CHAPTER 4

TESTIMONIALS AND CASE STUDY EVIDENCE

PLACEBO EFFECTS AND THE AMAZING RANDI

Cut to the *Oprah Winfrey Show,* one of the most popular audience participation television talk shows of the 1990s. Today's guest is Dr. Alfred Pontificate, director of the Oedipus Institute of Human Potential. Oprah moves among the audience, eliciting questions about the doctor's provocative new Theory of Birth Order, which is based on the idea that the course of one's life is irrevocably set by family interactions that are determined by birth order. The discussion inevitably turns from theoretical concerns to requests for explanations of personal events of importance to members of the audience. The doctor complies without much prodding.

For example, "Doctor, my brother is a self-destructive workaholic. He ignores his wife and family and places work-related problems above everything else. He has an ulcer and a drinking problem that he refuses to acknowledge. His family hasn't been on a real vacation in two years. He's headed for divorce and doesn't seem to care. Why has he chosen such a self-destructive course?"

To which the doctor replies, "What is his birth order, my dear?"

"Oh, he is the oldest of the children."

"Yes," the doctor says, "this is quite common. We see it often in the clinic. The underlying dynamics of a situation like this arise because parents transfer their life hopes and frustrations to their firstborn child. Through a process of unconscious wish transference, the child absorbs these hopes and frustrations, even if the parents never articulate them. Then, through the unconscious process that I call the dynamic expectation spiral, the aspirations of the parents become manifest as a pathological need for achievement in the child."

Although the audience on the *Oprah* show sometimes asks hostile questions when the guest challenges their beliefs, this rarely happens when a behavioral "expert" seems to confirm conventional wisdom. Once in a while, however, the show is animated by an audience member who questions the evidence behind the guest's declarations. In this case, an eager,

forthright questioner is in the studio. "But wait a minute, Doctor," the questioner begins. "My brother is a firstborn, too. My parents sent the bum to Harvard and told me to go to a two-year school to be a dental hygienist. So, this great brain of theirs drops out after one year, goes to some mountaintop in Colorado, and the last time we saw him he was weaving baskets! I don't understand what you're saying about firstborns."

The audience tenses for the confrontation, but alas, the doctor always wins in the end: "Oh, yes, I have seen many cases like your brother. Yes, I often meet them in my practice. They are people for whom the dynamic expectation spiral has short-circuited, creating an unconscious desire to thwart wish transference. Thus the individual's life develops in such a way as to reject conventional achievement aspirations." A hushed pause follows; then on we go to the next "case."

Of course, we are dealing with something quite familiar here. This is another example of the Benjamin Rush problem. This "theory" of birth order is structured so that no observation can disconfirm it. Because it is an unfalsifiable theory, the confirmations put forth to prove it are meaningless because nothing is ruled out by the theory.

However, our concern in this chapter is not with the theory itself, but with the nature of the evidence that is presented to support it. When pressed for evidence, Dr. Pontificate presents his own "clinical experience" or "case studies" as proof. This is an extremely common occurrence in the realm of media psychology. Talk shows and paperback book racks are full of psychological theories based on the clinical experience of the originator. Many of the therapies presented to the public through these outlets are backed by nothing more than the testimonials of individuals who have undergone them and consider themselves improved or cured. In this chapter, we shall develop a principle of great use to consumers of psychological information: Case studies and testimonials are virtually worthless as evidence for the evaluation of psychological theories and treatments. In this chapter, we will demonstrate why this is true, and we will also discuss the proper role of the case study in psychology.

STAGES OF SCIENTIFIC INVESTIGATION

The usefulness of case study information is strongly determined by how far scientific investigation is advanced in a particular area. The insights gained from case studies or clinical experience may be quite useful in the early stages of the investigation of certain problems, as indicators of which variables deserve more intense study. Case studies have played a prominent role in opening up new areas of study in psychology. Well-known examples occur in the work of Jean Piaget.

However, when we move from the early stages of scientific investigation, where case studies may be very useful, to the more mature stages of theory testing, the situation changes drastically. Case studies and testimonials are not useful at the later stages of scientific investigation because they cannot be used as confirming or disconfirming evidence in the test of a particular theory. The reason is that case studies and testimonials are isolated events that lack the comparative information necessary to rule out alternative explanations.

The problem of relying on testimonial evidence is that if testimonials accumulate to support virtually every therapy tried, then it is impossible to use them to support any *specific* remedy. All the competing remedies also have supporting testimonials. What we want to know, of course, is which remedy is *best*, and we cannot determine this by using testimonial evidence. As psychologist Ray Nickerson (1998) has said in his review of the cognitive processes we use to deceive ourselves: "Every practitioner of a form of pseudomedicine can point to a cadre of patients who will testify, in all sincerity, to having benefited from the treatment" (p. 192).

