Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland, 15,* 246–263.
Gawande, A. (1999, February 8). The cancer-cluster myth. *New Yorker,* pp. 34–37.
Gigerenzer, G. (2006). Out of the frying pan into the fire: Behavioral reactions to terrorist attacks. *Risk Analysis, 26,* 347–351.
Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314.
Hotelling, H. (1933). Review of *The Triumph of Mediocrity in Business. Journal of the American Statistical Association, 28,* 463–465.
Kareev, Y. (1992). Not that bad after all: Generation of random sequences. *Journal of Experimental Psychology: Perception and Performance, 18,* 1189–1194.
Koehler, J. J., & Conley, C. A. (2003). The "hot hand" myth in professional basketball. *Journal of Sport & Exercise Psychology, 25,* 253–259.
Koning, R. (2003). An econometric evaluation of the effect of firing a coach on team performance. *Applied Economics, 35,* 555–564.
Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32,* 311–328.
Langer, E. J., & Roth, J. (1975). Heads I win, tails is chance: The illusion of control is a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology, 32,* 951–955.
Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the "hot hand." *Chance,* 2(4), 22–30.
Lopes, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 626–636.
McNamar, Q. (1940). A critical examination of the University of Iowa studies of environmental influences on IQ. *Psychological Bulletin, 18,* 63–92.
Meisler, S. (1977, December 30). Spain lottery—Not even war stops it. *Los Angeles Times,* p. D1.
Oops! Who's excellent now? (1984, November 5). *BusinessWeek,* 76–88.
Peters, T., & Waterman, R., Jr. (1984). *In search of excellence.* New York: Harper & Row.
Poincaré, H. (1952). *Science and method* (F. Maitland, Trans.). London: Dover. (Original work published 1914)
Sagan, C. (1997). *The demon-haunted world: Science as a candle in the dark.* New York: Ballantine.
Secrist, H. (1933). *The triumph of mediocrity in business.* Chicago: Bureau of Business Research, Northwestern University.
Sivak, M., & Flannagan, M. J. (2003). Flying and driving after the September 11 attacks. *American Scientist, 91,* 6–8.
Tversky, A., & Gilovich, T. (1989). The "hot hand": Statistical reality or cognitive illusion. *Chance,* 2(4), 31–34.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

8

Thinking Rationally About Uncertainty

The actual science of logic is conversant at present only with things either certain, or impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the Calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

—James Clerk Maxwell

8.1 What to Do About the Biases

Ulysses wisely had himself chained to his ship's mast before coming within earshot of the Sirens. He did so not because he feared the Sirens per se, but because he feared his own reaction to their singing. In effect, he took a precaution against himself, because he knew what he would be likely to do if he heard the Sirens. Similarly, the cognitive biases of automatic thinking can lead us astray, in a predictable direction. We must take precautions to avoid the pitfalls of such unexamined judgment.

One of the goals of this book is to teach analytical thinking about judgment processes. The best way we know to think systematically about judgment is to learn the fundamentals of probability theory and statistics and to apply those concepts when making important judgments. Anyone who has taken or taught introductory probability theory realizes that Laplace's (1814/1951) famous dictum, that "the theory of probabilities is at bottom only common sense reduced to calculus" (p. 196), is certainly false. Probability theory was not invented until recent times, and our minds do not seem to be naturally "wired" to think according to its precepts. The first seven chapters in this book could be interpreted as a catalog of cognitive habits of thought that deviate, sometimes radically, from the laws of probability theory. We provide a summary of the basics of probability theory in the appendix of this book, but in the present chapter we will try to convey the essence of elementary probabilistic thinking illustrated with concrete examples.

Attempts to train people not to think representatively and not to be influenced by availability or other biases have been mostly unsuccessful. Associations are ubiquitous in our thinking processes (although perhaps they are not its sole "building blocks," as the English empiricists believed). Moreover, making judgments on the basis of one's experience is perfectly reasonable, and essential to our survival.

So, what is needed is some kind of alternative way of making these judgments, a method that "affirmatively" diverts us from relying on intuitions and associations and heuristics, at least when the judgments are important. One such precaution against our biases is the use of external aids. For example, a clinical psychologist can record instances (e.g., of suicide threats) on paper or in a computer file and then compile the data, using a symbolic formula or diagram, when he or she wishes to estimate the frequency. A simple charting of "good" and "bad" weeks can reveal a pattern—or the lack of one. Alternatively, even just writing down base rates and trying to apply the ratio rule can help us avoid irrational judgments.

The greatest obstacle to using external aids, such as the ones we will illustrate in this chapter, is the difficulty of convincing ourselves that we should take precautions against ourselves as Ulysses did. The idea that a self-imposed external constraint on action can actually enhance our freedom by releasing us from predictable and undesirable internal constraints is not an obvious one. It is hard to be Ulysses. The idea that such internal constraints can be cognitive, as well as emotional, is even less palatable. Thus, to allow our judgment to be constrained by the "mere numbers" or pictures or external aids offered by computer printouts is anathema to many people. In fact, there is even evidence that when such aids are offered, many experts attempt intuitively to improve upon these aids' predictions—and then they do worse than they would have had they "mindlessly" adhered to them. Estimating likelihood does in fact involve mere numbers, but as Paul Meehl (1986) pointed out, "When you come out of a supermarket, you don't eyeball a heap of purchases and say to the clerk, 'Well, it looks to me as if it's about $17.00 worth; what do you think?' No, you add it up" (p. 372). Adding,

keeping track, and writing down the rules of probabilistic inference explicitly are of great help in overcoming the systematic errors introduced by representative thinking, availability, anchor-and-adjust, and other biases. If we do so, we might even be able to learn a little bit from experience.

8.2 Getting Started Thinking in Terms of Probabilities

Modern probability theory got its start when wealthy noblemen hired mathematicians to advise them on how to win games of chance at which they gambled with their acquaintances (as noted in Chapter 1, in Cardano's case, the advice was for himself). Perhaps the fundamental precept of probabilistic analysis is the exhortation to take a bird's-eye, distributional view of the situation under analysis (e.g., a dice game, the traffic in Boulder, crimes in Pittsburgh, the situation with that troublesome knee) and to define a sample space of *all* the possible events and their logical, set membership interrelations. This step is exactly where rational analysis and judgments based on availability, similarity, and scenario construction diverge: When we judge intuitively, the mind is drawn to a limited, systematically skewed subset of the possible events. In the case of scenario construction, for example, we are often caught in our detailed scenario—focused on just one preposterously specific outcome path.

Daniel Kahneman and Dan Lovallo (1993) note that decision makers are prone to treat each problem as unique and to take an "inside view." Their remedy, similar to ours, is to deliberately take an *outside view,* in other words, to think of the current problem as a member of a category of many similar problems and to apply the rules of probabilistic thinking. To illustrate the importance of the outside view, Kahneman tells the story of a project he was involved in to design a curriculum for a new course:

> When the project team had been in operation for about a year, with some significant achievements already to its credit, the discussion at one of the team meetings turned to the question of how long the project would take. To make the debate more useful, I asked everyone to indicate on a slip of paper their best estimate of the number of months that would be needed to bring the project to a well-defined stage of completion: a complete draft ready for submission to the Ministry of Education. The estimates, including my own, ranged from 18 to 30 months. At this point I had the idea of turning to one of our members, a distinguished expert on curriculum development, asking him a question phrased about as follows: "We are surely not the only team to have tried to develop a curriculum where none existed before. Please try to recall as many such cases as you can. Think of them as they were in a stage comparable to the one at present. How long did it take them to complete

their projects?" After a long silence, something like the following answer was given, with obvious signs of discomfort: "First, I should say that not all teams that I can think of in a comparable stage ever did complete their task. About 40% of them eventually gave up. Of the remaining, I cannot think of one that was completed in less than seven years, nor of any that took more than ten." In response to a further question, he answered: "No, I cannot think of any relevant factor that distinguishes us favorably from the teams I have been thinking about. Indeed, my impression is that we are slightly below average in terms of our resources and potential." (Kahneman & Lovallo, 1993, p. 24)

The point is that judgments are likely to be more accurate if the judge can step back, take an outside view, and think distributionally and probabilistically, even if the thought process is only qualitative. Judgments will be even better if they can be based on systematically collected data and follow the quantitative rules of probability theory.

Probability theory starts with a precise vocabulary with which to describe elementary events, sets of events, and the relationships between them. Let's start with the well-defined example of throwing two dice. First, there is the *simplest event,* a value on a single upper face of the die; for example, "I throw a 1." Second, there is a *conjunction* of two simple events; for example, "I throw a 1 *and* a 6" (in either order, on either die). Third, there is a *disjunction* of two simple events; for example, "I throw a 1 *or* a 6, *or both*" (sometimes called the "inclusive or"). Fourth, there are *conditional* events, the occurrence of one event, *given* that another has occurred; for example, "I throw a 1 (on either die) *given* that I threw a total of 7 on the two dice." In the case of perfect dice, we can systematically describe the entire *sample space* of 36 possible, equal-probability outcomes: You might throw the numbers 1 to 6 on the first die, and the same numbers 1 to 6 on the second; so there are 6 × 6 possible pairings or conjunction events.

For present purposes, once we have conceptualized the sample space of possible events, we want to assign frequencies and probabilities to the simple events and relational events in the space. In the case of idealized situations, we can reason logically about the kinds of events, their frequencies, and probabilities as dice, card games, and other honest gambling devices closely approximate these ideals. So, since there is 1 face out of 6 that matches the description "1 is thrown," we would say the probability of a 1 is 1/6. And there are two outcomes where a 1 and a 6 are thrown, out of 36, so we have a 2/36 probability of throwing that conjunction; and there are 20 outcomes out of 36 where a 1 or a 6 or both are thrown, so we have a 20/36 probability of the disjunction event. Finally, for the conditional event, "1 on either die given a total of 7," we focus only on the given "total = 7" and calculate the probability as 2/6, since there are six events where the total = 7 and for two of them a 1 is thrown on one of the dice.

