Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311–327.

Heath, C., & Heath, D. (2007). *Made to stick: Why some ideas survive and others die.* New York: Random House.

Jardel Co. v. Hughes, Del. Supr., 523 A.2d 518 (1987).

Lewinsohn, P. M., & Rosenbaum, M. (1987). Recall of parental behavior by acute depressives, remitted depressives, and nondepressives. *Journal of Personality and Social Psychology, 52,* 611–620.

March, J. G. (1972). Model bias in social action. *Review of Education Research, 42,* 413–429.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9,* 1–22.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 521–533.

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review, 13,* 519–557.

Schank, R. C., & Abelson, R. P. (1995). Knowledge and memory: The real story. In R. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 8, pp. 1–86). Hillsdale, NJ: Lawrence Erlbaum.

Silkwood v. Kerr-McGee Corp., 464 U.S. 238 (1984).

Spence, D. F. (1982). *Narrative truth and historical truth: Meaning and interpretation in psychoanalysis.* New York: Norton.

Spence, G. (1994, November 29). Winning attorneys. *New York Times,* p. E1.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Bulletin, 90,* 293–315.

Van den Broek, P., & Thurlow, R. (1991). The role and structure of personal narratives. *Journal of Cognitive Psychotherapy, 5,* 257–274.

Wasserman, D., Lempert, R. O., & Hastie, R. (1991). Hindsight and causality. *Personality and Social Psychology Bulletin, 17,* 30–35.

7

Chance and Cause

Say you're thinking about a plate of shrimp. Suddenly someone says plate, or shrimp, or plate of shrimp. Out of the blue. No use looking for one either. It's part of the lattice of coincidence that lays on top of everything.

—From the film *Repo Man,* written and
directed by Alex Cox, 1984

7.1 Misconceptions About Chance

On January 26, 1972, Vesna Vulovic, a 22-year-old Yugoslavian flight attendant, was serving drinks to passengers on JAT Flight 367 when the plane was demolished by a bomb planted by a Croatian nationalist group. Most people would think she was extremely unlucky—first, to be on a rare flight destroyed by a terrorist bomb and second, because of a name confusion, she had been assigned to work the wrong flight. Definitely the wrong flight. But there is a positive side to this story. Ms. Vulovic lived and now holds the world record for surviving the highest fall without a parachute—33,000 feet (10,000 meters). Just a little more than a year after the fall, she declared herself ready to return to work, a self-described "optimist" with a newfound belief in God. So, many people would describe her as exceptionally lucky. Ms. Vulovic goes with our first assessment, "I'm not lucky. Everybody thinks I am lucky, but they are mistaken. If I were lucky I would never have had this accident" (Bilefsky, 2008).

It is not surprising that people often think and talk about unexpected events in different, sometimes contradictory ways. After all, these events are unpredictable and by definition mysterious and poorly understood. But, even beyond that, our minds do not seem to be designed to reason systematically about chance and uncertainty. Perhaps for evolutionary reasons, we are inclined to over-explain uncertain events and, even when we recognize they are inherently unpredictable, we have some queer notions about how they behave, including many superstitious beliefs (Sagan, 1997). Because we have natural misconceptions about uncertainty and randomness, this is one case in which learning about the rudiments of a technical framework—probability theory—can make a big difference in how we see the world. But without special training, no one thinks about the world in terms of probabilities. Rather, the world seems to be a bunch of events and objects glued together by causal relationships, and most of us think about causation deterministically and in terms of degrees of causal force, but not in terms of probabilities.

We have been careful not to refer to the world as probabilistic or random. Probability theory is a language we can use to *describe* the world or, more precisely, to describe the relationships among our beliefs about the world. It is an unfamiliar language to most people, with a special symbolic vocabulary and rules of grammar (see the Appendix for an introduction to probability theory). As we noted earlier, probability theory was not invented until recently in the history of Western civilization, and words like *probability* don't seem to have entered the English lexicon until the 17th century. (Lexicographers believe it was derived from the expression "approvable," e.g., a *probable* husband was originally an acceptable or morally "approvable" husband.)

Sometimes we do talk about chance, luck, probability, or randomness in everyday events—we say, "she was lucky," "it happened by chance," "that was a random event." But the most sensible interpretation of these expressions is that they indicate the state of knowledge within the mind of the person speaking. Harking back to a very wise essay on the nature of chance by the philosopher Poincaré (1914/1952), the events that we refer to in everyday life are all brought about by deterministic, physical processes. What singles out the events that we refer to as random, chance, or probabilistic is that the causal context is hidden, complex, or unknown to the person who describes the event as such. We can't specify the physical events that occurred to preserve Vesna Vulovic's life, but we believe that she survived because of physical conditions that *could* be specified, if we had enough information. If we'd been able to observe her fall, including the minute details concerning her contact with the ground and her internal body state immediately before contact, we should be able to account for her remarkable escape from death in terms of physical causality.

