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.