2

Something Out of Nothing

The Misperception and Misinterpretation of Random Data

The *human understanding supposes a greater degree of order and equality in things than it really finds; and although many things in nature be sui generis and most irregular, will yet invest parallels and conjugates and relatives where no such thing is.*

Francis Bacon, *Novum Organum*

In 1677, Baruch Spinoza wrote his famous words, "Nature abhors a vacuum," to describe a host of physical phenomena. Three hundred years later, it seems that his statement applies as well to human nature, for it too abhors a vacuum. We are predisposed to see order, pattern, and meaning in the world, and we find randomness, chaos, and meaninglessness unsatisfying. Human nature abhors a lack of predictability and the absence of meaning. As a consequence, we tend to "see" order where there is none, and we spot meaningful patterns where only the vagaries of chance are operating.

People look at the irregularities of heavenly bodies and see a face on the surface of the moon or a series of canals on Mars. Parents listen to their teenagers' music backwards and claim to hear Satanic messages in the chaotic waves of noise that are produced.[1] While praying for his critically ill son, a man looks at the wood grain on the hospital room door and claims to see the face of Jesus; hundreds now visit the clinic each year and confirm the miraculous likeness.[2] Gamblers claim that they experience hot and cold streaks in random rolls of the dice, and they alter their bets accordingly.

The more one thinks about Spinoza's phrase, the better it fits as a description of human nature. Nature does not "abhor" a vacuum in the sense of "to loathe" or "to regard with extreme repugnance" (Webster's definition). Nature has no rooting interest. The same is largely true of human nature as well. Often we impose order even when there is no motive to do so. We do not "want" to see a man in the moon. We do not profit from the illusion. We just see it.

The tendency to impute order to ambiguous stimuli is simply built into the cognitive machinery we use to apprehend the world. It may have been bred into us through evolution because of its general adaptiveness: We can capitalize on ordered phenomena in ways that we cannot on those that are random. The predisposition to detect patterns and make connections is what leads to discovery and advance. The problem, however, is that the tendency is so strong and so automatic that we sometimes detect coherence even when it does not exist.

This touches on a theme that will be raised repeatedly in this book. Many of the mechanisms that distort our judgments stem from basic cognitive processes that are usually quite helpful in accurately perceiving and understanding the world. The structuring and ordering of stimuli is no exception. Ignaz Semmelweis detected a pattern in the occurrence of childbed fever among women who were assisted in giving birth by doctors who had just finished a dissection. His observation led to the practice of antisepsis. Charles Darwin saw order in the distribution of different species of finches in the Galapagos, and his insight furthered his thinking about evolution and natural selection.

Clearly, the tendency to look for order and to spot patterns is enormously helpful, particularly when we subject whatever hunches it generates to further, more rigorous test (as both Semmelweis and Darwin did, for example). Many times, however, we treat the products of this tendency not as hypotheses, but as established facts. The predisposition to impose order can be so automatic and so unchecked that we often end up believing in the existence of phenomena that just aren't there.

To get a better sense of how our structuring of events can go awry, it is helpful to take a closer look at a specific example. The

example comes from the world of sports, but the reader who is not a sports fan need not dismay. The example is easy to follow even if one knows nothing about sports, and the lessons it conveys are quite general.

THE MISPERCEPTION OF RANDOM EVENTS

> "If I'm on, I find that confidence just builds. . . . you feel nobody can stop you. It's important to hit that first one, especially if it's a swish. Then you hit another, and . . . you feel like you can do anything."
>
> —World B. Free

I must caution the reader not to construe the sentences above as two distinct quotations, the first a statement about confidence, and the second an anti-imperialist slogan. Known as Lloyd Free before legally changing his first name, World B. Free is a professional basketball player. His statement captures a belief held by nearly everyone who plays or watches the sport of basketball, a belief in a phenomenon known as the "hqtjiand." The term refers to the putative tendency for success (and failure) in basketball to be self-promoting or self-sustaining. After making a couple of shots, players are thought to become relaxed, to feel confident, and to "get in a groove" such that subsequent success becomes more likely. In contrast, after missing several shots a player is considered to have "gone cold" and is thought to become tense, hesitant, and less likely to make his next few shots.

The belief in the hot hand, then, is really one version of a wider conviction that ".success breeds success" and "failure breeds failure" in many walks of life. In certain areas it surely does. Financial success promotes further financial success because one's initial good fortune provides more capital with which to wheel and deal. Success in the art world promotes further success because it earns an artist a reputation that exerts a powerful influence over people's judgments of inherently ambiguous stimuli. However, there are other areas—gambling games immediately come to mind—where the belief may be just as strongly held, but where the phenomenon simply does not exist. What about the game of basketball? Does success in this sport tend to be self-promoting?