For another example, we refer to the toss of a fair coin as a random process and assign the (ideal) probability value of .50 to the event of *heads*, although we believe that the hidden biological and physical events that *cause* the outcome of the toss are all deterministic. In fact, skilled sleight-of-hand magicians, like the mathematician Persi Diaconis, have developed their manual skills to the point where they can execute apparently uncontrolled coin tosses and reliably produce the desired result, heads or tails (Bayer & Diaconis, 1992; Diaconis, Holmes, & Montgomery, 2007). Of course, there are levels of physical analysis, for example, at the quantum level, where scientists do not believe causality maps directly onto the mechanical principles of causality we experience. But we do not experience the world at that level, and it is a rare conversation that refers to those events.

Of course, there are parts of our environment that approximate the idealized behavior of theoretical random processes; events in casinos and lotteries are "caused" by deterministic physical processes, but the causal mechanism is so complex and the determinants of the events are so subtle that the best way to think about these situations is in terms of probability theory. An important message of this book is that we should use probability theory to organize our thinking about all judgments under uncertainty, even where we know much more (or less) about the relevant causes than we do in a casino. But we tend to deny the random components even in trivial events that we *know* to be the result of chance. There is a wonderful story about the winner of a national lottery in Spain. When interviewed about how he won, the winner said that he had deliberately selected a ticket that ended with the numbers 4 and 8. He explained, "I dreamed of the number 7 for seven straight nights. And 7 times 7 is 48" (Meisler, 1977).

7.2 Illusions of Control

In a clever series of experiments, Ellen Langer (1975) of Harvard University demonstrated that—automatically, without any conscious awareness—we often treat chance events as if they involve skill and are hence controllable. For example, gamblers tend to throw dice with greater force when they are attempting to roll high numbers than when they are attempting to roll lower numbers. Langer conducted a lottery in which each participant was given a card containing the name and picture of a National Football League player; an identical card was put into a bag; and the person holding the card matching the one drawn from the bag won the lottery. In fact, Langer conducted two lotteries. In one, the participants chose which player would constitute their ticket; in the other, players were assigned to the participants by the experimenter. Of course, whether or not the entrants were able to choose their own players had no effect on the probability of their winning the lottery, because the winning cards were drawn at random from the bag. Nevertheless, when an experimenter approached the participants offering to buy their card, those who had chosen their own player on the average demanded *more than 4 times as much money* for their card as did those with randomly assigned cards. Upon questioning, no one claimed that being allowed to choose a player influenced his or her probability of winning. The participants just *behaved* as if it had.

In another striking experiment, Langer and Susan Roth (1975) were able to convince Yale undergraduates that they were better or worse than the average person at predicting the outcome of coin tosses. The subjects were given rigged feedback that indicated they did not perform any better than at a chance level—that they were correct on 15 of 30 trials. What the experimenters did was manipulate whether the subjects tended to be correct toward the beginning of the 30-trial sequence or toward the end. Consistent with a primacy effect (or anchoring-and-[insufficient]-adjustment), those subjects who tended to be correct toward the beginning were apt to think of themselves as "better than average" at predicting, while those who did not do well at the beginning judged themselves to be worse. (Of course, due to random fluctuations, the probability of success in predicting the outcome of coin tosses cannot be expected to be invariant across a sequence as short as 30 trials.) In addition, "over 25% of the subjects reported that performance would be hampered by distraction and 40% of all the subjects felt that performance would improve with practice." Thus, not only do people behave as if they can control random events; they also express the conscious belief that doing so is a skill, which, like other skills, is hampered by distractions and improves with practice. It is important to remember that these subjects were from one of the most elite universities in the world, yet they treated the prediction of coin tosses as if it involved some type of ability, not just dumb luck.

Moreover, as with most everyday applications of psychology, practitioners like the managers of casinos and lotteries already have an intuitive understanding of these principles. Commercial games of chance often contain deceptive skill elements, deliberately designed to confuse the players about the skill and opportunity for control involved in games of chance. In most states, lottery players can choose the numbers they bet their money on, and the lotteries often have skill-evoking cover stories: "Hit a home run and win Major League bucks," "Just by buying a Bowling for Bucks ticket, you're a winner."

A more serious consequence of the illusion of control is revealed in our preference for driving over flying. At least part of this irrational—from a survival point of view—habit is due to the fact that we "feel in control" when

driving, but not when flying. The probability of dying in a cross-country flight is approximately equal to the probability of dying in a 12-mile drive—in many cases, the most dangerous part of the trip is over when you reach the airport (Sivak & Flannagan, 2003). Gerd Gigerenzer (2006) estimates that the post-9/11 shift from flying to driving in the United States resulted in an additional 1,500 deaths, beyond the original 3,000 immediate victims of the terrorist attacks.

One of the most compelling studies of the illusion of control demonstrated that it was related to consequential, poor performance in a real-world investment situation. Four British finance experts asked traders from four investment banks to play a computer game in which they attempted to influence the price of a fictional investment index (Fenton-O'Creevy, Nicholson, Sloane, & Willman, 2003). The movements of the index were completely independent of the actions by the trader-players—it was a random walk with a slight positive trend. The traders played the game for four rounds and rated their personal success in raising the index—because the index movements were independent of the actions of the traders, this is a measure of individual illusions of control. On average, the traders fell prey to the illusion that they had influenced the movement of the price index. More interesting, the level of individual illusions of control negatively predicted the traders' earnings and their managers' ratings of their talents and performance. Traders with a greater illusion of control earned substantially less than their more realistic peers ($100,000s); they contributed less to their bank's profits; and their managers rated them lower on risk management, analytical ability, and people skills.