Now, let's work through another, less precisely defined situation: Suppose we are interested in events that involve the characteristics of college students. If we pick a student at random from the college student body at the University of Chicago, what is the probability that student will be female? Out of the approximately 5,026 undergraduates at Chicago in 2008, a total of 2,513 were women; so the probability a randomly picked undergraduate is a woman is 2,513/5,026 or approximately .50. What about the probability of majoring in the physical sciences? A total of 815 students declared a physical science as their major, so the probability of a randomly sampled college student at Chicago being in the physical sciences is 815/5,026 or .16. Now, what is the probability of the conjunction event of being both female and a physical sciences major? A total of 211 students were both female and physical science majors, so the probability is 211/5,026 or .04. As for the disjunction, female or physical science major or both, 3,117 meet this description, so the probability is 3,117/5,026 or .62. And what about the conditional, "female *given* physical sciences major"? The probability is 211/815 or .26—we consider only students who are physical science majors (there are 815 of them) and then ask, What is the probability a physical science major is a woman? Here's another case where the categories or sets that define the events are well-defined (let's assume that femaleness and physical science major can be defined precisely), so we can count empirical frequencies to infer probabilities (not the idealized, logical frequencies of the dice game).

Notice that the inverse conditional probability, "physical science major *given female,*" is *not* the same (211/2,513 or .08) as "female *given physical science major*" (211/815 or .26). In general, a conditional probability is not the same as its inverse, as illustrated by the ratio rule from Chapter 5—for example, $p(\text{female}|\text{physical science}) \neq p(\text{physical science}|\text{female})$.

What about an even murkier situation, where we can define sets and categories but there are no obvious frequencies to count? Suppose we are trying to decide if a Republican will win the presidential election in the year 2012. As we wrote this book (in 2008), the lead candidate for the Democratic party would be Barack Obama (incumbent president), but there is massive ignorance about potential Republican nominees 4 years hence. Some names, mostly successful state governors, are in the air, such as Sarah Palin (Alaska), General David Petraeus (the most visible military leader at the moment), and Newt Gingrich (former congressman, now conservative pundit), but no one has a clue who might be electable in 4 years. Even

Obama's candidacy is uncertain, as his first 4 years in office are sure to be full of surprises. Nonetheless, the distributional approach is still the best method to analyze the situation and to make predictions. We can list most of the plausible outcome categories, starting with the slates of potential candidates for each party and the still uncertain events that are likely to have an impact on the party and electorate's votes (economic conditions, personal scandal, health problems, charisma factor, campaign funds, etc.). In a situation like this, the systematic listing is unlikely to make us confident about precise probabilities, but it will remind us just how uncertain that future is and keep us from myopically developing one scenario and then believing in it too much. Despite the murkiness, distributional representations and probabilistic analysis are a big improvement over spontaneous judgments. However, we are unlikely to rely primarily on relative frequencies about this saliently vivid, unique event—though, if we're really at a loss, we might resort to some statistics that *may* be relevant, for example, *p*(incumbent wins). However, even when we reason based on scenarios and "reasons for" the possible outcomes, a deliberate attempt to represent the problem systematically will improve the coherence and accuracy of our judgments.

Let's consider one more case: Will there be nuclear weapons deployed by one nation against another in the next decade? Here even the outcomes are poorly defined: Is it a nuclear deployment if terrorists (perhaps not even identified with a single nation) detonate a nuclear device in a Middle Eastern country? Our scenarios, really descriptions of concrete possible outcomes, are nebulous: "A minor skirmish between United Nations peacekeeping troops and an African Warlord escalates . . ."; "An assassination attempt on an Israeli leader fails, and the reprisal . . ." Now, it seems there are no relevant frequencies to count. The future situation will be different from any previous historical situation that comes to mind. But we still believe that the systematic distributional approach is the best method by which to make educated, though vastly uncertain, probability estimates. In fact, there is evidence from psychological studies conducted by Asher Koriat, Sarah Lichtenstein, and Baruch Fischhoff (1980) that simply spelling out many of the relevant events and systematically thinking through the "reasons for" and "reasons against" the occurrence of each event increase the quality of judgments under ignorance.

What points did we want to make with these examples? First, we introduced the basic set membership relationships that are used to describe events to which technical probabilities can be assigned. Second, we introduced four kinds of situations to which we might want to attach probabilities: (i) situations, like conventional games of chance (e.g., throwing dice), where idealized random devices provide good descriptions of the underlying structure and where logical analysis can be applied to deduce probabilities; (ii) well-defined "empirical" situations where statistical relative frequencies can be used to measure probabilities (e.g., our judgments about kinds of students at the University of Chicago); (iii) moderately well-defined situations, where we must reason about causation and propensities (rather than relative frequencies—e.g., predicting the outcome of the next U.S. presidential election), but where a fairly complete sample space of relevant events can be defined with a little thought; and (iv) situations of huge ignorance, where even a sample space of relevant events is difficult to construct, and where there seem to be no relevant frequencies (e.g., international conflict in the next decade?).

One of the remarkable characteristics of probability theory is that four simple axioms (see Appendix) provide the rules for how to reason rationally and probabilistically, even though there is massive disagreement about what the numbers refer to. Our four examples were chosen to give a feeling for the spectrum of interpretations of probabilities: as an extension of elementary deductive logic; as real numbers based on frequencies of events in the external world; or as indices of subjective confidence *in our heads,* but not in the external world.

Third, many errors in judging and reasoning about uncertainty stem from mistakes that are made at the very beginning of the process, when comprehending the to-be-judged situation. If people could generate veridical representations of the to-be-judged situations and then keep the (mostly) set membership relationships straight throughout their reasoning, many errors would be eliminated. Of course, there are also misconceptions about probabilities and about random processes, but many times judgments under uncertainty are already off-track even before a person has tried to integrate the uncertainties. Our primary advice about how to make better judgments under uncertainty is focused on creating effective external (diagrammatic and symbolic) representations of the situation being judged.

8.3 Comprehending the Situation Being Judged

It may seem difficult to start from a written description of a novel uncertain situation and to create a comprehensive representation of the situation, although it is probably even harder to create situation models from direct experience. Raymond Nickerson (1996) has cataloged many of the errors that occur at the comprehension stage in an essay that focuses on the ambiguities in verbal probability problems. Some of the best-known examples have been enshrined in the popular literature on brain teasers. Let's start with a simple problem that has been the subject of research on probabilistic reasoning (introduced to psychology by Maya Bar-Hillel and Ruma Falk, 1982, p. 119); play the game and make a personal estimate before you read on:

> Three cards are in a hat. One is red on both sides ("red-red"), one is white on both sides ("white-white"), and one is red on one side and white on the other ("red-white"). A single card is drawn at random and tossed into the air and lands red-side up. What is the probability that it is the "red-red" card?

A common response is "1/2" or ".50" (given by 66% to 79% of the participants in experiments conducted by Bar-Hillel). Interviews with participants revealed that a typical justification for this answer is, "Well, since the card landed red-side up, we know it's not the 'white-white' card. There are two cards left, so it's 50-50 whether it's the 'red-red' card." The implication is that the written problem led these subjects to create a "first three, then two cards remaining" problem representation. However, the (umambiguously) correct representation is in terms of the *sides* of the cards, not the whole cards (see probability tree representation in Figure 8.1; Brase, Cosmides, & Tooby, 1998, also make this point). The sample space for all outcomes comprises six events—one for each side of each card. And, after red is observed face up, there are three events in the "active sample space" where a red side ends facing up: "red-white" (red up), "red_{side1}-red_{side2}" (red_{side1} up), or "red_{side1}-red_{side2}" (red_{side2} up), so the correct answer is 2/3—in 2 of the 3 equally probable events, the card is truly "red-red."

A more complicated problem was published in the "Ask Marilyn" column of a popular magazine and received much attention because the answer is surprising to most people and subtle enough to provoke disagreements among some famous mathematicians (vos Savant, 1991; Deborah Bennett provides a good summary to this problem in her popular introduction to probability theory, *Randomness* [1998]):

> Suppose you're on a game show, and you're given a choice of three doors. Behind one door is a car; behind the others, goats. You pick a door—say No. 1—and the host, who knows what's behind the doors, opens another door—say No. 3—which has a goat. He says to you, "Do you want to pick Door No. 2?" Is it to your advantage to switch your choice? (vos Savant, 1991, p. 12)

The first difficulty with this brain teaser is the surprising complexity of the situation of "possible events" to which it refers. Try to diagram the situation by systematically listing each of the relevant events: There are three doors that the contestant could pick; there are three possibilities for where the car is located; there are several options for which door the host could open (and they differ in number depending on which of these nine "situations" is encountered). Then

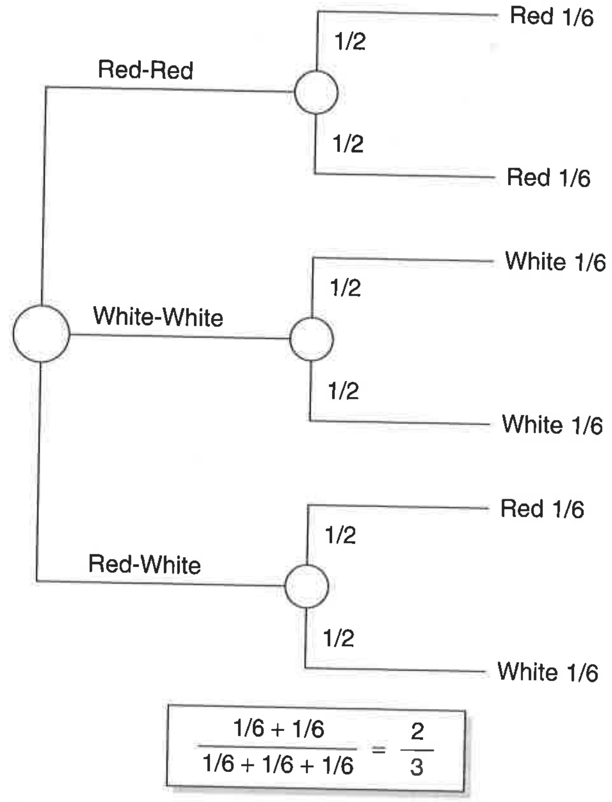


Figure 8.1 Probability tree representation of the Three Cards Problem

there is a further complexity created by which policy (stay or switch) is followed by the contestant; there are between 18 and 36 situations to keep in mind—*depending on which representation of the problem* a solver has settled on.

There is yet a further difficulty with this brain teaser because of the ambiguity of the written statement concerning the *host's rule for choosing a door to open;* unless this ambiguity is resolved, there is no unique sample space representation of the problem. There are at least three plausible interpretations of the host's rule given the problem statement. One rule is that the host always opens one of the non-chosen doors at random (e.g., by tossing a coin to choose Door No. 2 or Door No. 3 in the situation described in the written problem above). This means that he could open a door and reveal the car—presumably, then, he (and the audience) will just laugh at you for having chosen the wrong door and the game is over. But there is a second rule for the host, which is also consistent with the written problem statement:

Suppose the host always selects a door concealing a goat; never opens the door selected; and when the contestant has chosen the door concealing the car, the host picks a door at random. Now, there is a more complex dependency between the contestant's choice and the door opened by the host. An even more complex third rule has also been proposed: Suppose the host always selects a door concealing a goat; never opens the door selected by the contestant; and when the contestant has chosen door with the car behind it, the host has a bias to pick the remaining door with the lowest number (and there are other possible biases for this kind of rule). The underlying probabilities are different for each of these three rules—though all three are consistent with the original verbal problem statement.