My colleagues and I have conducted a series of studies to answer this question.[3] The first step, as always, involved translating the idea of the hot hand into a testable hypothesis. If a player's performance is subject to periods of hot and cold shooting, then he should be more likely to make a shot after making his previous shot (or previous several shots) than after missing his previous shot. This implies, in turn, that a player's hits (and misses) should cluster together more than one would expect by chance. We interviewed 100 knowledgeable basketball fans to determine whether this constitutes an appropriate interpretation of what people mean by the hot hand. Their responses indicated that it does: 91% thought that a player has "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." In fact, when asked to consider a hypothetical player who makes 50% of his shots, they estimated that his shooting percentage would be 61% "after having just made a shot," and 42% "after having just missed a shot." Finally, 84% of the respondents thought that "it is important to pass the ball to someone who has just made several shots in arow."

To find out whether players actually shoot in streaks, we obtained the shooting records of the Philadelphia 76ers during the 1980-81 season. (The 76ers are the only team, we were told, who keep records of the order in which a player's hits and misses occurred, rather than simple cumulative totals.) We then analyzed these data to determine whether players' hits tended to cluster together more than one would expect by chance. Table 2.1 presents the relevant data. Contrary to the expectations expressed by our sample of fans, players were *not* more likely to make a shot after making their last one, two, or three shots than after missing their last one, two, or three shots. In fact, there was a slight tendency for players to shoot better after *missing* their last shot. They made 51% of their shots after making their previous shot, compared to 54% after missing their previous shot; 50% after making their previous two shots, compared to 53% after missing their previous two; 46% after making three in a row, compared to 56% after missing three in a row. These data flatly contradict the notion that "success breeds success" in basketball and that hits tend to follow hits and misses tend to follow misses.

We also examined each player's performance record to determine

Table 2.1 Probability of Making a Shot Conditioned on the Outcome of Previous Shots for Nine Members of the 76ers

Player	*P(x ooo)*	*P(x\oo]*	*P(xo)*	*P(x]*	*P(xx)*	*P(x xx)*	*P(x xxx)*	*r*
C. Richardson	.50	.47	.56	.50	.49	.50	.48	-.02
J.Erving	.52	.51	.51	.52	.53	.52	.48	.02
L. Hollins	.50	.49	.46	.46	.46	.46	.32	.00
M. Cheeks	.77	.60	.60	.56	.55	.54	.59	-.04
C. Jones	.50	.48	.47	.47	.45	.43	.27	-.02
A. Toney	.52	.53	.51	.46	.43	.40	.34	-.08
B. Jones	.61	.58	.58	.54	.53	.47	.53	-.05
S. Mix	.70	.56	.52	.52	.51	.48	.36	-.02
D. Dawkins	.88	.73	.71	.62	.57	.58	.51	-.14
Mean =	.56	.53	.54	.52	.51	.50	.46	-.04

NOTE: x = a hit; o = a miss, r - the correlation between the outcomes of consecutive shots

whether the number of streaks of various lengths exceeded the number to be expected if individual shots were statistically independent. Were there more streaks of, say, 4, 5, or 6 hits in a row than chance would allow? Were there more, for example, than the number of streaks of 4, 5, or 6 heads in a row that one observes when flipping coins? The relevant statistical tests indicated that there was no such tendency. A variety of additional, more complicated, analyses led to the same conclusion: A player's performance on a given shot is independent of his performance on previous shots. (It is interesting to note that an interview with eight members of the 76ers that year revealed that these very players believed that they tended to shoot in streaks.)

How can we reconcile the widespread belief in the hot hand with the startling disconfirmation provided by these data? Most people's first response is to insist that the belief is valid and the data are not. The hot hand exists, the argument goes, it just did not show up in our sample of data. Perhaps it did not appear because being hot is perfectly compensated for by a hot player's tendency to take more difficult shots or receive more attention by the defensive team. The hot hand may have been masked, in other words, by other phenomena that work in the opposite direction. To test such an alternative interpretation, one must examine players' performance records when the difficulty of the shot and the amount of defensive pressure have been held constant. The most direct way of doing so is to examine players' "free-throw" records—penalty shots taken in pairs from the same distance and without defensive pressure. If success promotes success, then we would expect a player's shooting percentage on his second shot to be higher after making his first shot than after missing his first. It is not. Our analysis of two seasons of free-throw statistics by the Boston Celtics indicate that the outcomes of consecutive free throws are independent. On average, the players made 75% of their second free throws after making their first, and 75% after missing their first.

Still unconvinced, a number of people have tried to salvage their belief in the hot hand by suggesting that perhaps we have not adequately captured what is meant by the term (our initial survey results notwithstanding). Perhaps players' hits and misses do not cluster together more than do heads and tails, but, unlike coin flips, the player can predict in advance whether he is likely to make the next shot. In other words, maybe the hot hand really refers to the predictability of hits and misses rather than the clustering together of success with success and failure with failure.