7.3 Seeing Causal Structure Where It Isn't

A pernicious result of representative and scenario-based thinking is that they make us see structure (nonrandomness) where none exists. This occurs because our naïve conceptions of randomness involve *too much* variation—often to the point where we conclude that a generating process is *not* random, even when it represents an ideal random trial. Consider one of the simplest, most familiar processes we would describe as random, a coin toss. When asked to "behave like a coin" and to generate a sequence of heads and tails that would be typical of the behavior of a fairly tossed coin, most people produce too much alternation—nonrandomly too many heads–tails and tails–heads transitions. (They exhibit the same bias when shown sequences and asked to pick the "real coin" [Lopes, 1982].) Representativeness enters in because when we are faced with the task of distinguishing between random and nonrandom "generators" of events, we rely on our stereotype of a random process (analogous to our stereotype of a feminist or a bank teller or an art history major) and use *similarity* to judge or produce a sequence. Thus, when we encounter a truly random sequence, we are likely to decide it is *nonrandom* because it does not look haphazard enough—because it shows less alternation than our *incorrect* stereotype of a random sequence.

Suppose you're playing Langer and Roth's (1975) coin toss game with a fair coin (which you pulled out of your own pocket) and you are trying to predict the next outcome, heads or tails, after the coin has been tossed 8 times. Remarkably, the coin has come up heads on each toss, a run of 8 heads. If you're like most people, you'll have a feeling that tails is more likely on the ninth toss—you feel "it's due"—and you'd probably even bet some money on the prediction of tails. Another example of this feeling is the common, but incorrect advice about how to gamble: "When you're in Vegas and you see a roulette wheel come up with a run of three or more reds, bet black. You're sure to win." There is even a rationale for this belief: Nine heads (or reds) in a row is very rare; the odds are strongly against this happening ($[1/2]^9$ or 1/512 or approximately .002 for the coin, less for the roulette wheel), so if you're looking at 8 in a row, it's very unlikely you'll get 9 in a row. This intuition and the rationale are an error called the *gambler's fallacy*—the notion that "chances of [independent, random] events mature" if they have not occurred for a while. Fair coins and roulette wheels have no memories; the chance of each event is independent of all the other events in a sequence, and the probability of tails or red is constant.

Many people believe airplane accidents happen in "bunches"—usually threes. (One clinical psychologist we know cites such coincidences as evidence for "Jungian synchronicity.") Russell Vaught and Dawes obtained data from the FAA describing all commercial airline crashes between 1950 and 1970. They examined the number of days between the occurrences of the crashes. A totally random model begins with the assumption that the probability of a crash on any given day is a constant, p. Hence, the probability of a crash occurring the day following another crash is p. The probability that the next crash occurs on the second day subsequent to a crash is $(1 - p)p$, because there must be no crash on the succeeding day and then a crash on the next one. (Note that $[1 - p]p$ is less than p, a result that some people find counterintuitive, perhaps analogous to "Linda the feminist bank teller" from Chapter 5.) Similarly, the probability that the next crash will occur on the third day following a crash is $(1 - p)(1 - p)p = (1 - p)^2p$, and in general the probability that the next crash will occur on the *nth* succeeding day is $(1 - p)^{n-1}p$.

Examining all crashes and fatal crashes separately, Vaught and Dawes (unpublished research) discovered that the fit to the theoretical random prediction based on a constant p was almost perfect. Yet crashes seem to occur in "bunches." Why? Because $(1 - p)^j > (1 - p)^k p$ when $j < k$. Hence, truly random sequences actually contain "bunches" of events. The problem is that representative thinking leads us to conclude that such random patterns are *not* random. Instead, we hypothesize positive feedback mechanisms such as "momentum" to account for them. (Those of us hypothesizing "Jungian synchronicity" are in a minority.) While, for example, the maxim that "nothing succeeds like success or fails like failure" may be true, phony evidence for it can be found in the bunching of successes in patterns of people or organizations with high probabilities of success, and of failures in those with high probabilities of failure—even when the pattern is of independent events.

A well-defined situation in which people clearly see patterns that are not in the data is the *hot hand* phenomenon in basketball. The hot hand does not merely refer to the fact that some players are more accurate shooters than others, but to the (hypothetical) positive feedback performance process that makes players more likely to score after scoring and to miss after missing. (Note that the same term—a hot hand—is used to describe successful crap shooters, despite general acknowledgment that in well-run games, they cannot control the outcome of a roll.) Tom Gilovich, Robert Vallone, and Amos Tversky (1985) demonstrated empirically that the hot hand does not exist; that a success following a success is just as likely for an individual player as a success following a failure. At least, neither the floor shots of the Philadelphia '76ers, the free throws of the Boston Celtics, or the experimentally controlled floor shots of men and women on the Cornell varsity basketball teams showed evidence of a hot hand. But the players' *predictions* of their success showed a hot-hand effect, even though their performance did not. More than 90% of a sample of basketball players and sports reporters answered "yes" to the following question: Does a player have a better chance of making a shot after having just made his last two or three shots than he does after having just missed his last two or three shots?