The most popular representation to the problem is to interpret the problem statement to mean that the host always opens a door other than the one chosen originally by the contestant, and never opens a door revealing the car (i.e., follows Rule 2 above); then it is possible to say that the "switch doors" strategy may increase and will never decrease the probability of getting the car. So, under this representation of the problem, the answer is the contestant should switch. We provide an unambiguous representation of the problem in a probability tree format in Figure 8.2. The point is that representation is the essential, determinative first step in probabilistic reasoning. In the case of the Three Doors Problem, much confusion and controversy, and many academic journal articles, ensued from the ambiguity in the problem statement, though an unambiguous statement of the problem is complex and confusing in its own right. And real-world uncertainties and decisions are even more dauntingly ambiguous than probability word problems.

One major benefit of enrollment in probability and statistics courses is that it provides practice in translating situations into more precise and complete representations—or, in the case of real-world complexities, extracting the essential uncertainties and causal contingencies. We recommend the tables, probability (or decision) trees, and Venn diagrams that we use to illustrate most of the major judgment and decision situations described in this book. Unfortunately, the creation of appropriate and effective tables and diagrams is contingent on the specific problem being solved. We try the tree diagrams first, as they are the most generally effective, but sometimes one of the other formats is more illuminating. Fortunately, constructing these representations is a skill that any motivated student can learn with some practice. The first step is to study the examples in this book.

Furthermore, it is usually helpful to think about probabilities concretely in terms of *frequencies* of individuals in the relevant subgroups. People are much better able to keep track of the relationships between the partitions of the overall population when they imagine frequencies of individuals, objects, or events. In fact, some of the errors of judgment we presented in the previous chapters are

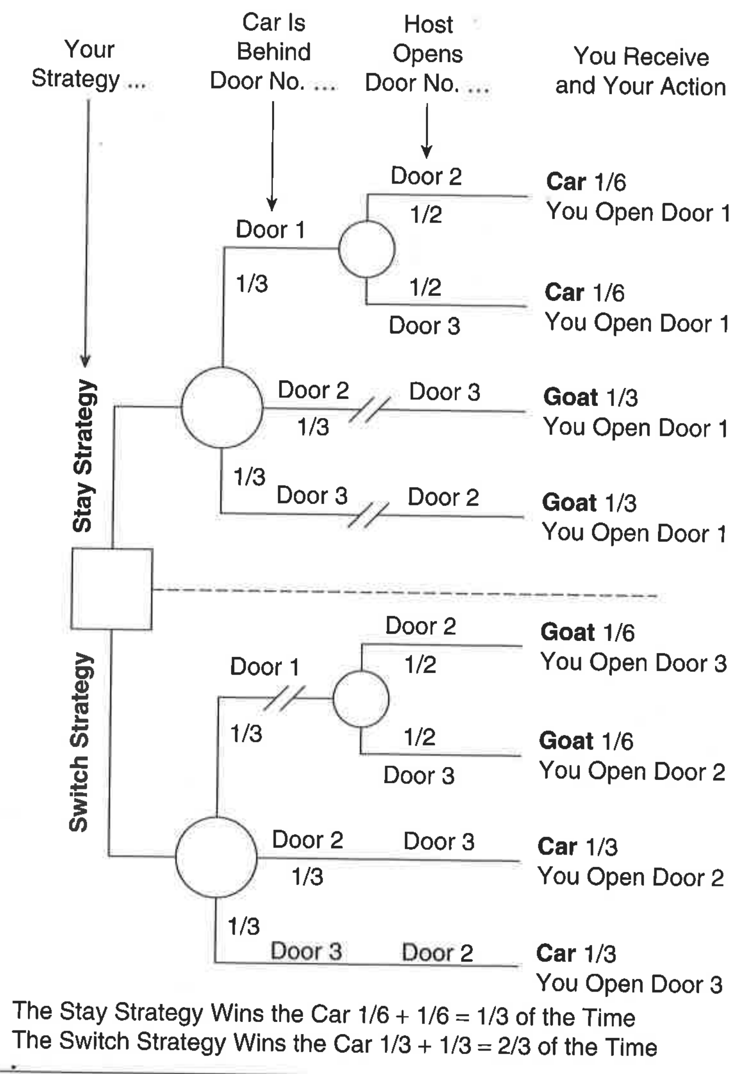


Figure 8.2 Probability tree representation for the Three Doors Problem when you choose Door No. 1 initially (This tree represents one-third of the possibilities—two more similar trees can be constructed for the cases where you choose Door No. 2 and Door No. 3, which exhausts the "sample space" of possibilities.)

reduced dramatically when people are encouraged to represent the situations in terms of frequencies instead of probabilities (e.g., Gigerenzer & Hoffrage, 1995; Sedlmeier & Betsch, 2002). Frequency formats are useful for preventing confusions of conditional probabilities (e.g., probability[cancer|positive test] versus probability [positive test|cancer]) and the conjunction error ("Linda is more likely to be a feminist bank teller than simply a bank teller of any kind.").

We will return to this theme of how to represent to-be-judged situations unambiguously and distributionally after we review the concept of rationality in judgments under uncertainty.

8.4 Testing for Rationality

In the first half of this book, we've provided many examples of judgments that are inaccurate or irrational. On what basis can we make such evaluations? The conditions necessary to conclude a judgment is *inaccurate* are relatively straightforward: (i) We need to have some measurable criterion event or condition in mind that is the target of the judgment; (ii) we need to be sure the person making the judgment is in agreement with us on the nature of the target and is trying to estimate, predict, or judge the same criterion value that we have in mind; and (iii) we also want to be sure that the judge is motivated to minimize error in the prediction and that the "costs" of errors are symmetric so the judge will not be biased to over- or underestimate the criterion. (For example, one of the authors of this book indicates that his judgments of acquaintances' ages [the criterion] are inaccurate and they tend to be systematically too low. But it is also important to know that this bias is partly deliberate to avoid offending people who are sensitive about being viewed as older than they really are.) This logic for assessing the quality of judgments has been dubbed the *correspondence framework* for accuracy, and it is the framework that underlies the Lens Model approach that we introduced in Chapter 3. (See Hammond, 1996, or Hastie & Rasinski, 1988, for further discussion.)

However, we also talk about irrationality or *incoherence* in judgments, when it is not obvious that a correspondence test can be applied. For example, we said that people who ranked "Linda is a feminist bank teller" as more probably true than "Linda is a bank teller" were irrational and made judgment errors, even though *there is no real Linda out there,* whose occupation and attitudes could serve as a criterion for a correspondence test of accuracy. In these cases, we are evaluating the quality of judgments, and we can only apply the approach to two or more judgments by considering their coherence or logical consistency with one another. We evaluate the judgments with reference to their consistency with the laws of logic and probability theory, which we accept as a standard of rational reasoning. By the way, if we are sure that a collection of judgments is incoherent, we can be sure that some are also inaccurate, though we often cannot say exactly which of the individual judgments are in error. (And, more generally, as

noted in Chapter 2, that which is self-contradictory cannot constitute a true description of the world.)

Another convincing argument that the judgment errors we have identified are truly irrational is that experimental participants shown their own responses and told the rule they have violated often conclude, "I made a mistake," or even, "Boy, that was stupid, I'm embarrassed." Kahneman and Tversky (1982, 1996), who first identified many of the errors we discussed, label the these judgment errors *illusions,* because they are behavioral habits that we know are mistakes when we think carefully about them, but they still persist when we do not exercise deliberate self-control to counteract our intuitive tendencies—very much like the many familiar, but still irresistible optical illusions.

This dissociation of deliberate reasoning and mostly automatic behaviors is the basis for the separation of analytic versus intuitive reasoning and memory processes (Kahneman, 2003). Seymour Epstein and his students (Denes-Raj & Epstein, 1994) have reported several studies in which some of the errors demonstrated by Kahneman and Tversky, such as in the Linda problem in Chapter 5 and other probability brain teasers, were reduced or eliminated by simply instructing experimental participants to answer "how a completely logical person would respond." They aptly titled their paper, "When do people behave against their better judgment?" However, in general, simply instructing someone to "behave logically" is not sufficient to induce rational thinking.

Once we have committed ourselves to using logic, mathematics, and decision theory as the standards to evaluate rationality in judgments and choices, there is much more work to be done to evaluate rationality in practice. First, it is not always obvious how to represent a decision situation objectively so that rational principles can be applied. Even when we have clear verbal descriptions, as in the brain teasers described at the beginning of this chapter, there is still often incompleteness, ambiguity, and even contradiction in our knowledge about the to-be-analyzed situation. Furthermore, it is often difficult to specify exactly what an actor's goals are in a situation, and most rational analysis requires knowing what the actor is trying to "maximize" to define a rational standard for evaluation. So, even if we have a well-specified standard for rationality, there can be problems in deciding if and how a response is irrational.

Second, it is not always appropriate to focus on the short-run performance of a fully informed person with plenty of time to think in an ideally quiet environment. We should be more interested in performance in the long run over many judgments, made in noisy environments with distractions, interruptions, and missing information. It may well be that the optimal, ideally rational judgment calculation is not the *adaptively* best judgment process under more realistic conditions. This theme has been developed recently by researchers led by John Payne (Payne, Bettman, & Johnson, 1993), by Lola Lopes and Gregg Oden (1991), and by Gerd Gigerenzer (Gigerenzer, Todd, & the ABC Research Group, 1999). These groups of scientists have argued that "fast-and-frugal" algorithms or heuristics for judgments and choices may be more robust, sturdier, and have better survival value than optimal calculations that are superior only when lots of information, computational capacity, and time are available.

So far, we have presented the "behavioral side" of our story, illustrating these judgment errors in the last four chapters, organized according to the cognitive processes and heuristics that underlie these judgments and produce the errors. Now we will discuss the judgment errors with reference to the rules of probability and logic that are violated, with some advice about how to avoid these irrationalities. We should warn the reader that sometimes it is difficult to infer exactly which rule of probability theory was violated first in a person's judgment process. Because the rules are all inextricably interrelated, it is difficult to know for sure whether the primary error is a misrepresentation of the set membership relations among the events being judged, an error in assuming two different contingencies or probabilities are the same, the ignoring of a critical piece of judgment-relevant information such as a background base rate, or something else.