This too was tested and found wanting. We asked a group of college basketball players to take 100 shots from along an arc that was everywhere an equal distance from the basket. Before each shot the players chose either a risky or conservative bet corresponding to whether they felt more or less likely to make their upcoming shot. The results indicated that the players believed that they shot in streaks: They tended to make risky bets after hitting their previous shot and conservative bets after missing their previous shot. However, there was no correlation between the outcome of consecutive shots, and hence no connection between their bets and the outcome of the next shot. In other words, not only do players fail to shoot in streaks, but they cannot predict in advance whether they are likely to make a given shot. Even according to this revised definition, the hot hand does not seem to exist.

Why Players Seem *to Shoot in Streaks*. It is important to note that although a player's performance record does not contain more or longer streaks than chance would allow, it does not mean that the player's performance is chance *determined*. It is not. Whether a given shot is hit or missed is determined by a host of non-chance

factors, foremost among them being the skill of the offensive and defensive players involved. However, one factor that does *not* influence the outcome, or does not have any *predictable* influence, is the outcome of the previous shot(s). That is what our research shows.

This qualification aside, why do people believe in the hot hand when it does not exist? There are at least two possible explanations. The first involves the tendency for people's preconceptions to bias their interpretations of what they see. Because people have theories about how confidence affects performance, they may expect to see streak shooting even before watching their first basketball game. This preconception could then influence their interpretation and memory of the game's events. Streaks of successive hits or misses may stand out and be remembered, while sequences of frequent alternation between the two may go unnoticed and be forgotten. Or, the common occurrence of a shot popping out of the basket after having seemingly been made might be counted as a "near miss" if the player had made his last several shots, but as evidence of being extremely cold if the player had missed his last several shots.[4] (The biasing effects of people's theories and preconceptions is discussed more thoroughly in Chapter 4.)

A second explanation involves a process that appears to be more fundamental, and thus operates even in the absence of any explicit theories people might have. Psychologists have discovered that people have faulty intuitions about what chance sequences look like.[5] People expect sequences of coin flips, for example, to alternate between heads and tails more than they actually do. Because chance produces less alternation than our intuition leads us to expect, truly random sequences look too ordered or "lumpy." Streaks of 4, 5, or 6 heads in a row clash with our expectations about the behavior of a fair coin, although in a series of 20 tosses there is a 50-50 chance of getting 4 heads in a row, a 25 percent chance of five in a row, and a 10 percent chance of a streak of six. Because the average basketball player makes about 50% of his shots, he has a reasonably good chance of looking like he has the hot hand by making four, five, or even six shots in a row if he takes 20 shots in a game (as many players do).

To determine whether this general misconception of the laws of chance might be responsible for the belief in the hot hand, we showed basketball fans sequences of X's and O's that we told them represented a player's hits and misses in a basketball game. We also asked them to indicate whether each sequence constituted an example of streak shooting. For instance, one of the sequences was OXXXOXXXOXXOOOXOOXXOO, a sequence in which the order of hits and misses is perfectly random.* Nevertheless, 62% of our subjects thought that it constituted streak shooting.

Note that although these judgments are wrong, it is easy to see why they were made. The sequence above does *look* like streak shooting. Six of the first eight shots were hits, as were eight of the first eleven! Thus, players and fans are not mistaken in what they see: Basketball players do shoot in streaks. But the length and frequency of such streaks do not exceed the laws of chance. and thus do not warrant an explanation involving factors like confidence and relaxation that comprise the mythical concept of the hot hand. Chance works in strange ways, and the mistake made by players and fans lies in how they interpret what they see.

The Clustering Illusion. The intuition that random events such as coin flips should alternate between heads and tails more than they do has been described by statisticians as a "clustering illusion." Random distributions seem to us to have too many clusters or streaks of consecutive outcomes of the same type, and so we have difficulty accepting their true origins. The term illusion is well-chosen because, like a perceptual illusion, it is not eliminated by repeated examination.[6]

Consider the picture of St. Louis's Gateway Arch depicted in Figure 2.1.[7] The arch is one of the world's largest optical illusions: It appears to be much taller than it is wide, although its height and base are equal in length. More important, even when one is told that the height and base are equal, they still do not seem to be. The illusion cannot be overcome simply by taking another look; only an objective measurement will do. (The reader is encouraged to make the necessary measurements.)

