDS. In fact, in an editorial in the *British Medical Journal* a few weeks after the trial, the chances of two siblings' dying of SIDS were estimated at 2.75 million to 1.[11] Those are still very long odds.

The key to understanding why Sally Clark was wrongly imprisoned is again to consider the inversion error: it is not the probability that two children will die of SIDS that we seek but the probability that the two children who died, died of SIDS. Two years after Clark was imprisoned, the Royal Statistical Society weighed in on this subject with a press release, declaring that the jury's decision was based on "a serious error of logic known as the Prosecutor's Fallacy. The jury needs to weigh up two competing explanations for the babies' deaths: SIDS or murder. Two deaths by SIDS or two murders are each quite unlikely, but one has apparently happened in this case. What matters is the relative likelihood of the deaths . . . , not just how unlikely . . . [the SIDS explanation is]."[12] A mathematician later estimated the relative likelihood of a family's losing two babies by SIDS or by murder. He concluded, based on the available data, that two infants are 9 times more likely to be SIDS victims than murder victims.[13]

The Clarks appealed the case and, for the appeal, hired their own statisticians as expert witnesses. They lost the appeal, but they continued to seek medical explanations for the deaths and in the process uncovered the fact that the pathologist working for the prosecution had withheld the fact that the second child had been suffering from a bacterial infection at the time of death, an infection that might have caused the infant's death. Based on that discovery, a judge quashed the conviction, and after nearly three and a half years, Sally Clark was released from prison.

The renowned attorney and Harvard Law School professor Alan Dershowitz also successfully employed the prosecutor's fallacy—to help defend O. J. Simpson in his trial for the murder of Simpson's ex-wife, Nicole Brown Simpson, and a male companion. The trial of Simpson, a former football star, was one of the biggest media events of 1994–95. The police had plenty of evidence against him. They found a bloody glove at his estate that seemed to match one found at the murder scene. Bloodstains matching Nicole's blood were found on the gloves, in his white Ford Bronco, on a pair of socks in his bedroom, and in his driveway and house. Moreover, DNA samples taken from blood at the crime scene matched O. J.'s. The defense could do little more than accuse the Los Angeles Police Department of racism—O. J. is African American—and criticize the integrity of the police and the authenticity of their evidence.

The prosecution made a decision to focus the opening of its case on O. J.'s propensity toward violence against Nicole. Prosecutors spent the first ten days of the trial entering evidence of his history of abusing her and claimed that this alone was a good reason to suspect him of her murder. As they put it, "a slap is a prelude to homicide."[14] The defense attorneys used this strategy as a launchpad for their accusations of duplicity, arguing that the prosecution had spent two weeks trying to mislead the jury and that the evidence that O. J. had battered Nicole on previous occasions meant nothing. Here is Dershowitz's reasoning: 4 million women are battered annually by husbands and boyfriends in the United States, yet in 1992, according to the FBI Uniform Crime Reports, a total of 1,432, or 1 in 2,500, were killed by their husbands or boyfriends.[15] Therefore, the defense retorted, few men who slap or beat their domestic partners go on to murder them. True? Yes. Convincing? Yes. Relevant? No. The relevant number is not the probability that a man who batters his wife will go on to kill her (1 in 2,500) but rather the probability that a battered wife who was murdered was murdered by her abuser. According to the Uniform Crime Reports for the United States and Its Possessions in 1993, the probability Dershowitz (or the prosecution) should have reported was this one: of all the battered women murdered in the United States in 1993, some 90 percent were killed by their abuser. That statistic was not mentioned at the trial.