8.5 How to Think About Inverse Probabilities

We have given several examples of judgment errors that arise because people (including your authors) are not careful to keep separate the easily confusable inverse probabilities (see especially Section 5.10). Let's spend some time dissecting a detailed example (reported in a news article by Gay McGee, 1979).

> BAY CITY, MICHIGAN, 1979: A surgeon here is one of a handful in the country who is taking a pioneering approach to the treatment of breast cancer. Charles S. Rogers, M.D., is removing "high-risk" breasts before cancer has developed.
>
> The risk factor is determined by mammogram "patterns" of milk ducts and lobules, which show that just over half of the women in the highest-risk group are likely to develop cancer between the ages of 40 and 59. The mammogram patterns are the work of Detroit radiologist John N. Wolfe, M.D.
>
> The surgery, called prophylactic (preventive) mastectomy, involves removal of the breast tissue between the skin and the chest wall as well as the nipple.
>
> Reconstruction of the breast with the remaining skin is usually done at the time of the mastectomy. Silicone implants and replacement of the areola (the pigmented skin around the nipple) leave the patient "looking like a woman," according to the surgeon.

> He has performed the surgical procedure on 90 women in two years.
>
> The rationale for the procedure is found in the *surgeon's interpretation* of studies by the radiologist, Wolfe. The newspaper article continues:
>
> In his research Wolfe found that one in thirteen women in the general population will develop breast cancer but that one in two or three DY (highest-risk) women *will develop it between the ages of 40 and 59.* [Italics added; Wolfe did *not* find that. What he discovered is explained in the next paragraph.]
>
> The low-risk women (NI) account for 42 percent of the population, but only 7.5% of the carcinomas. By examining the DY women and those in the next-lower risk groups, P1 and P2, Wolfe felt that 93% of the breast cancers could be found in 57% of the population.

One remedy to these confusions is to shift to systematic symbolic representations. Translating each to-be-judged situation into probability theory notation and then carefully applying basic rules from probability theory can help. (See the discussion of the useful Ratio Rule in Section 5.10 and the more general discussion of Probability Theory in the Appendix.) Let's look at that approach and a probability tree representation applied to the Rogers example. Using Rogers's figures, it is possible to construct results for 1,000 "typical" women (see the table in Figure 8.3). No other numbers satisfy the constraints. Note that 499 + 71 = 570, or 57% of the population, which is the stated proportion in the high-risk category. Also, 71/(71 + 6) = .92. Thus, as stated, 92% of the cancers are discovered in 57% of the population. The overall breast cancer rate of the population is (71 + 6)/1,000 = .077, and so on.

To return to Rogers on breast cancer, while it is true that 92% of the cancers are found in the high-risk group, *the estimated probability that someone in this group will develop cancer is only 71/570 or .12.* (Remember, these calculations are based on Rogers's own proportions.)

The .12 figure can be determined even more easily by applying the ratio rule. According to Wolfe's figures, p(cancer) = .075, p(high risk|cancer) = .93, and p(high risk) = .58. Thus,

$$\frac{p(\text{cancer}|\text{high risk})}{.93} = \frac{.075}{.580}\text{; therefore,}$$

$$p(\text{cancer}|\text{high risk}) = .12.$$

The most informative statistic is negative—the estimated probability of developing breast cancer if the woman is from the low-risk group

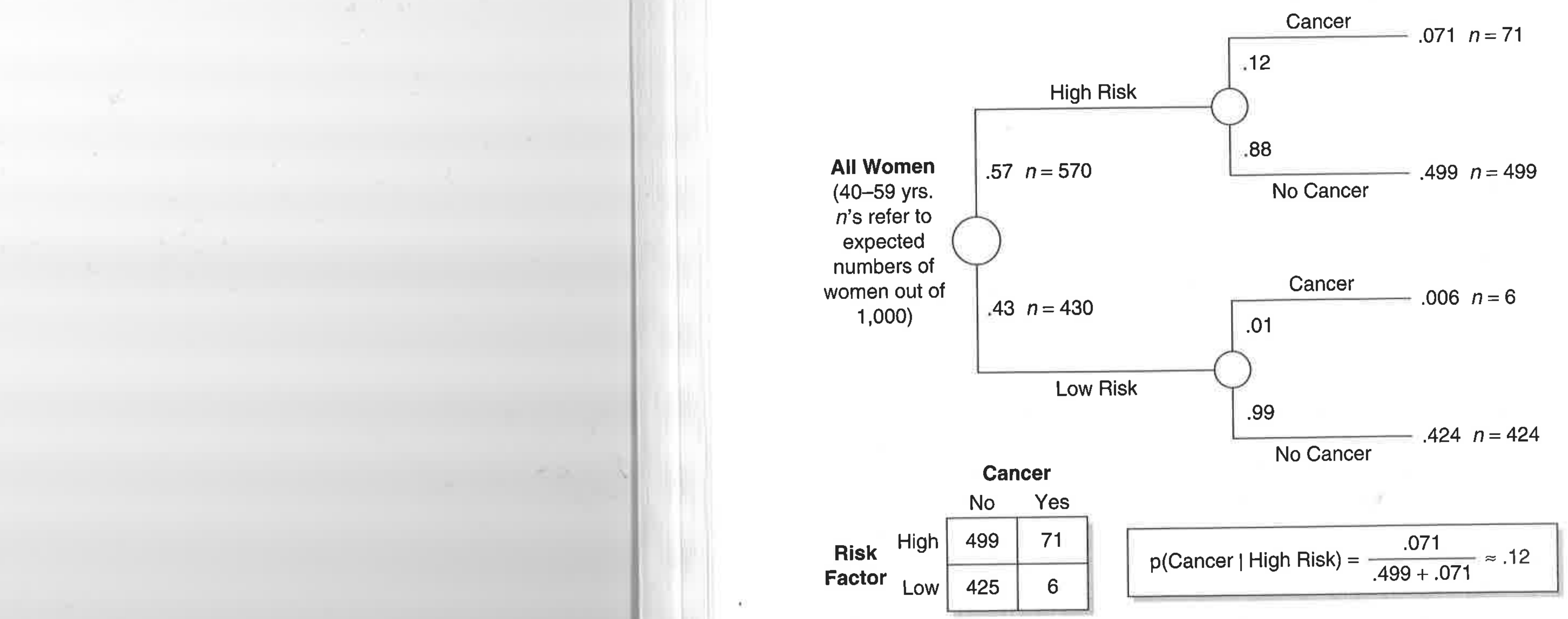


Figure 8.3 Probability tree representation of the Rogers breast cancer problem

being 6/(425 + 6), or .014. It is not possible based on the newspaper article to evaluate the claim about the very highest-risk group, DY.

Dr. Rogers does not stress the value of a negative inference. After urging *all women* over 30 to have an annual mammography examination, he is quoted as saying, "The greatest danger is in having a mammogram without a medical exam by a doctor. There are too many times when the surgeon feels a lesion that wasn't picked up on a mammogram. . . . This is definitely a case where one plus one equals more than three" (McGee, 1979).

Agreed. But, incidentally, his mammogram advice is also based on a confusion of inverse probabilities. Roughly 20% of cancers were not detected by mammography—that is, surgeons discovered a lesion "that wasn't picked up on a mammogram." But that is much different from the percentage of times women have cancer given a normal ("negative") mammogram result; $p(\text{cancer}|\text{negative}) \neq p(\text{negative}|\text{cancer})$. In fact, this former figure at the time the article was written was approximately .5% (1 in 200) according to figures

from the Hartford Insurance Project that had just been completed and published—which most of us would not regard as a "great danger." (In fairness to Rogers, it must be pointed out that the article did not specify how seriously high risk the "high-risk" patients would have to be before Rogers would operate. The point of the present critique is that his *reasoning* used to justify the procedure *at all* is, from a rational perspective, unpersuasive.)

In general, words are poor vehicles for thinking about inverse probabilities. It is clear that some verbal links are not symmetric; for example, "roses are red" does not mean that all red flowers are roses. Other verbal links, however, are symmetric; "dirigibles with hydrogen gas are the type that explode" also means that the type of dirigibles that explode are filled with hydrogen gas. It is easy to confuse symmetric and asymmetric verbal links. In fact, linguistic links are notorious for their ambiguity. (Does "the skies are not cloudy all day" mean that the skies are cloudy for only a portion of the day or never cloudy?) And it is possible to express sincere belief in a linguistic phrase without knowing what it means. (How many schoolchildren singing our national anthem know that *o'er* refers to "over" rather than "or"? Or when asked, "How many animals of each kind did Moses take on the Ark?" how many of us confidently answer, "two," without noticing that it was Noah, not Moses, who was supposed to have survived the Biblical flood on an ark.)

But it is difficult for many people to think without words. In fact, some eminent thinkers maintain that it is virtually impossible: "How do we know that there is a sky and that it is blue? Should we know of a sky if we had no name for it?" (Max Muller). "Language is generated by the intellect and generates intellect" (Abalard). "The essence of man is speech" (the Charodogya Upanishad). "In the beginning was the word" (Genesis 1:1). But perhaps the advice of the Lankavatara Sutra is more useful and correct: "Disciples should be on their guard against the seduction of words and sentences and their illusive meaning, for by them the ignorant and dull-witted become entangled and helpless as an elephant floundering around in deep mud." Or perhaps we should cultivate nonverbal thinking patterns such as those of Albert Einstein, who wrote, "The words or language, as they are written or spoken, do not seem to play any role in my mechanism of thought." But concrete, visual images are often no better than words, and images can still produce biased judgments.

Symbolic, especially algebraic representations are effective, but many people are not adept at algebra. Fortunately, graphical methods can be very helpful in representing probability problems and everyday situations. We have used Venn diagrams several times to clarify logical relationships, especially when conditional probabilities are involved. But for most problems we recommend decision trees and probability trees because they are more generally applicable and they are more useful for organizing numerical information relevant to decision problems.

8.6 Avoiding Subadditivity and Conjunction Errors

Another flagrant error we have described in judgments, especially those that depend on our sense of similarity and involve category memberships, is the habit of making estimates of several exclusive event probabilities that add up to more than 100%. For example, the probabilities that your car has failed to start because the battery is dead, or because a wire is loose, or because the gas line is plugged, or because the gas tank is empty, or because there is a seat-belt security bar on the ignition sum to 1.55. In its extreme form, subadditivity involves estimating that the probability of a subset, nested event is greater than the probability of a superset, superordinate event in which the subset event is nested (e.g., that Linda is more likely to be a feminist bank teller than any kind of bank teller at all). The problem is termed *subadditivity* because the probability of the whole is judged to be less than that of the sum of its parts—in the case of the conjuncture fallacy, less than that of a single part.