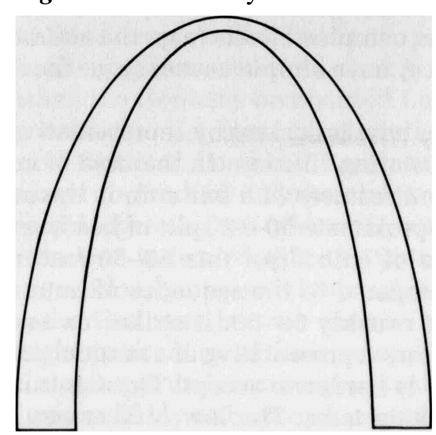
The reaction of the professional basketball world to our research on the hot hand is instructive in this regard. Do those close to the game give up their belief in the hot hand when confronted

* The sequence is random in the sense that there is no correlation between the outcomes of consecutive shots. The number of adjacent shots with the same outcome (i.e., **xx** or oo) in the sequence is equal to the number of adjacent shots with different outcomes (i.e., xo or ox).

with the relevant data? Hardly. Red Auerbach, the brains behind what is arguably the most successful franchise in American sports history, the Boston Celtics, had this to say upon hearing about our results: "Who is this guy? So he makes a study. I couldn't care less." Another prominent coach, Bobby Knight of the 1987 NCAA champion Indiana Hoosiers, responded by saying". . .there are so many variables involved in shooting the basketball that a paper like this really doesn't mean anything." These comments are not terribly surprising. Because a truly random arrangement of hits and misses contains a number of streaks of various lengths, the belief in the hot hand should be held most strongly by those closest to the game. Furthermore, simply hearing that the hot hand does not exist, or merely taking another look at the game is not sufficient to disabuse oneself of this belief. It is only through the kind of objective assessment we performed that the illusion can be overcome.

Judgment by Representativeness. In the grand scheme of things, whether or not basketball players shoot in streaks is not particularly important. What is important is the suggestion — conveyed with unusual clarity by the basketball example — that people chronically misconstrue random events, and that there may be other cases in which truly random phenomena are erroneously thought to be ordered and "real." If so, we arrive at the more critical question of why people expect random sequences to alternate more than

Figure 2.1 Gateway Arch



they do. Why, beyond noting that human nature abhors a vacuum, do people fall prey to the clustering illusion?

The best explanation to date of the misperception of random sequences is offered by psychologists Daniel Kahneman and Amos Tversky, who attribute it to people's tendency to be overly influenced by judgments of "representativeness."[8] Representativeness can be thought of as the reflexive tendency to assess the similarity of outcomes, instances, and categories on relatively salient and even superficial features, and then to use these assessments of similarity as a basis of judgment. People assume that "like goes with like": Things that go together should look as though they go together. We expect instances to look like the categories of which they are members; thus, we expect someone who is a librarian to resemble the prototypical librarian. We expect effects to look like their causes; thus, we are more likely to attribute a case of heartburn to spicy rather than bland food, and we are more inclined to see jagged handwriting as a sign of a tense rather than a relaxed personality.

Judgment by representativeness is often valid and helpful because objects, instances, and categories that go together often do in fact share a resemblance. Many librarians fit the prototype of a librarian—after all, the prototype came from somewhere. Causes often resemble their effects: All else being equal, "bigger" effects require "bigger" causes, complex effects stem from complex causes, etc. It is the *overapplication* of representativeness that gets us into trouble. All else is not always equal. Not all librarians are prototypical; Some big effects (e.g., an epidemic) have humble causes (e.g., a virus) and some complex effects (e.g., the alteration of a region's ecological balance) have simple causes (e.g., the introduction of a single pesticide).

It is easy to see how judgment by representativeness could contribute to the clustering illusion. In the case of coin flipping, one of the most salient features of a fair coin is the set of outcomes it produces—an approximate 50—50 split of heads and tails. In examining a sequence of coin flips, this 50—50 feature of the coin is automatically compared to the sequence of outcomes itself. If the sequence is split roughly 50-50, it strikes us as random because the outcome appears representative of a random generating process. A less even split is harder to accept. These intuitions are correct, but only in the long term. The law of averages (called the "law of large numbers" by statisticians) ensures that there will be close

to a 50-50 split after a large number of tosses. After only a few tosses, however, even very unbalanced splits are quite likely. There is no "law of small numbers."

The clustering illusion thus stems from a form of over-generalization: We expect the correct proportion of heads and tails or hits and misses to be present not only globally in a long sequence, but also locally in each of its parts. A sequence like the one shown previously with 8 hits in the first 11 shots does not look random because it deviates from the expected 50-50 split. In such a short sequence, however, such a split is not terribly unlikely.