Jay Koehler and Caryn Conley (2003) followed up the original studies with an analysis seeking nonrandom patterns in the NBA Long Distance Shootout Contest from 4 years of the competition. In this event, the best field goal shooters in the NBA attempt to score as often as possible within a 60-second time limit from the 3-point shot arc (the area of the court from which shots will count for 3 points instead of 2). Again, there was no evidence of nonrandomness. Even when the researchers conditioned their analysis on the announcers' assertions of "hotness," there were no patterns. It is notable that nonrandom streaks have been verified in other sports such as bowling, archery, billiards, and golf, suggesting that the statistics are sensitive enough to pick up patterns if they are there in the data. (It looks like there might be a bigger picture here: In nonreactive, uniform-playing-field sports, subtle sequential dependencies manifest themselves in performance; in chaotic, in-your-face, player-on-player reactive sports, there are no such patterns.)

These studies do not prove the *universal nonexistence* of the hot hand in basketball (which would be difficult to do, if you think about it), but their results imply that if it exists, it is small, unreliable, or very rare. The claim that any particular set of data is random is tenuous; it is more defensible to claim that a process that generates the data is random, in the sense that the observers of the data could not know the information necessary to predict the events in the data with any degree of specificity—that to these observers, the best description is a probabilistic or random process. The example of the hot hand in basketball is especially surprising because it is so easy to imagine a causal process that might generate the expected (but not observed) patterns. For example, one reply to Gilovich et al.'s (1985) and Tversky and Gilovich's (1989) original claim was that they had missed the true hot-hand pattern that was hidden in their data because they had ignored the timing of baskets. Patrick Larkey, Richard Smith, and Jay Kadane (1989) published a reanalysis consisting only of runs of shots occurring in close temporal proximity. They found one player, Vinnie "Microwave" Johnson of the Detroit Pistons, who departed from the random model. Microwave earned his nickname because of his reputation for streak shooting. However, Gilovich et al. (1985), in rebuttal, noted that the reanalysis found only one "hot" player, and that his statistically distinctive streakiness was due entirely to a single run of seven baskets. Then they pointed out that a review of the original game videotapes showed that the seven-basket run had *not* occurred. In fact, Microwave had a run of four baskets, missed a shot but scored on his own rebound, and then made one more score. After correcting for this data collection error, even Microwave did not depart from the random model.

Do 3 good weeks in a row indicate therapeutic success with a patient? Do 3 bad weeks indicate failure (or, more sanguinely, "coming to face problems")? Does losing three games in a row mean the coach should be fired? Or do three down quarters mean a CEO should be fined? No, no more than three heads in a row within a sequence of coin tosses indicate that the coin is biased. Yet, knowing the person's base rate for success—and expecting more alternation than in fact occurs if these weeks or quarters are totally *unrelated*—makes the temptation to impute causal factors to such strings almost overpowering, especially causal factors related to the actor's own behavior. (Another speculation is this: Could it be that the perceptual salience of "streaks" of hits and misses is the key temptation to see "hot" or "cold" patterns in performance? In professional basketball where fans talk avidly about "hot hands," the success rate for shots is well over 50%, and so, runs of "hits" would be common and violate our expectation for too many reversals [hit–miss and miss–hit transitions]. But consider baseball batting where the fans are likely to talk about "slumps" and where batting averages are all well below 50% so that runs of "misses" would be most salient.)

Why do we expect too much alternation? Tversky and Kahneman (1974) ascribe this expectation to the belief that even very small sequences must be representative of a population, that is, the proportion of events in a small frame must match—be representative of—the proportion in the population. When, for example, we are tossing a fair coin, we know that the entire population of possible sequences contains 50% heads; therefore, we expect 50% heads in a sample of four tosses. That requires more alternation than is found when each toss is independent. (At the extreme, 50% heads in a sequence of two tosses requires that each head is followed by a tail and vice versa.) Here, representative thinking takes us from schema to characteristic, rather than the reverse. Again, however, the basic belief is due to similarity matching—that is, to association. Moreover, the effect is compounded by our relatively brief span of attention—we want the short sequences *we can remember or imagine* to be representative (Kareev, 1992).

Consider the following question from a study by Tversky and Kahneman (1974):

> All families of six children in a city were surveyed. In 72 families, the exact order of births of boys and girls was G B G B B G. What is your estimate of the number of families in which the exact order of births was B G B B B B? What about the number of families with the exact order B B B G G G?