If we diagram the exclusive subset relations for "reasons our car didn't start," we are much less likely to distribute more than 100% of the probability space across the subset events, and we are also more sensible about estimating the basic probabilities of various failures (Figure 8.4). In fact, just reminding people verbally at the time they are making multiple subevent judgments that the sum of mutually exclusive events must not exceed 1.00, if they are using probability numbers correctly, is sufficient to produce more rational, comparative reasoning about event probabilities. Lori Van Wallendael and Hastie (1990) asked college students to solve some "whodunit" mysteries. Students who had not been reminded that the guilt judgments for different, mutually exclusive suspects needed to sum to unity, exhibited massive subadditivity. When they learned about new incriminating evidence, they increased their belief in the guilt of the most relevant suspect without reducing their suspicions about other suspects. However, when they were reminded of the "hydraulic" property of the mutually exclusive events, then they were more additive, and they traded off guilt and innocence judgments much more reasonably.

Representations like the probability tree or Venn diagram (see Figure 8.5) also reduce conjunction errors. In Section 5.8, we noted that if we draw a Venn diagram of the relationship between "bank tellers" and "bank tellers who are feminists," it is unlikely we will judge the probability (or frequency) of "feminist bank teller" to be higher than "bank teller." A probability tree

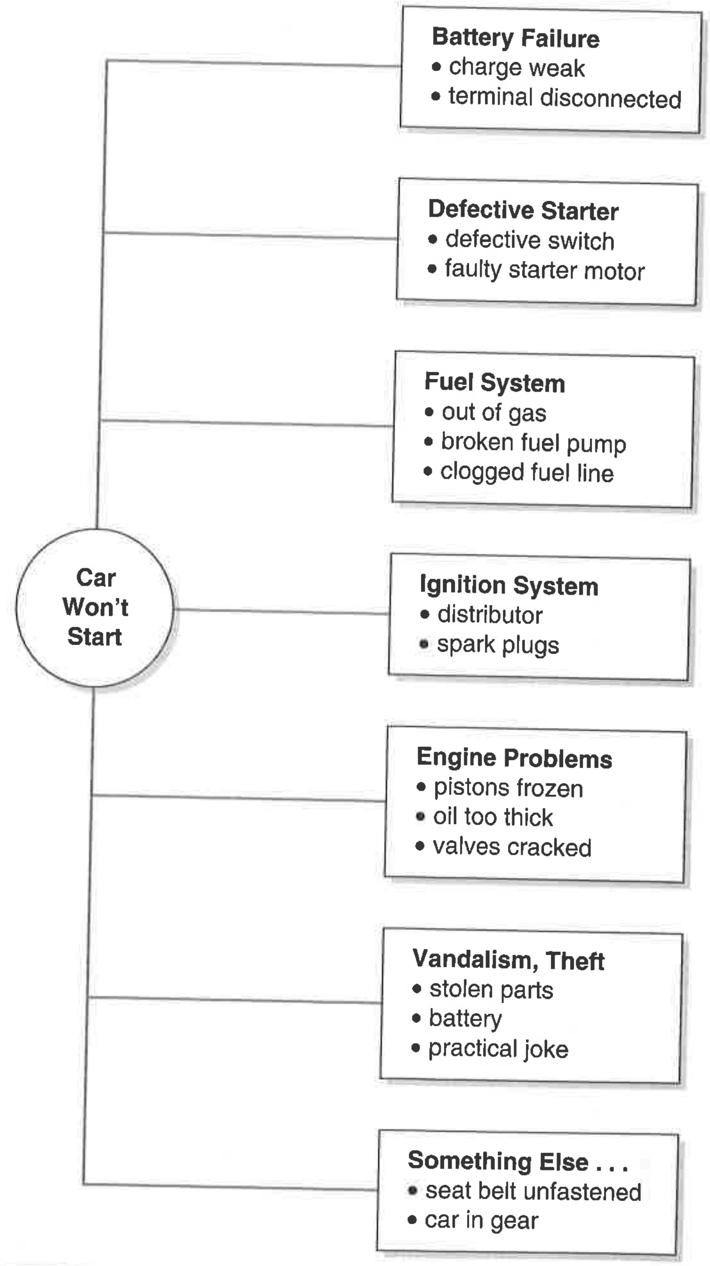


Figure 8.4 A plausible, but incomplete probability tree ("fault tree") to represent the reasons a car would not start

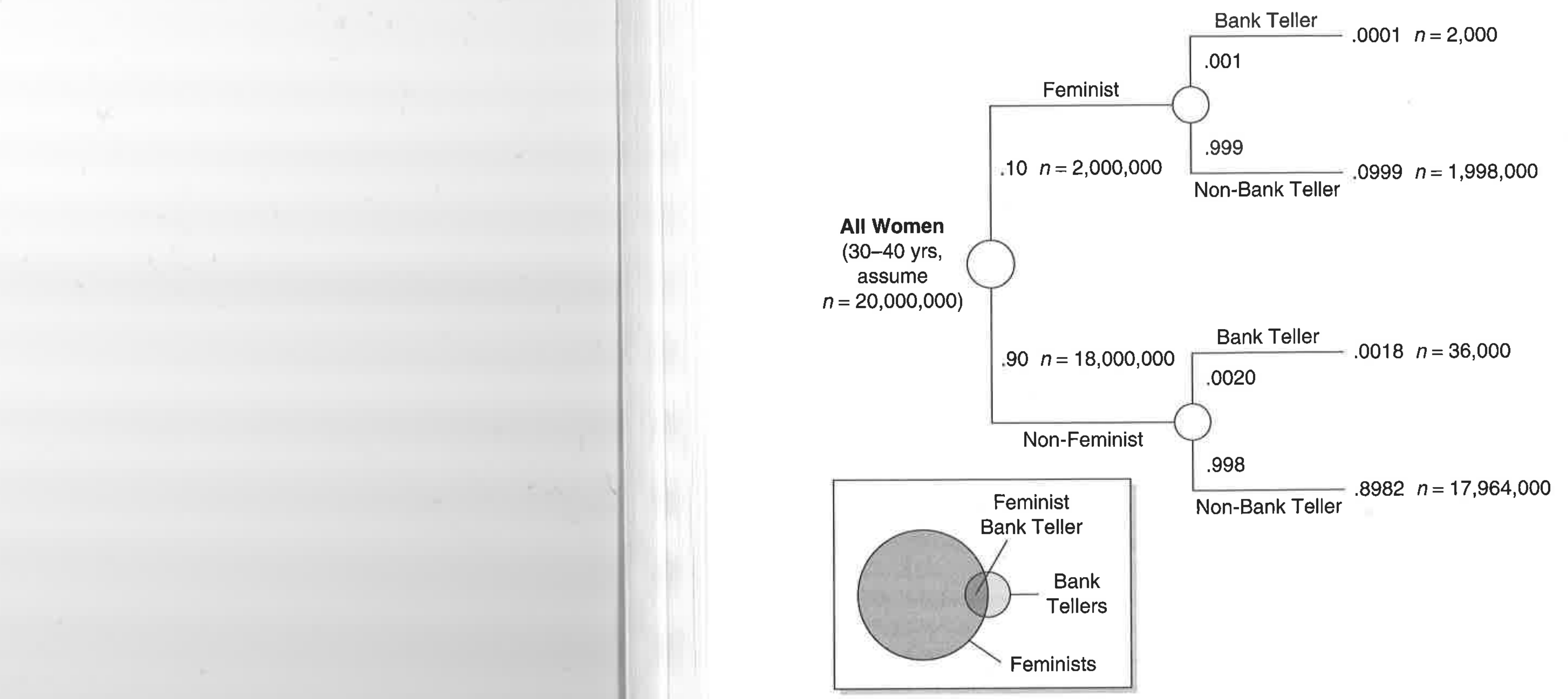


Figure 8.5 Probability tree and Venn diagram representations of the Linda the Feminist Bank Teller Problem (Some hypothetical assumptions about frequencies have been made to create plausible, but probably not exactly correct, frequencies: 20,000,000 women in the U.S. population in Linda's age cohort, 20 times more likely that a female bank teller is not a feminist than that she is a feminist, and 2 bank tellers out of every 1,000 women in the population.)

also protects us from this representative thinking error, and thinking about potential Lindas in terms of frequencies further de-biases our thinking. The conjunction error for the Linda the Feminist Bank Teller Problem was committed by 86% of the college students in the original probability format, but when Klaus Fiedler (1982) re-expressed the problem in a *frequency format,* the error rate dropped to about 20% (e.g., "Suppose that there are 100 people who fit Linda's description; how many of them are bank tellers? Bank tellers and active in the feminist movement?").

8.7 The Other Side of the Coin: The Probability of a Disjunction of Events

Consider a set of events 1, 2, . . . k. Suppose, moreover, that these events are *independent*—that is, whether or not one occurs has no effect on whether any of the others occurs, singly or in combination. (For a more precise definition of *independence,* see the Appendix.) Let the probabilities that the events occur be. $p_1 \times p_2 \times \ldots \times p_k$ What is the probability that *at least one* will occur? That is, what is the probability of the *disjunction* (as opposed to the conjunction) of these events? The probability of the disjunction is equal to 1 minus the probability that none will occur. But the probability that the first will *not* occur is $(1 - p_1)$, the probability that the second will not occur is $(1 - p_2)$, and so on. Hence, the probability that none will occur is $(1 - p_1) \times (1 - p_2) \times \ldots (1 - p_k)$. (This also is explained in the Appendix.) The product may be quite small, even though each $(1 - p_i)$ is quite large, because each p_i is small. For example, let the probabilities of six events be .10, .20, .15, .20, .15, and .10, respectively. Then the product of the $(1 - p_i)$'s is, once again, .90 × .80 × .85 × .80 × .85 × .90 = .37, so the probability that at least one of these events will occur is 1 − .37 = .63. The result occurs even though each separate event is quite improbable (the average being .15).