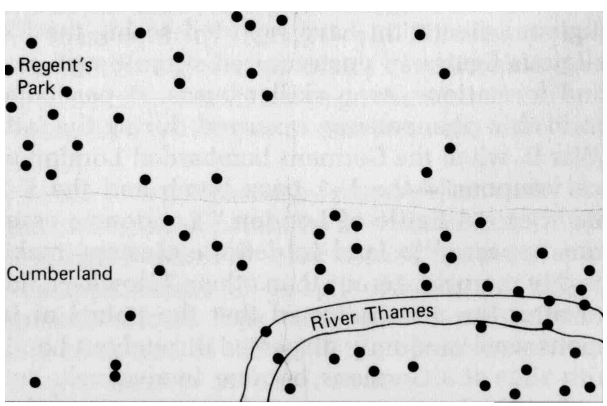
Misperceptions of Random Dispersions. The hot hand is not the only erroneous belief that stems from the compelling nature of the clustering illusion. People believe that fluctuations in the prices of stocks on Wall Street are far more patterned and predictable than they really are. A random series of changes in stock prices simply does not look random; it seems to contain enough coherence to enable a wily investor to make profitable predictions of future value from past performance. People who work in maternity wards witness streaks of boy births followed by streaks of girl birthslhat they attribute to a variety of mysterious forces like the phases of the moon. Here too, the random sequences of births to which they are exposed simply do not look random.

The clustering illusion also affects our assessments of spatial dispersions. As noted earlier, people "see" a face on the surface of the moon and a series of canals on Mars, and many people with a religious orientation have reported seeing the likeness of various religious figures in unstructured stimuli such as grains of wood, cloud formations, even skillet burns. A particularly clear illustration of this phenomenon occurred during the latter stages of World War II, when the Germans bombarded London with their "vengeance weapons"—the V-l buzz bomb and the V-2 rocket. During this "Second Battle of London," Londoners asserted that the weapons appeared to land in definite clusters, making some areas of the city more dangerous than others.[9] However, an analysis carried out after the war indicated that the points of impact of these weapons were randomly dispersed throughout London.[10] Although with time the Germans became increasingly accurate in terms of having a higher percentage of these weapons strike London, within this general target area their accuracy was sufficiently limited that any location was as likely to be struck as any other.

Still, it is hard not to empathize with those who thought the weapons fell in clusters. A random dispersion of events often does not look random, as Figure 2.2 indicates. This figure shows the points of impact of 67 V-l bombs in Central London.[11] Even after learning the results of the proper statistical analysis, the points do not look randomly dispersed. The lower right quadrant looks devastated and the upper left quadrant also looks rather hard hit; the upper right and lower left quadrants, however, appear to be relatively tranquil. We can easily imagine how the presence of special target areas could have seemed to Londoners to be an "irresistible product of their own experience."

A close inspection of Figure 2.2 sheds further light on why people "detect" order in random dispersions. Imagine Figure 2.2 being bisected both vertically and horizontally, creating four quadrants of equal area. As already discussed, this results in an abundance of points in the upper-left and lower-right quadrants, and a dearth of points in the other two areas. In fact, the appropriate statistical test shows this clustering to be a significant departure from an independent, random dispersion.* In other words, when the dispersion of points is carved up in this particular way, non-chance clusters can be found. It is the existence of such clusters, no doubt, that creates the impression that the bombs did not fall randomly over London.

Figure 2.2 Points of Impact of 67 V-l Bombs in Central London



* The appropriate test in this case is the chi-square test, and the obtained chi-square value is 20.69. The probability of obtaining a chi-square value this large by chance alone is less than 1 in 1,000.

But why carve the map this way? (Indeed, why conduct the statistical analysis only on the data from this particular area of London?) Why not bisect this figure with two diagonal lines? Bisected that way, there are no significant clusters.

The important point here is that with *hindsight* it is always possible to spot the most anomalous features of the data and build a favorable statistical analysis around them. However, a properly-trained scientist (or simply a wise person) avoids doing so because he or she recognizes that constructing a statistical analysis retrospectively capitalizes too much on chance and renders the analysis meaningless. To the scientist, such apparent anomalies merely suggest hypotheses that are subsequently tested on other, *independent* sets of data. Only if the anomaly persists is the hypothesis to be taken seriously.

Unfortunately, the intuitive assessments of the average person are not bound by these constraints. Hypotheses that are formed on the basis of one set of results are considered to have been proven by those very same results. By retrospectively and selectively perusing the data in this way, people tend to make too much of apparent anomalies and too often end up detecting order where none exists.

CEMENTING OUR MISPERCEPTIONS WITH CAUSAL THEORIES

The main thrust of these examples, and the major point of this chapter, lies in the inescapable conclusion that our difficulty in accurately recognizing random arrangements of events can lead us to believe things that are not true—to believe something is systematic, ordered, and "real" when it is really random, chaotic, and illusory. Thus, one of the most fundamental tasks that we face in accurately perceiving and understanding our world—that of determining whether there is a phenomenon "out there" that warrants attention and explanation—is a task that we perform imperfectly.