Almost everyone (80% or more of respondents) judges the latter birth sequences to be less likely than the first. However, all exact sequences are equally likely (the probability of any exact sequence is simply .5 × .5 × .5 × .5 × .5 × .5 or 0.015625, implying approximately 16 families out of a sample of 1,000 six-child families). Why do people have the strong intuition that G B G B B G is much more frequent? Because this short sequence captures all of our intuitions about what the result of a random process will look like: The sequence exhibits the correct proportion (half boys, half girls), it looks haphazard, and it has lots of alternation—in short, it looks "really random." (It is also the kind of sequence of hits and misses we would expect an ordinary basketball player to generate—too many short alternating runs, so that when we see a performance with longer runs, we are prone to say, "That can't be random. This player must really be 'hot.'") In contrast, the second sequence looks less likely because it violates *the law of small numbers* by having the wrong ratio of births (too many boys), while the third sequence is okay for proportion, but looks too orderly (three in a row, then three in a row).

Occasionally, this belief in alternation in random sequences (the gambler's fallacy that "red is due" because the last 6 outcomes on the roulette wheel were black) reaches ludicrous extremes. Consider, for example, the beginning of a "Dear Abby" letter:

> DEAR ABBY: My husband and I just had our eighth child. Another girl, and I am really one disappointed woman. I suppose I should thank God that she was healthy, but, Abby, this one was supposed to have been a boy. Even the doctor told me the law of averages were [sic] in our favor 100 to one.

A "graphic" example of the tendency to see patterns (and infer causes) where there surely weren't any occurred during the World War II bombing of London by German V-1 and V-2 missiles. London newspapers published maps of the missile impact sites (see Figure 7.1), and citizens immediately saw clusters of strikes and interpreted them with reference to the intentions of the hostile forces. What kind of stories did they tell to explain these patterns? The British citizens reasoned that the patterns they saw were the result of deliberate efforts to miss the areas of the city in which German spies lived. However, a classic probability modeling analysis demonstrated that the clusters were completely consistent with a random Poisson process–generating device, that there was no reason to infer a systematic motive or cause behind the patterns (see William Feller's classic textbook, *An Introduction to Probability Theory and Its Applications,* Vol. 1, pp. 160f, for a mathematical analysis).

A timely example of this tendency to infer causes for geographic patterns is part of the psychology of "cancer cluster" hysterias. During the past two decades, reports of communities in which there seem to be an unusual number of cancer incidents have soared (see Gawande, 1999). A community that notices an unusual number of cancers quite naturally looks for a cause in the environment—something in the water or the ground or the air. But investigating isolated neighborhood cancer clusters is almost always an exercise in futility. Public health agencies deploy thousands of "hot pursuit" studies every year in response to reports of raised local cancer rates. But Raymond Richard Neutra, California's chief environmental health investigator (in 1999), notes that among the hundreds of published reports of such investigations, *not one* has convincingly identified an environmental cause (cited in Gawande, 1999). And only one investigation resulted in the discovery of an unrecognized carcinogen. Neutra points out that in a typical Public Health Service registry of 80 different cancers, probability theory predicts you would expect to observe 2,750 of California's 5,000 census tracts to have statistically significant but random elevations of some form of cancer. So, if you check to see if your neighborhood has a statistically significant elevation in the rate of at least 1 of the 80 cancers, the chances are better than .50 it will—but that discovery will be perfectly consistent with a random model of the distribution of incidences, assuming *no* environmental causes. Commenting on the hot-pursuit investigations that result from neighborhood cluster alarms, Alan Bender (quoted in Gawande, 1999), an epidemiologist in the Minnesota Department of Health, says, "The reality is they're a total waste of taxpayer dollars."