As the hour of the verdict's announcement approached, long-distance call volume dropped by half, trading volume on the New York Stock Exchange fell by 40 percent, and an estimated 100 million people turned to their televisions and radios to hear the verdict: not guilty. Dershowitz may have felt justified in misleading the jury because, in his words, "the courtroom oath—'to tell the truth, the whole truth and nothing but the truth'—is applicable only to witnesses. Defense attorneys, prosecutors, and judges don't take this oath . . . indeed, it is fair to say the American justice system is built on a foundation of *not* telling the whole truth."[16]

THOUGH CONDITIONAL PROBABILITY represented a revolution in ideas about randomness, Thomas Bayes was no revolutionary, and his work languished unattended despite its publication in the prestigious *Philosophical Transactions* in 1764. And so it fell to another man, the French scientist and mathematician Pierre-Simon de Laplace, to bring Bayes's ideas to scientists' attention and fulfill the goal of revealing to the world how the probabilities that underlie real-world situations could be inferred from the outcomes we observe.

You may remember that Bernoulli's golden theorem will tell you *before* you conduct a series of coin tosses how certain you can be, if the coin is fair, that you will observe some given outcome. You may also remember that it will not tell you *after* you've made a given series of tosses the chances that the coin was a fair one. Along the same lines, if you know that the chances that an eighty-five-year-old will survive to ninety are $^{50}/_{50}$, the golden theorem tells you the probability that half the eighty-five-year-olds in a group of 1,000 will die in the next five years, but if half the people in some group died in the five years after their eighty-fifth birthday, it cannot tell you how likely it is that the underlying chances of survival for the people in that group were $^{50}/_{50}$. Or if Ford knows that 1 in 100 of its automobiles has a defective transmission, the golden theorem can tell Ford the chances that, in a batch of 1,000 autos, 10 or more of the transmissions will be defective, but if Ford finds 10 defective transmissions in a sample of 1,000 autos, it does not tell the automaker the likelihood that the average number of defective transmissions is 1 in 100. In these cases it is the latter scenario that is more often useful in life: outside situations involving gambling, we are not normally provided with theoretical knowledge of the odds but rather must estimate them after making a series of observations. Scientists, too, find themselves in this position: they do not generally seek to know, given the value of a physical quantity, the probability that a measurement will come out one way or another but instead seek to discern the true value of a physical quantity, given a set of measurements.

I have stressed this distinction because it is an important one. It defines the fundamental difference between probability and statistics: the former concerns predictions based on fixed probabilities; the latter concerns the inference of those probabilities based on observed data.

It is the latter set of issues that was addressed by Laplace. He was not aware of Bayes's theory and therefore had to reinvent it. As he framed it, the issue was this: given a series of measurements, what is the best guess you can make of the true value of the measured quantity, and what are the chances that this guess will be "near" the true value, however demanding you are in your definition of *near*?

Laplace's analysis began with a paper in 1774 but spread over four decades. A brilliant and sometimes generous man, he also occasionally borrowed without acknowledgment from the works of others and was a tireless self-promoter. Most important, though, Laplace was a flexible reed that bent with the breeze, a characteristic that allowed him to continue his groundbreaking work virtually undisturbed by the turbulent events transpiring around him. Prior to the French Revolution, Laplace obtained the lucrative post of examiner to the royal artillery, in which he had the luck to examine a promising sixteen-year-old candidate named Napoléon Bonaparte. When the revolution came, in 1789, he fell briefly under suspicion but unlike many others emerged unscathed, declaring his "inextinguishable hatred to royalty" and eventually winning new honors from the republic. Then, when his acquaintance Napoléon crowned himself emperor in 1804, he immediately shed his republicanism and in 1806 was given the title count. After the Bourbons returned, Laplace slammed Napoléon in the 1814 edition of his treatise *Théorie analytique des probabilités*, writing that "the fall of empires which aspired to universal dominion could be predicted with very high probability by one versed in the calculus of chance."[17] The previous, 1812, edition had been dedicated to "Napoleon the Great."

Laplace's political dexterity was fortunate for mathematics, for in the end his analysis was richer and more complete than Bayes's. With the foundation provided by Laplace's work, in the next chapter we shall leave the realm of probability and enter that of statistics. Their joining point is one of the most important curves in all of mathematics and science, the bell curve, otherwise known as the normal distribution. That, and the new theory of measurement that came with it, are the subjects of the following chapter.