Furthermore, once we suspect that a phenomenon exists, we generally have little trouble explaining *why* it exists or what it means. People are extraordinarily good at ad hoc explanation. According to past resarch, if people are erroneously led to believe that they are either above or below average at some task, they can explain either their superior or inferior performance with little difficulty.[12] If they are asked to account for how a childhood experience such as running away from home could lead during adulthood to outcomes as diverse as suicide or a job in the Peace Corps, they can do so quite readily and convincingly.[13] To live, it seems, is to explain, to justify, and to find coherence among diverse outcomes, characteristics, and causes. With practice, we have learned to perform these tasks quickly and effectively.

A dramatic illustration of our facility with ad hoc explanation comes from research on split-brain patients. In nearly all of these patients, language ability is localized in the left cerebral hemisphere, as it is in most people. The one difference between split-brain patients and other individuals is that communication between the two hemispheres is prevented in the split-brain patient because of a severed corpus callosum. Imagine, then, that two different pictures are presented to the two hemispheres of a split-brain patient. A picture of a snow-filled meadow is presented to the non-verbal right hemisphere (by presenting it in the left visual field). Simultaneously, a picture of a bird's claw is presented to the verbal left hemisphere (by presenting it in the right visual field). Afterwards, the patient is asked to select from an array of pictures the one that goes with the stimuli he or she had just seen.

What happens? The usual response is that the patient selects two pictures. In this instance, the person's left hand (controlled by the right hemisphere) might select a shovel to go with the snow scene originally presented to the right hemisphere. At the same time, the right hand (controlled by the left hemisphere) might select a picture of a chicken to go with the claw originally presented to the left hemisphere. Both responses fit the relevant stimulus because the response mode—pointing—is one that can be controlled by each cerebral hemisphere. The most interesting response occurs when the patient is asked to explain the choices he or she made. Here we might expect some difficulty because the verbal response mode is controlled solely by the left hemisphere. However, the person generally provides an explanation without hesitation: "Oh, that's easy. The chicken claw goes with the chicken and you need a shovel to clean out the chicken shed."[14] Note that the real reason the subject pointed to the shovel was not given, because the snow scene that prompted the response is inaccessible to the left hemisphere that must fashion the verbal explanation. This does not stop the person from giving a "sensible" response: He or she examines the relevant output and invents a story to account for it. It is as if the left hemisphere contains an explanation module along with, or as part of, its language center—an explanation module that can quickly and easily make sense of even the most bizarre patterns of information.[15]

This work has important implications for the ideas developed in this chapter. It suggests that once a person has (mis)identified a random pattern as a "real" phenomenon, it will not exist as a puzzling, isolated fact about the world. Rather, it is quickly explained and readily integrated into the person's pre-existing theories and beliefs. These theories, furthermore, then serve to bias the person's evaluation of new information in such a way that the initial belief becomes solidly entrenched. Indeed, as the astute reader has probably discerned, the story of our research on the hot hand is only partly a story about the misperception of random events. As the common response to our research makes clear, it is also a story about how people cling tenaciously to their beliefs in the face of hostile evidence. In Chapter 4 we return to the subject of how people's theories and expectations influence their evaluation of evidence.

MISUNDERSTANDING INSTANCES OF STATISTICAL REGRESSION

An important lesson taught in nearly every introductory statistics course is that when two variables are related, but imperfectly so, extreme values on one of the variables tend to be matched by less extreme values on the other. This is the regression effect. The heights of parents and children are related, but the relationship is not perfect—it is subject to variability and fluctuation. The same is true of a student's grades in high school and in college, a company's profits in consecutive years, a musician's performance from concert to concert, etc. As a consequence, very tall parents tend to have tall children, but not as tall (on average) as they are themselves; high school valedictorians tend to do well in college, but not as well (on average) as they did in high school; a company's disastrous years tend to be followed by more profitable ones, and its banner years by those that are less profitable. When one score is extreme, its counterpart tends to be closer to the average. It is a simple statistical fact.*

The concept of statistical regression is not terribly difficult, and most people who take a statistics course can learn to answer correctly the standard classroom questions about the heights of fathers and sons, the IQs of mothers and daughters, and the SAT scores and grade point averages of college students. People have more difficulty, however, acquiring a truly general and deep understanding that whenever *any* two variables are imperfectly correlated, extreme values of one of the variables are matched, on the average, by less extreme values of the other. Without this deeper understanding, people encounter two problems when they venture out in the world and deal with less familiar instances of regression.