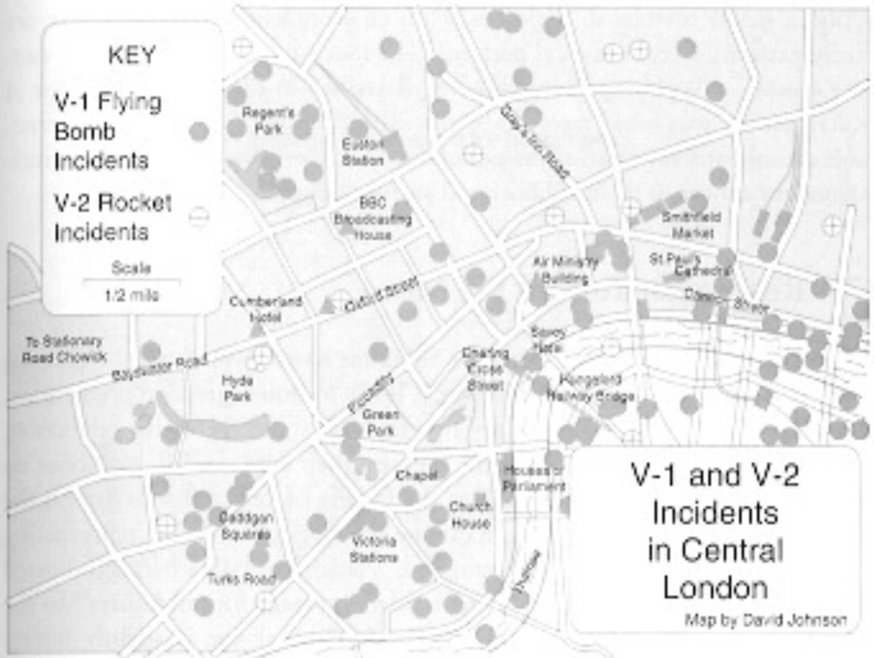

Figure 7.1 London V-1 and V-2 rocket impact pattern

But what can we do to maintain public trust and to identify true environmental health hazards? The fact that a random probability theory model is *consistent with* the patterns does not prove that there are no causal effects—It's that "How do you prove it doesn't exist anywhere, ever?" problem again. But we are wasting a lot of public funds responding to emotionally and symbolically important events and discovering many false correlations between clusters and their contexts. The strategy of analyzing individual clusters and looking for correlations with some (any) environmental cause is called the *Texas sharpshooter fallacy* by epidemiologists, after the story about a rifleman who shoots a cluster of bullet holes in the side of a barn and then draws a bull's-eye around the holes. This is a case where we should go with the advice of statistically sophisticated experts and only respond when there are good a priori reasons to hypothesize an environmental cause, or there are truly extraordinary statistical patterns. The much-publicized case of the cancer cluster in Woburn, Massachusetts, described in the book and movie *A Civil Action,* was never resolved by the identification of a scientifically credible causal pathway relating the pollutants from the Riley Tannery to the incidences of cancer in the neighborhood surrounding the factory.

7.4 Regression Toward the Mean

A final problem with representative thinking about events with a random (unknown causes) component is that it leads to non-regressive predictions. To understand why, it is necessary first to understand regressive prediction.

Consider very tall fathers. On the average, their sons are tall, but about an inch shorter than their fathers. Also, the fathers of very tall sons are on the average shorter than their sons. Examine the vertical solid line representing Tall Fathers in Figure 7.2. The average son's height for tall fathers is indicated by tracing the horizontal broken line labeled "average for tall fathers" to the ordinate—the y-axis—in the graph. (The horizontal line is slightly higher than the midpoint of the vertical line between the top and bottom edges of the ellipse representing "the data" because the distribution of sons' heights on that vertical dimension is probably not exactly symmetric, but is likely to have a longer tail downward toward shorter sons' heights.) Tracing this path for a typical "Tall Father" simply works through the logic of identifying the mean height for sons of such fathers and shows that the mean "regresses"—that is, it is less extreme than the extreme father's height. The difference between D and d' is an index of the degree of regression for this data set. Exactly the same abstract pattern in reverse is revealed if we work from Tall Sons, following the horizontal solid line for a typical Tall Son and tracing the vertical broken line path downward to the abscissa—the x-axis—for the average father's height for Tall Sons.

The British scientist Sir Francis Galton (1886) was the first to notice this relationship, which he labeled "filial regression towards mediocrity" (p. 246). At first, he thought the relationship was the result of some genetic process that made organisms shift toward average attributes, but after considering the reverse relationship (backward in time), he concluded it was a statistical property of all correlational relationships. The relationship is illustrated in Figure 7.2. What you see is a simple averaging effect. Because the heights of fathers and sons are not perfectly correlated (for whatever reasons), there is *regression. Non-regressive prediction* refers to people's tendency to miss the subtle regression relationship and to predict that extreme values will be associated with too-extreme values—as we will see in a moment.