Just as we tend to overestimate the probability of conjunctions of events (to the point of committing the conjunction probability fallacy), we tend to *underestimate* the probability of disjunctions of events. There seem to be two reasons for this. First, our judgments tend to be made on the basis of the probabilities of individual components; as illustrated, even though those probabilities may be quite low, the probability of the disjunction may be quite high. We attribute this error primarily to the anchor-and-(under-) adjust estimation process. Second, any irrational factors that lead us to underestimate the probabilities of the component events—such as difficulty of imagining the event—may lead us to underestimate the probability of the disjunction as a whole. Occasionally, this underestimation problem is intuitively understood. For example, in their summations, lawyers avoid arguing from disjunctions in favor of conjunctions. (The great trial attorney, Richard "Racehorse" Haynes, illustrated the error of "arguing in the alternative" with the humorous example: "Say you sue me because you say my dog bit you. Well, now this is my defense: My dog doesn't bite. And, second, in the alternative, my dog was tied up that night. And, third, I don't believe you really got bit. And, fourth, I don't have a dog." Or more simply, Bart Simpson's famous defense: "I didn't do it; no one saw me do it; you can't prove anything.") Rationally, of course, disjunctions are *much* more probable than are conjunctions.

There is evidence for a *disjunction probability fallacy* comparable to the conjunction probability error—such a fallacy consisting of the belief that a disjunction of events is *less* probable than a single event comprising it (Bar-Hillel & Neter, 1993). But, of course, logically when the probability of A *and* B is higher than the probability of A alone (the conjunction fallacy), then the probability of not-A would be less than that of not-A *or* not-B. This is because the probability of not-A is 1 minus that of A and the probability of not-A or not-B is 1 minus that of A and B. So the former fallacy implies the latter. In fact, if we can arbitrarily decide what we call A and not-A (for example, call A not-being-a-feminist and not-A, being-a-feminist) and B and not-B (call B not-being-a-bank-teller and not-B, being-a-bank-teller), then aren't the two fallacious inequalities equivalent? Our answer is that they are logically equivalent, but not psychologically equivalent. We think in terms of categories, not their complements (negations). To a trained logician, not-A is as well-defined a category as A, but A's (which may have many associations) rather than not-A's (which tend to have few) crowd our minds. It takes a Sherlock Holmes to understand that the fact that the dog *did not bark* constitutes a crucial clue (implying that the dog was familiar with the criminal)—that is, to treat not-barking as an event.

8.8 Changing Our Minds: Bayes's Theorem

A very common judgment problem arises when we receive some new information about a hypothesis that we want to evaluate, and we need to update our judgment about the likelihood of the hypothesis. Consider a medical example that was introduced by physicians interested in how doctors and patients would interpret the new information provided by medical tests (Casscells, Schoenberger, & Graboys, 1978):

> The prevalence of breast cancer is 1% for women over age 40. A widely used test, mammography, gives a positive result in 10% of women without breast cancer, and in 80% of women with breast cancer. What is the probability that a woman in this age class who tests positive actually has breast cancer? (p. 999)

When David Eddy (1988) presented this problem to practicing physicians, an amazing 95 out of 100 responded with the answer, "About 75%." That estimate is dramatically incorrect—and in a context where these physicians deal with this type of judgment on a daily basis and where the numbers in the problem reflect the actual conditions surrounding mammography test results. What is the correct answer? About 7%—an order of magnitude lower than the physicians' modal answer!

One way to calculate the correct answer is symbolic, algebraic. If we study the rules of probability, it is not difficult to see that the following formula applies to this question (an informal derivation is provided in the Appendix, Section A.5):

$$p(\text{cancer}|\text{positive test}) = \frac{p(\text{cancer}_{\text{before the test}}) \times p(\text{positive test}|\text{cancer})}{p(\text{positive test}_{\text{with or without cancer}})}$$

The original problem statement gives us the probabilities we need to plug in on the right-hand side of the equation: $p(\text{cancer}_{\text{before the test}}) = .01$; $p(\text{positive test}|\text{cancer}) = .80$, and $p(\text{positive test}_{\text{with or without cancer}}) = .107$. The last term requires a little precalculation: If the person has cancer (1% of the women we are concerned with), the numerator gives us the probability .008 (= .01 × .80); if the person does *not* have cancer (99% of the women we are concerned with), we get the probability .099 (= .99 × .10); and since the only possibilities are having cancer or not having cancer, we add those two probabilities, .099 + .008 = .107. If we put all the numbers into the right-hand side of the equation, we get the following: (.01 × .80)/.107 or .075. That conclusion follows from the even simpler one, that $p(\text{cancer}|\text{positive test}) \times p(\text{positive test}) = p(\text{cancer}) \times p(\text{positive test}|\text{cancer})$.

The famous and useful formula for updating beliefs about a hypothesis (e.g., that an event is true or will occur) given evidence is called *Bayes' theorem* after Thomas Bayes, the British clergyman who derived it algebraically in his quest for a rational means to assess the probability that God exists *given* the (to him) abundant evidence of God's works. (Amazingly, almost any reader of this book can derive this profound theorem from the four basic principles of probability theory, once the derivation problem has been clearly stated; see the Appendix. The formula is also easily expressed as a probability tree; see Figure 8.6 for the application to the Eddy cancer diagnosis problem.)

$$p(\text{hypothesis}|\text{evidence}) = \frac{p(\text{evidence}|\text{hypothesis}) \times p(\text{hypothesis})}{p(\text{evidence})}$$

What systematic errors do people make as they try to update their beliefs about an event when they receive new information relevant to the judgment? We want to repeat our admonition that it is often difficult to figure out exactly which part of the judgment process is the fundamental error, and even harder to assign the error to a specific misconception or misapplication of probability theory. In the Eddy mammography example, we would describe the error as a failure to consider the alternative hypothesis, and to

Test Result	Cancer: No	Cancer: Yes
Positive	99	8
Negative	891	2

$$p(\text{Cancer}|\text{Positive Result}) = \frac{.008}{.008 + .099} \approx .07$$

Figure 8.6 A probability tree and a table to represent the situation described in the Eddy cancer diagnosis problem

ignore the probability that the evidence would be observed even if the hypothesis is false—that is, what we symbolized as *p*(positive test|no cancer) in our example above is often ignored. This focus on the salient hypothesis is a general habit of our attention and reasoning systems; we might even attribute it to the general bias toward available, salient information, which we allow to dominate our judgments. (Nickerson, 1998, provides a thorough introduction to this *confirmation bias.*) A second error is to ignore the base rates of occurrence of simple events (e.g., to underweight the fact that only 1% of the patients walking into the clinic are going to have breast cancer—before we know the results of any test).

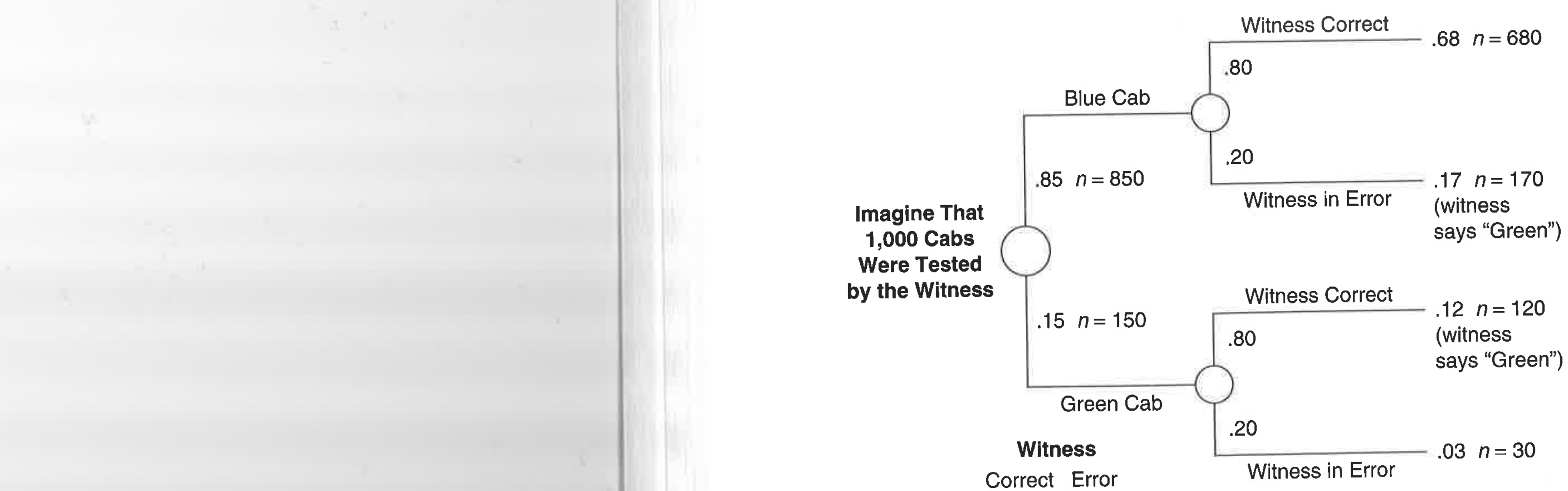
We've encountered the bad habit of ignoring base rates before, most obviously in Section 5.8 where we found that errors in judging Penelope's major field of study and errors in judging occupations of engineers versus lawyers

were due to a reliance on similarity between personality sketches and social stereotypes. But conceptualizing the errors with reference to probability theory, instead of psychology, we would say people were under-using or ignoring background base rates. Here's another example judgment from Bar-Hillel (1980) in which it is obvious that base rates are being ignored (again, make your own judgment before you read our analysis):

> Two cab companies operate in River City, the Blue and the Green, named according to the colors of the cabs they run. [A total of] 85% of the cabs are Blue and the remaining 15% are Green.
>
> A cab was involved in a hit-and-run accident at night. An eyewitness later identified the cab as Green. The Court tested the witness's ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found the witness was able to identify each color correctly about 80% of the time, but he confused it with the other color about 20% of the time.
>
> What do you think is the probability that the cab in the accident was Green, as the witness claimed? (p. 211)

Let's represent the information of the problem in terms of Bayes' theorem: In this problem, the most relevant base rate corresponds to the proportions of green and blue cabs on the streets, and it should be used as the starting point for the judgment—the "prior probability" of green before any case-specific evidence (e.g., the witness's testimony) is heard. What Bar-Hillel (1980) found, when she presented the problem to a varied sample of people, was almost universal failure to consider the base rate; once they heard about the concrete, case-specific eyewitness testimony, the base rate faded into the background. Thus, Bar-Hillel found that the modal response was the eyewitness's accuracy rate (.80), with no adjustment for the base rate information. If we plug the numbers into the Bayes' theorem formula (see Figure 8.7), we get the correct answer: .41.