First, people tend to be insufficiently conservative or "regressive" when making predictions. Parents expect a child who excels in school one year to do as well or better the following year; shareholders expect a company that has had a banner year to earn as much or more the next. In each case, the predicted performance is simply matched to initial performance without taking into account the likely effects of regression. This tendency for people's predictions to be insufficiently regressive has been implicated in the high rate of business failures, in disastrous personnel hiring decisions, and

* To understand why regression occurs, consider the relation between a person's scores on the Scholastic Aptitude Test (SAT) on two occasions. Each score can be thought of as a reflection of the person's true ability level plus some "chance error" that either improves or lowers the observed result (e.g., some answers may have been mere guesses that turned out to be correct or incorrect, the room might be unusually noisy or quiet, the person might have slept poorly or well the previous evening, etc.). A very high score is more likely to be the result of a less extraordinary true ability that has been helped by chance error, than of an even more extraordinary true ability that has been hurt by it—simply because there are more of the former than the latter (truly extraordinary ability is rare by definition). As a consequence, an extraordinarily high score at one time will tend to be less extreme the next time because it is unlikely to be paired again with such a favorable chance error. To see this more clearly, consider the case in which someone receives the highest score possible on the SAT, 800 points. Because those who receive such scores cannot score any higher the next time, their scores on a subsequent test will either be the same (the person has true 800 "aptitude") or lower (the person has less "aptitude" but was lucky the first time). On average, then, the SAT scores of those getting an 800 the first time will be lower than 800 the second. Analogous logic explains why those who do poorly the first time tend to do better the second.

in non-conservative risk estimates made by certified public accountants.

A particularly striking demonstration of people's insensitivity to regression effects was provided by an experiment in which the participants were asked to predict the grade-point averages (GPAs) of ten hypothetical students on the basis of one of two types of information.[16] Some were given information that is perfectly predictive of GPA (the targets' GPA not in "raw" form such as "4.0," but in "percentile" form such as "99th percentile"). Others were given information that was described as less diagnostic of GPA (the targets' score on a test of sense of humor). Statistical theory dictates that the better one's basis of prediction, the less regressive one needs to be. Thus, those who based their estimates on the perfectly"predictive information need not have been regressive at all; in contrast, the estimates based on the students' sense of humor should have been regressed considerably (i.e., a nearly-average GPA should have been predicted for each student, regardless of the student's score on the relatively uninformative test of sense of humor).

That is not what happened. The predictions made by the respondents in the two groups were nearly identical, and only minimally regressive. Students who supposedly scored at the 90th percentile, for example, were predicted to have the same GPA, regardless of whether their percentile ranking referred to their GPA or their sense of humor. The regression effect was just not incorporated into the participants' predictions.

This tendency to make non-regressive predictions, like the clustering illusion, can be attributed to the compelling nature of judgment by representativeness. In this case, people's judgments reflect the intuition that the prediction ought to resemble the predictor as much as possible, and thus that it should deviate from the average to the same extent. The most representative son of a 6'5" father is one who is 6'5" himself—a height that is reached by only a minority of such fathers' sons. Once again, judgment by representativeness produces overgeneralization. In this case, people correctly recognize that if variables x and y are related, the value of x is helpful in predicting y, and that therefore relatively extreme values of y should be predicted for extreme values of x (e.g., we expect tall parents to have tall children, and our expectation is usually confirmed). However, this intuition is often taken too far, and the predictions made about y tend to be as *extreme* as the input variable x rather than regressed toward the average of y (e.g., few parents who are 6'5" have children as tall as they are).

A second, related problem that people have with regression is known as the regression fallacy. The regression fallacy refers to the tendency to fail to recognize statistical regression when it occurs, and instead to "explain" the observed phenomena with superfluous and often complicated causal theories. A lesser performance that follows a brilliant one is attributed to slacking off; a slight improvement in felony statistics following a crime wave is attributed to a new law enforcement policy. The regression fallacy is analogous to the clustering illusion: Both represent cases of people extracting too much meaning from chance events. By developing elaborate explanations for phenomena that are the predictable result of statistical regression, people form spurious beliefs about phenomena and causal relations in everyday life.

Examples of erroneous beliefs produced by the regression fallacy pervade many walks of life. There are many such examples in the sports world, for instance, one of the best being the widespread belief in the "*Sports Illustrated* jinx." Many individuals associated with the world of athletics believe that it is bad luck to be pictured on the cover of *Sports Illustrated* magazine.[17] Doing so is thought to spell doom for whatever success was responsible for getting oneself or one's team on the cover in the first place. Olympic medalist Shirley Babashoff, for example, reportedly balked at getting her picture taken for *Sports Illustrated* before the 1976 Olympics because of her fear of the jinx (she was eventually persuaded to pose when reminded that a cover story on Mark Spitz had not prevented him from winning seven gold medals in the previous Olympic games).

It does not take much statistical sophistication to see how regression effects may be responsible for the belief in the *Sports Illustrated* jinx. Athletes' performances at different times are imperfectly correlated. Thus, due to regression alone, we can expect an extraordinarily good performance to be followed, on the average, by a somewhat less extraordinary performance. Athletes appear on the cover of *Sports Illustrated* when they are newsworthy—i.e., when their performance is extraordinary. Thus, an athlete's superior performance in the weeks preceding a cover story is very likely to be followed by somewhat poorer performance in the weeks after. Those who believe in the jinx, like those who believe in the hot

hand, are mistaken, not in what they observe, but in how they interpret what they see. Many athletes do suffer a deterioration in their performance after being pictured on the cover of *Sports Illustrated,* and the mistake lies in citing a jinx, rather than citing regression as the proper interpretation of this phenomenon.