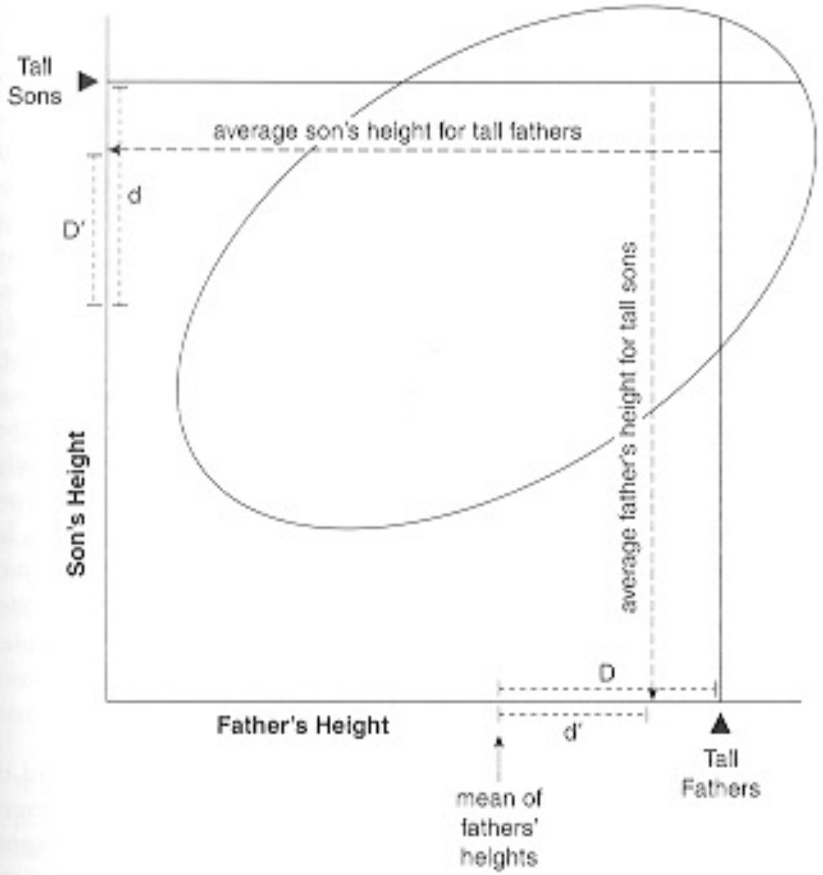

Figure 7.2 Illustration of statistical regression

Consider another example (based on the work of Quinn McNemar [1940], a psychologist who was one of the first to point out this statistical result and its implications for research on human behavior): Suppose that an intelligence test is administered to all the children in an orphanage on two occasions, a year apart. Assume, plausibly, that the group mean and standard deviation are the same on both tests; but that the correlation between scores on the two tests is not perfect (the actual correlation would be about +.80). Now consider only the children with the highest scores on the first test: Their scores on the second test will be on average lower. (Since the correlation is below +1.00, we expect some change; *since the two distributions of scores are the same,* the first test high scorers must get lower scores on average. The same was true for the children with the lowest scores: The average of the lowest-scoring children on the first test will be higher on the second. What if we reverse time and look backward from the second to the first test? The same relationships will apply: Extreme scores will be less extreme. Regression toward the mean is inevitable for scaled variables that are not perfectly correlated.

Perhaps it is easiest to understand regression by considering the extreme case in which we obtain perfect regression. Toss a fair coin 8 times; now toss it another 8 times. No matter how many heads are obtained in the first sequence of tosses, the expected (average) number of heads in the second sequence is 4. Because the coin is fair, the number of heads in the first sequence is totally uncorrelated with the number in the second—hence, average, of 4. That is total *regression to the mean.* As variables become more predictable from each other, there is less regression; for example, on average, the sons of very tall fathers are taller than the average person, but not as tall as their fathers. It is only when one variable is perfectly predictable from the other that there is no regression. In fact, the (squared value of the) standard correlation coefficient can be defined quite simply as the degree to which a linear prediction of one variable from another is not regressive. The technical definition of regression toward the mean is the difference between a perfect relationship (+/–1.00) and the linear correlation:

$$\text{regression} = \text{perfect relationship} - \text{correlation}$$

There are many examples of failure to appreciate regression toward the mean in everyday judgments. We are constantly surprised when an exceptional performance on Wall Street, a hit movie, a #1 pop song, or a sports achievement is followed by something more mediocre. The *Sports Illustrated cover jinx* is one of the classic examples. Readers noticed that when an athlete or a team was featured on the cover of *Sports Illustrated,* always for some exceptional achievement, the individual or team was likely to experience a slump in performance or some other misfortune afterward. Statistical analysis only served to reinforce the impression, and fans generated many plausible explanations for the phenomenon—the athlete became overconfident because of the publicity, the athlete was distracted by the media attention, and so forth. Of course, we know that most if not all of "the effect" was due to selecting extreme cases and observing regression toward the mean. No special explanation beyond noting "selection for exceptionality" is needed.

A classic academic example is provided by Horace Secrist's 1933 book, *The Triumph of Mediocrity in Business.* Secrist's thesis was that successful and unsuccessful businesses "tend towards mediocrity." The thesis is supported by hundreds of graphics showing that when businesses are selected in Year 1 for exceptional performance, on average the most successful become less successful and the least successful become more successful. Howard Hotelling, a prominent statistician, commented, "The seeming convergence is a statistical fallacy, resulting from the method of grouping. These diagrams really prove nothing more than that the ratios in question have a tendency to wander about." He points out that the true test of convergence toward mediocrity would be a consistent decrease in the variance among the groups over time—which was not observed. This same mistake was manifested in Tom Peters's and Robert Waterman's 1984 best-seller *In Search of Excellence.* These management consultants selected 43 exceptionally successful companies and reviewed the distinctive features that they believed made them "excellent." But, 5 years later, *BusinessWeek*'s cover story, "Oops! Who's Excellent Now?" pointed out that over one-third of the original, sampled-because-they-were-extreme companies were in financial difficulty or bankrupt.

In many cases, we are interested in the effects of some treatment on performance—an educational enrichment treatment for low-performing school-children, bonuses for high-performing employees, a dietary supplement for the least healthy. Again, there is a problem of separating the true effects of a treatment, applied only to extreme cases, from simple regression. Some of the subsequent errors can be quite subtle. For example, when Daniel Kahneman (Tversky & Kahneman, 1974) was explaining to Israeli Defense Force flight instructors in the mid-1960s that reward is a better motivator than punishment, he was told by one instructor that he was wrong.

> With all due respect, Sir, what you are saying is literally for the birds. I've often praised people warmly for beautifully executed maneuvers, and the next time they almost always do worse. And I've screamed at pupils for badly executed maneuvers, and by and large, the next time they improve. Don't tell me that reward works and punishment doesn't. My experience contradicts it.