We should acknowledge here that there is ambiguity in the written problem statement: Is the witness accuracy already conditioned on the base rate of green cabs, a posterior probability, because the witness was tested under "15% green cabs" conditions? Moreover, there are other interpretations that depart from the information in the problem statement by assuming that readers import various kinds information into their problem representations from their background experiences with taxicabs or traffic accidents or eyewitnesses (see, for example, Birnbaum, 1983). However, there is no direct evidence that these alternative representations are conceived of by anyone except the experts intent on criticizing Bar-Hillel's conclusions by speculating about alternate representations. In fact, unpublished data collected by



Cab Color	Witness Correct	Witness Error
Green	120	30
Blue	680	170

$$p(\text{Green}|\text{Witness Said "Green"}) = \frac{.12}{.12 + .17} \approx .41$$

Figure 8.7 Probability tree and table to represent the situation described in the Cab identification problem

one of the authors (Hastie) is most consistent with Bar-Hillel's interpretation that college students comprehend the problem as in the Bayes formula format presented above, but ignore the base rate information.

How can these errors be remedied? First, we noted in Section 5.10 that when the problem statement links the base rate information more strongly to the outcomes in the situation, especially when causal relationships make the connection, people are more likely to incorporate the base rates into their judgments. Bar-Hillel (1980) created a version of the cab problem that stated, "Police statistics show that in 15% of traffic accidents involving taxi cabs, a Green cab was involved." With this causal connection, the majority of people presented with the problem used the base rate information to adjust down from the 80% value implied by the witness's testimony, though

the adjustment (as we would expect) was insufficient. Perhaps this finding can be interpreted as a vote for the underlying rationality of our natural tendency to create and rely on situation models in the form of causal scenarios (see also Krynski & Tenenbaum, 2007). We speculate that causal scenario–based reasoning may be an intuitive way to keep track of the most important relationships among events—important when we need to make predictions, diagnoses, or just update our "situation models." However, spontaneous scenario-based reasoning cannot be sufficient by itself; most of the probability errors we have discussed are still prevalent when this mode of judgment is adopted.

Second, use of symbolic algebraic representations, like those we provide above, has a big impact on judgments. In medical diagnosis situations, there are now software decision aids that query physicians for relevant "prior probabilities" and "evidence diagnosticity" estimates and then compute the posterior probabilities. These systems improve performance in repeated clinical judgment situations, although there still seems to be a psychological mismatch between physicians' intuitive reasoning and the systems' response formats. People still have difficulty estimating the probability of observing the evidence (test result, witness testimony, symptom) *given* that the condition or disease was *not* present. But if a person making a judgment can deliberately spell out the problem in terms of the Bayes equation and then identify all the relevant information, performance improves. Even if the person only uses the formula to organize his or her thinking (but not to calculate), we expect improvements from (i) identification of incomplete and ambiguous descriptions of the judgment problem, (ii) consideration of nonobvious information necessary to make the calculation, and (iii) motivation to search for specific information and to think about focal hypothesis–disconfirming information (e.g., the probability that the witness says "green," *given* the cab was really blue; the probability that the test is positive, *given* that the patient did not have cancer; the probability of a DNA match, *even if* the defendant was not the criminal).

Third, and most helpful, we recommend the use of diagrams to represent the to-be-judged situation and to guide information search, inferences, and calculations as in Figures 8.6 and 8.7. Note that it is usually best to order the tree causally and temporally. In the mammogram diagnostic test situation, start with the fact that the prevalence of breast cancer is 1% for women over age 40. Then, consider the fact that the mammogram test is performed and it gives a positive result in 10% of women without breast cancer, and in 80% of women with breast cancer. What is the probability that a woman in this age class who tests positive actually has breast cancer? Finally, we recommend thinking about the situation in terms of frequencies; for example, imagine that 1,000 women are tested.

The task of reasoning coherently and rationally about probabilities is not a classroom homework problem anymore. All of us are more and more likely to encounter probabilistic evidence, presented as probability numbers in courtrooms and hospitals and financial investments. Consider the months of testimony and debate about DNA match and blood type evidence in the O. J. Simpson criminal and civil trials—or this woman reporter's story about a consultation with her physician after a mass was discovered in her breast (Kushner, 1976):

> "I'd like you to get a xero-mammogram. It's a new way to make mammograms—pictures of the breasts."
>
> "Is it accurate?"
>
> He shrugged, "Probably about as accurate as any picture can be. You know," he warned, "even if the reading is negative—which means the lump isn't malignant—the only way to be certain is to cut the thing out and to look at it under a microscope."

The woman then discussed the problem with her husband.

> "What did the doctor say?"
>
> "He wants to do a xero-mammogram. Then, whatever the result is, the lump will have to come out."
>
> "So why get the X-ray in the first place?"
>
> "It's something to go on, I guess. And our doctor says it's right about 85 percent of the time. . . . So, first I've scheduled an appointment to have a thermogram. If that's either positive or negative, and if it agrees with the Xerox pictures from the mammogram, the statistics say the diagnosis will be 95 percent reliable."

Is there any possibility that this patient will not have the tests? Or that she will decide not to get the lump biopsied *no matter what the test results?*

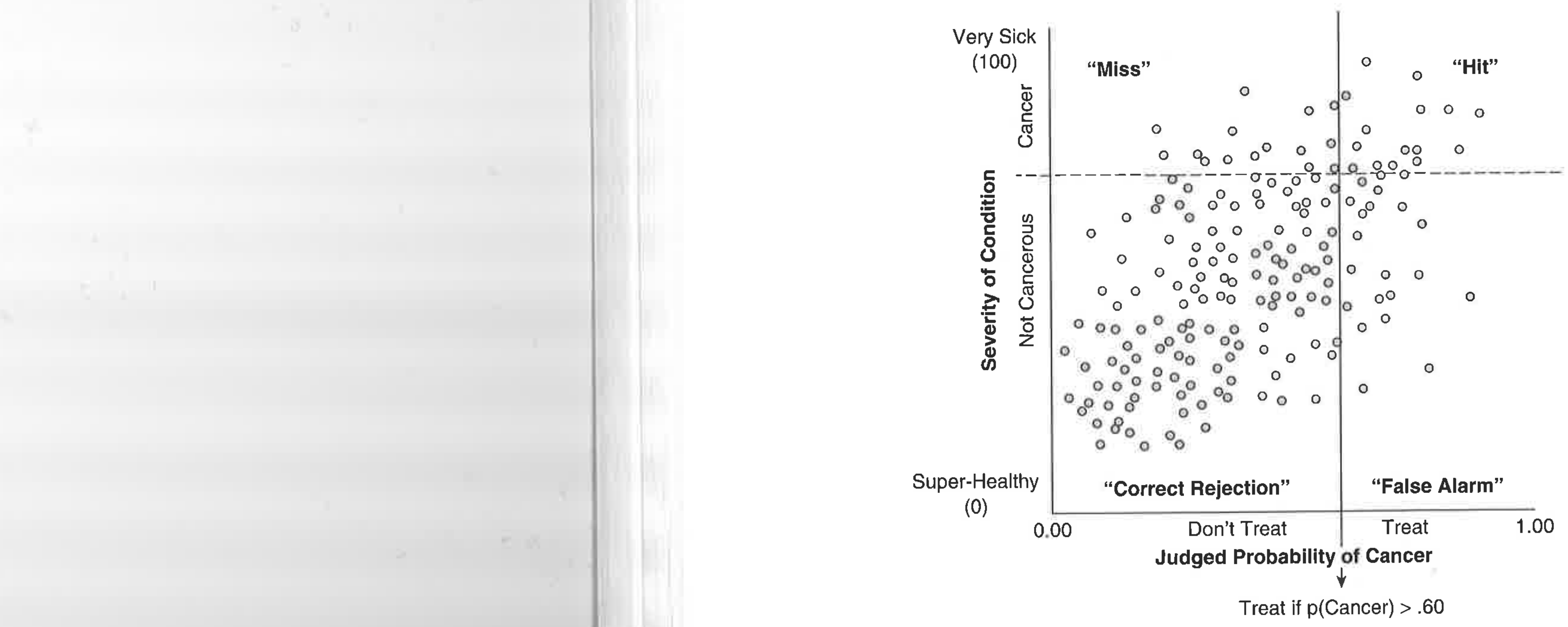
8.9 Statistical Decision Theory

Our discussion of estimates and judgments under uncertainty raises an important practical and theoretical question: How should we use judgments to decide whether or not to take consequential actions? The normative "should do" answer is provided by *statistical decision theory.* (We can only hint at the importance and sophistication of this area of theory; see also Macmillan & Creelman, 2004; Swets, Dawes, & Monahan, 2000). Let's consider a simple case where a physician assesses the probability that a patient has a serious condition like cancer and has to decide whether to operate or not. (Today, this is usually a joint physician–patient decision, although most patients want the physician to make the decision for them.) The bigger picture for this situation is provided by the scatterplot in Figure 8.8, representing this judgment being made for many similar patients. Millions of decisions can be summarized in this format, and the key question is, "How high does the probability have to be to take action?" Accept or reject, invest or don't invest, commit or withdraw, convict or acquit, evacuate or don't evacuate, retaliate or don't retaliate, and so on.

The answer to the "Should I take action?" question depends on these probabilities (relating your current knowledge and the true condition you're trying to infer) and how much you care about each of the four possible outcomes. (Note that in this simple but realistic example, if we knew for sure what the true condition was, we would know how to act; but since there's uncertainty, we face a tough decision.) More specifically, if we know how we value the outcomes, we can work backward and calculate the threshold probability that prescribes when we should shift from inaction to action, to maximize those values.

Figure 8.8 includes the popular labels for each of the four judgment-outcome possibilities: (a) a "hit" or "true positive" is when the judgment correctly identifies the focal target condition, for example, when the judgment correctly identifies the cancerous condition; (b) a "miss" or "false negative" is when the judgment incorrectly implies the patient does *not* have the condition; (c) a "false alarm" or "false positive" is when the judgment incorrectly implies the patient *does* have the condition; and (d) a "correct rejection" or "true negative" is when the judgment correctly indicates the patient does *not* have the condition. (The figure represents a situation in which 30 people truly have cancer and 170 are healthy out of 200, and where the correlation between the physician's judgment and the presence/absence of cancer is approximately +.65.)