The regression fallacy also plays a role in shaping parents' and teachers' beliefs about the relative effectiveness of reward and punishment in producing desired behavior and learning. Psychologists have known for some time that rewarding desirable responses is generally more effective in shaping behavior than punishing undesirable responses.[19] However, the average person tends to find this" fact surprising, and punishment has been the preferred reinforcer for the majority of parents in both modern society[19] and in earlier periods.[20] One explanation for this discrepancy between common practice and the recommendation of psychologists is that regression effects may mask the true effectiveness of reward, and spuriously boost the apparent effectiveness of punishment. Rewards are most likely to be given following another person's extraordinarily good performance. However, regression guarantees that on the average such extraordinary performances will be followed by deterioration. The reward will thus appear ineffective or counter-productive. In contrast, because bad performances tend to be followed by improvement, any punishment meted out after a disappointing performance will appear to have been beneficial. Regression effects, in other words, serve to "punish the administration of reward, and to reward the administration of punishment."[21]

An intriguing demonstration of this phenomenon was provided by an experiment in which the participants played the role of a teacher trying to encourage a hypothetical student to arrive for school on time at 8:30 A.M.[22] A computer displayed tne "student's" arrival time, which varied from 8:20 to 8:40, for each of 15 consecutive days, one at a time. On each day, the participants were allowed to praise, reprimand, or issue no comment to the student. Predictably, the participants elected to praise the student whenever he was early or on time, and to reprimand him when he was late. The student's arrival time, however, was pre-programmed and thus was not connected to the subject's response for the previous day. Nevertheless, due to regression alone, the student's arrival time tended to improve (to regress toward 8:30) after he was punished for being late, and to deteriorate (again, by regressing to 8:30) after being praised for arriving early. As a result, 70% of the subjects concluded that reprimand was more effective than praise in producing prompt attendance by the student. Regression effects teach us specious lessons about the relative effectiveness of reward and punishment.

CODA

Perhaps the reader has anticipated how the two difficulties discussed in this chapter—the clustering illusion and the regression fallacy—can combine to produce firmly-held, but questionable beliefs. In particular, they may combine to produce a variety of superstitious beliefs about how to end a bad streak or how to prolong a good one. A modest "streak" of good or bad performance may be assigned too much significance initially, making its likely regression even more salient and in even greater need of explanation. An episode I witnessed during a recent trip to Israel provides a good example.

A flurry of deaths by natural causes in the northern part of the country led to speculation about some new and unusual threat. It was not determined whether the increase in the number of deaths was within the normal fluctuation in the death rate that one can expect by chance. Instead, remedies for the problem were quickly put in place. In particular, a group of rabbis attributed the problem to the sacrilege of allowing women to attend funerals, formerly a forbidden practice. The remedy was a decree that subsequently barred women from funerals in the area. The decree was quickly enforced, and the rash of unusual deaths subsided—leaving one to wonder what the people in this area have concluded about the effectiveness of their remedy.[23]

Examples like this illustrate how the misperception of random sequences and the misinterpretation of regression can lead to the formation of superstitious beliefs. Furthermore, these beliefs and how they are accounted for do not remain as isolated convictions, but serve to bolster or create more general beliefs—in this case about the wisdom of religious officials, the "proper" role of women in society, and even the existence of a powerful and watchful god.

3

Too Much from Too Little

The Misinterpretation of Incomplete and Unrepresentative Data

They *still cling stubbornly to the idea that the only good answer is a* yes answer. *// they say, "Is the number between 5,000 and 10,000?" and I say* yes, they *cheer; if I say* no, *they groan, even though they get exactly the same amount of information in either case.*

John Holt, *Why Children Fail*

"I've seen it happen." "I know someone who did." "You see JL it all the time." What these statements have in common is that they are often cited in support of a person's beliefs. "I know horoscopes can predict the future, because I've seen it happen." "I am convinced you can cure cancer with positive thinking because I know somebody who whipped the Big C after practicing mental imagery." "Of course there's a second-year slump, you see it all the time." Sometimes these statements are offered as justifications for the speaker's own beliefs; at other times they are designed to convince the listener of some important truth. In either case, they represent a conviction that a particular belief is warranted in light of the evidence presented.

Such convictions are on the right track. Evidence of the type mentioned in these statements is certainly *necessary* for the beliefs to be true. If a phenomenon exists, there must be some positive evidence of its existence—"instances" of its existence must be visible to oneself or to others. But it should be clear that such