This flight instructor had witnessed a regression effect. People tend to do worse after a "beautifully executed maneuver" because performance at one time is not perfectly correlated with performance the next (again, for whatever reason). Performances also tend to improve each time after "badly executed maneuvers"—once more, simply because performance is not perfectly correlated from one occasion to the next. (The easiest way to obtain an award for "academic improvement" is to be right near the bottom of the class the semester prior to the one for which such awards are given, and the way to be labeled an "underachiever" is to score brilliantly on an aptitude test.) Unfortunately, as the flight instructor anecdote illustrates, teachers who do not understand regression effects may be systematically reinforced (by regression to better performance) for punishing students and disappointed (by regression to worse performance) for rewarding them. (Regression alone may be a sufficient explanation for some people's preference, like the flight instructor's, for punishment over reward as a means of behavior control.)

Another unhappy by-product of our ignorance of the inevitability of regression effects is our overconfidence in the success of interventions like firing coaches and CEOs. Consider the prototypical situation: A team performs poorly during the first half of the season. The owner reacts by firing the coach, and the team performs better during the second half of the season. Should we attribute the improvement to the firing and replacement of the coach or to simple regression effects? After all, mid-season firings are usually conditioned on an extreme, poor performance. Absent an experiment in which coaches are randomly fired, we cannot be sure (and such an experiment is unlikely to be performed). But careful statistical analyses consistently show that most of the improvement is due to regression (Koning, 2003), and the same is true for the firing of business executives. (The reality in sports is that, if a team performs extremely poorly during the first half of the season, it is likely to have been pitted against stronger teams, and the second half will involve weaker opponents, exaggerating the apparent success of the replacement coach even further.)

The rational way of dealing with regression effects is to "regress" when making predictions. Then, if there is some need or desire to evaluate discrepancy (e.g., to give awards for "overachievement" or therapy for "underachievement"), compare the actual value to the *predicted* value—not with the actual value of the variable used to make the prediction. For example, to determine patient "improvement" by comparing Minnesota Multiphasic Personality Inventory (MMPI) profiles at time 1 and time 2, first correlate the profiles to determine a (regressed) predicted score for each patient at time 2; then compare the actual profile with this predicted score, not with the score at time 1. Otherwise, patients who have high (pathological) profiles at time 1 may be mistakenly labeled "improved" ("they had nowhere to go but down"), while those with normal MMPI profiles may be mistakenly regarded as unresponsive to treatment. Representative thinking, in contrast, leads to comparing discrepancies without regressing first, and the results are predictable. For example, "Of particular significance was the fact that those scoring highest on symptom reductions . . . were those whose symptoms were initially more severe, and who were the less promising candidates for conventional types of therapy" (Dawes, 1986). (While Dawes was a clinical psychologist trainee, he asked the psychologists and psychiatrists at the hospital to dichotomize patients whose improvement was above average at discharge and those whose improvement was below average. Those they categorized as above average in improvement had higher scores on most of the MMPI scales on admission—significantly higher on the major clinical ones.)

Regression toward the mean is particularly insidious when we are trying to assess the success of some kind of intervention designed to improve the state of affairs—like the flight instructor's efforts to improve student performance by intervening to punish poor performances. The worst case scenarios for understanding the effects of interventions occur when the intervention is introduced because "we've got a problem." For instance, it is almost impossible to accurately assess the causal effects of the introduction of a strict traffic enforcement program *after* a flurry of tragic traffic accidents, or the hiring of a new CEO *after* several poor corporate performances, or the hiring of a new coach after a losing streak. The chances are, the interventions are going to show improvements, and it is almost certain that some or most of the effect will be due to regression toward the mean.

7.5 Reflections on Our Inability to Accept Randomness

Some of the errors in judgment we have just described are probably not so surprising. Why would we be smarter than casino operators who have spent hundreds of years perfecting diabolical probability games to trap unwary customers? Or why wouldn't sports fans confuse conditions under which streaks do occur (in some sports events) with similar situations in which they do not? But the pervasive tendency to see much more structure than is actually present and to imagine we have much more control over events than we do in hundreds of important naturally occurring situations is still a puzzle. In the next chapter, we'll introduce the best remedy we know for these hard-to-eradicate bad habits—thinking like a probability theorist.

References

Bayer, D., & Diaconis, P. (1992). Trailing the dovetail shuffle to its lair. *Annals of Applied Probability, 2,* 294–313.

Bilefsky, D. (2008, April 26). Serbia's most famous survivor fears that recent history will repeat itself. *New York Times.* Retrieved June 20, 2009, from http://www.nytimes.com/2008/04/26/world/europe/26vulovic.html

Dawes, R. M. (1986). Representative thinking in clinical judgment. *Clinical Psychology Review, 6,* 425–441.

Diaconis, P., Holmes, S., & Montgomery, R. (2007). Dynamical bias in the coin toss. *Society for Industrial and Applied Mathematics Review, 49,* 211–235.

Feller, W. (1968). *Introduction to probability theory and its applications* (3rd ed.). New York: Wiley.

Fenton-O'Creevy, M., Nicholson, N., Sloane, E., & Willman, P. (2003). Trading on illusions: Unrealistic perceptions of control and trading performance. *Journal of Occupational and Organizational Psychology, 76,* 53–68.

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