One insight that immediately comes from the scatterplot is that we can control the rates at which the various judgment-outcome combinations occur by varying our threshold to decide to operate. If we set the *operate threshold* at the point where we judge the probability of cancer is .60, we see 15 hits, but also 15 misses (7.5% overall; 15 out of 30, or 50% of those with cancer), but we pay for quite a few false alarms (unnecessary operations—about 10% of all cases; 20 out of 35, or 57% of the operations performed). If we lower the threshold for treatment to judged p(cancer) of .50, we



Truth	Judged p(Cancer) > .60 Treat	Judged p(Cancer) < .60 Don't Treat
Cancer	"Hit"	"Miss"
No Cancer	"False Alarm"	"Correct Rejection"

Truth	Judged p(Cancer) > .60 Treat	Judged p(Cancer) < .60 Don't Treat
Cancer	15	15
No Cancer	20	150

Figure 8.8 Statistical decision theory diagram (This diagram represents a medical decision under uncertainty where a physician makes a judgment of the probability that a patient has cancer and then acts to treat or not treat the patient based on that assessment. The problem is represented as a hypothetical series of judgments for 200 patients, who vary in health status. The judgment is moderately accurate; r = .65 between judgment and true health status. The judge is depicted as deciding to treat a patient if the judged *p*(CANCER) is greater than .60; that set of judgments and outcomes for the 200 patients is summarized in the table at the bottom of the figure in terms of the statistical decision theory concepts of "hit," "false alarm," "miss," and "correct rejection." Warning: Different applications of statistical decision theory may use different organizations of the summary table; the presentation here is consistent with the conventions of psychological *signal detection theory,* which is one useful version of statistical decision theory.)

increase hits to 20 (from 15) and reduce misses to 10, but pay for that with more false alarms (30, or 15% overall; 20 out of 35, or 57% of all operations, are on healthy patients).

This is a very important insight that seems to be ignored in many policy discussions: In many cases, more careful thought about what we care about, which errors we most want to avoid, can improve the decision. Often, we cannot increase the accuracy of a diagnosis or other judgment (in this example, we can't increase the diagnostic accuracy of the physician), but we can trade off the two types of errors (and "corrects," too). If misses are most costly, we can lower our threshold for action on the judgment dimension and reduce misses (but at the cost of more false alarms); if false alarms are the costly error, we can move the decision threshold up and reduce that error (trading it for more misses, of course). We often try to avoid these tragic trade-offs by increasing accuracy, so that there is less error of both types to trade off. Thus, we spend billions of dollars every year increasing the accuracy of medical, military, financial, and meteorological forecasts to accomplish just this. But there are very few policy situations in which we can abolish all uncertainty. In most circumstances, we should recognize we are stuck with trade-offs, proceed with a sensible discussion of what we value, and then set a decision threshold accordingly (Hammond, 1996).

If we do face these trade-offs, we need to try to value the various judgment-outcome combinations and then apply statistical decision theory to set a proper decision threshold. For example, suppose that the numbers +100, 0, +30, and +80 represent our values for each of the four outcome states (hit, miss, false alarm, and correct rejection; we stay with our convention of representing values on a 0–100 scale). Notice also that there may be significant disagreements between people about these values. A patient might place the highest value on a "hit" and lowest on a "miss" (as in our ordering on the value scale), while a policy maker might value "correct rejections" more highly and "false alarms" more negatively. To pursue our illustrative example, given a single numerical valuation, we can calculate the decision threshold that yields a maximum on the value function. In this example, the policy with the maximum overall value is attained by setting the decision threshold equal to a judged *p*(cancer) of approximately .55 (we omit the calculation because it involves calculus).

Thus, in many practical situations, we should be thinking harder about values, rather than accuracy. But the determination of values is itself a complex process, even when only one decision maker is involved (see the next two chapters), because everyday alternatives often have many attributes and serve multiple objectives. The task is even more daunting when we must perform an analysis across stakeholders with different personal values, as we must in any organizational or societal policy analysis.

However, these difficulties must not divert us from trying to think harder, more systematically, and from multiple perspectives about the unavoidable trade-offs.

8.10 Concluding Comment on Rationality

If a scientific theory cannot state when an event will occur, a skeptic might ask what good it is. In fact, a dedicated behaviorist (if any still exist) might critique this entire book on the grounds that since the phenomena discussed are not controllable, descriptions of them—and the mechanisms hypothesized—are of no scientific value. The answer is that insofar as we are dealing with mental events and decisions of real people in the booming, buzzing confusion of the real world, we can neither predict nor control them perfectly. The qualifier of *ceteris paribus* ("other things being equal") always applies to these phenomena. The uncertainty of predicting actual outcomes in the world is intrinsic to both the problem and the consequences of decision. Of course, it may be said that true scientists should not investigate such uncertain phenomena—that they should perhaps limit themselves to investigating the rate at which a rat presses a bar in an environment where the only moving part is the bar. (What, other than the consequences of manipulating the one thing that can be manipulated, could shape the rat's behavior?) But if all scientists followed this rule and stayed in their neat ivory towers, we would not have meteorology, agricultural science, genetic counseling, computer science, and many other useful applied sciences.

Of course, perfectly rational thought processes do not guarantee true conclusions. It is necessary to have realistic, valid inputs, too. When Hastie was teaching his judgment and decision making course for the first time, he encountered a middle-aged student in a class of 20-year-olds. After a few lectures, the older student introduced himself and explained why he was enrolled in the course. As the student told it, he had suffered a series of apparent misfortunes and was in the midst of a divorce and fighting his employer's efforts to fire him. He said that at first he was confused by the events that were happening to him, but that on reflection he realized that he was actually the subject of a huge "psychology experiment." (In fact, one reason he had come to Harvard to study was because he wanted to meet Professor B. F. Skinner, whom he believed was the experimenter who was manipulating his life.) He went on to cite dozens of instances of inexplicable behaviors and events that only made sense if the hypothesis that he was "in a psychology experiment" was true. Hastie asked for specific examples, but was not much impressed by the strength of the evidence for the student's hypothesis; most of the examples seemed at least as probable under the alternate

hypothesis that the student was *not* in a psychology experiment (e.g., "I was interrupted by my wife, and she spoke exactly the words I was thinking, before I could say them myself"; "I was having a drink with a coworker after work, and he mentioned the company was laying off workers, and this happened only days before they told me I was fired"). The bright side of the student's delusional system was that he believed that eventually the experiment would be finished and publicized and that his demonstrated aptitude (for being controlled by the experimenter) would certify his qualities as a leader, who would be trustworthy and accountable in a high public office.

Perhaps the most fascinating part of this anecdote concerns the student's explanation for why he had approached Hastie: He was concerned that he not be irrationally deluded in his interpretation of these events. So, in order to ensure that he not reach a false conclusion, he was attempting to apply the instructor's advice as carefully as he could. After enrolling in Hastie's class, he realized that he needed to deliberately apply Bayes' theorem to evaluate the posterior probability of the hypothesis that "I am the subject of a huge, secret psychology experiment," with reference to the many items of evidence he had accumulated. He wanted help with his calculations to evaluate that hypothesis!

The story did not end that semester. A few months later, the student called on Hastie to testify on his behalf in his lawsuit to retain his job. Psychiatrists for his employer had asserted that he was suffering from massive paranoid delusions (which Hastie found plausible); that Bayes' theorem was part of his delusional system; and furthermore, that the Reverend Thomas Bayes was a figment of his schizophrenic imagination. (Hastie provided a deposition rebutting the psychiatrists' claims that Thomas Bayes was a delusion, although he noted that he was dubious of that posterior probability: 999,999/1,000,000 and the conclusion that the student was the subject of a massive social experiment. Of course, the experience of producing the deposition also led Hastie to wonder why he was so confident in the existence of a vague historical figure whom he knew of only because a mathematical theorem was named after him.) "Delusions in, delusions out"—no matter how consistent the calculations connecting the two.

What we have attempted to do is to point out factors and thinking styles that lead us all to make irrational judgments and choices (based on those judgments). People will not necessarily engage in these thought processes, any more than a swimmer who panics necessarily attempts to keep his or her head above water. Like a swimmer with survival training, we can learn to counteract the intuitive response and to be more rational—but as with the swimming example, it takes knowledge, self-control, and effort. From a normative perspective, however, learning to specify conditions that facilitate or inhibit certain types of behavior, or to distinguish between productive and nonproductive ways of thinking, is quite an accomplishment for psychologists or other social scientists.

Finally, as will be pointed out in the concluding chapter, people who attempt to grasp the totality of situations in order to predict or control exactly what will happen seldom fare as well as those who seek the more modest goal of living with the uncertainty that is irreducible and determining what we can influence. A person who attempts to understand everything can easily end up understanding nothing. An understanding of irrational forms of thinking is not nothing, even though we cannot predict exactly when such irrationality will occur or always how to control it.

References

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44,* 211–233.

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition, 11*(2), 109–122.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology, 65,* 1119–1131.

Bennett, D. J. (1998). *Randomness.* Cambridge, MA: Harvard University Press.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General, 127,* 3–21.

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*(18), 999–1001.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processes: When do people behave against their own better judgment. *Journal of Personality and Social Psychology, 66,* 819–829.

Eddy, D. (1988). Variations in physician practice: The role of uncertainty. In J. Dowie & A. S. Elstein (Eds.), *Professional judgment: A reader in clinical decision making* (pp. 200–211). Cambridge, UK: Cambridge University Press.

Fiedler, K. (1982). Causal schemata: Review and criticism of research on a popular construct. *Journal of Personality and Social Psychology, 42,* 1001–1013.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* New York: Oxford University Press.

Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193–208). Cambridge, UK: Cambridge University Press.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk-taking. *Management Science, 39,* 17–31.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11,* 123–141.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103,* 582–591.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General, 136,* 430–450.

Kushner, R. (1976, March 24). Breast cancer—the night I found out. *San Francisco Chronicle,* p. C1.

Laplace, P. S. (1951). *A philosophical essay on probabilities* (F. W. Truscott & F. L. Emory, Trans.). New York: Dover. (Original work published 1814)

Lopes, L. L., & Oden, G. D. (1991). The rationality of intelligence. In E. Eels & T. Maruszewski (Eds.), *Poznan studies in the philosophy of the sciences and humanities* (Vol. 21, pp. 225–249). Amsterdam: Rodopi.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

McGee, G. (1979, February 6). Breast surgery before cancer. *Ann Arbor News,* p. B1 (reprinted from the *Bay City News*).

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Nickerson, R. S. (1996). Ambiguities and unstated assumptions in probabilistic reasoning. *Psychological Bulletin, 120,* 410–433.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments, & Computers, 29,* 328–336.

Sedlmeier, P., & Betsch, T. (2002). *Etc.: Frequency processing and cognition.* New York: Oxford University Press.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American, 283*(4), 70–75.

Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition, 18,* 240–250.

Vos Savant, M. (1991, February 17). Ask Marilyn. *Parade Magazine,* 12.