Nickerson's point is illustrated empirically in a study conducted by psychologist Anthony Greenwald and his colleagues (Greenwald, Spangenberg, Pratkanis, & Eskenazi, 1991). In this study, the authors tested the usefulness of subliminal self-help audiotapes (tapes that use messages below hearing threshold), which are commonly advertised in magazines and on television (Moore, 1995). They tested one tape program designed to improve memory and another designed to improve self-esteem. After taking memory and self-esteem tests, the subjects were given the tape and listened to it each day for a month (the amount of time that the advertisers of the tapes said was sufficient to produce the advertised effects). Some subjects were given a self-esteem tape labeled "self-esteem tape," and other subjects were given the memory tape labeled "memory tape." Importantly, however, two more groups of subjects were tested: one given a tape that was labeled "self-esteem tape" but had the content of the memory tape and another given a tape that was labeled "memory tape" but had the content of the self-esteem tape. These two conditions served as critical controls. What happened was that there was no improvement in actual memory or self-esteem. However, there were differences in the *self-perceptions* (testimonials) among the groups. Here, it was the *label* on the tape that was important, not the content. Both groups receiving a tape labeled "self-esteem" scored higher on the measure of self-perception of improvement in self-esteem (even though one group had received memory content), and both groups receiving a tape labeled "memory" scored higher on the measure of self-perception of improvement in memory (even though one group had received self-esteem content). In short, these tapes generated plenty of testimonials despite the fact that their content had absolutely no effect on memory or self-esteem (see also, Moore, 1995).

The idea of alternative explanations is critical to an understanding of theory testing. The goal of experimental design is to structure events so that support of one particular explanation simultaneously disconfirms other explanations. Scientific progress can occur only if the data that are collected rule out some explanations, as discussed in Chapter 2. Science sets up conditions for the natural selection of ideas. Some survive empirical testing and others do not. Those that remain are closer to the truth. This is the honing process by which ideas are sifted so that those that contain the most truth are found. But there must be selection in this process: Data collected as support for a particular theory must not leave many other alternative explanations as equally viable candidates. For this reason, scientists construct control or comparison groups in their experimentation. These groups are formed so that, when their results are compared with those from an experimental group, some alternative explanations are ruled out. How this is done will be a main topic in several later chapters.

Case studies and testimonials stand as isolated phenomena. They lack the comparative information necessary to prove that a particular theory or therapy is superior. It is thus wrong to cite a testimonial or a case study as support for a *particular* theory or therapy. Those who do so mislead the public if they do not point out that such evidence is open to a wide range of alternative explanations. In short, the isolated demonstration of a phenomenon may be highly misleading. This point can be illustrated more specifically by the example of placebo effects.

WHY TESTIMONIALS ARE WORTHLESS: PLACEBO EFFECTS

Virtually every therapy that has ever been devised in medicine and psychology has garnered supporters and has been able to produce individuals who will testify sincerely to its efficacy. Medical science has documented testimonials to the curative powers of swine teeth, crocodile dung, powdered Egyptian mummy, and many other even more imaginative remedies (Harrington, 1997; Shapiro, 1960). In fact, it has long been known that the mere suggestion that treatment is being administered is enough to make many people feel better.

The tendency of people to report that any treatment has helped them, regardless of whether it has a real therapeutic element, is known as the *placebo effect* (Dodes, 1997; Ernst & Abbot, 1999; Harrington, 1997; Roberts, Kewman, Mercier, & Hovell, 1993). The concept of the placebo effect was well illustrated in the movie *The Wizard of Oz.* The wizard did not *actually* give the tin man a heart, the scarecrow a brain, and the lion courage, but they all felt better nevertheless. In fact, because it is only in the last hundred years or so that medical science has developed a substantial number of treatments that actually have therapeutic efficacy, it has been said that "prior to this cen-

tury, the whole history of medicine was simply the history of the placebo effect" (Postman, 1988, p. 96).

We can illustrate the concept of a placebo effect by considering biomedical research, where all studies of new medical procedures must include controls for placebo effects. Typically, if a new drug is being tested on a group of patients, an equivalent group will also be formed and given a pill that does not contain the drug (a placebo). Neither group will know what they are receiving. Thus, when the two groups are compared, the placebo effect—that is, the tendency to feel better when any new treatment is introduced—is controlled for. It would not be sufficient merely to show that a percentage of patients receiving the new drug report relief from their symptoms, because in the absence of a control group it would be impossible to know what percentage is reporting relief due to a placebo effect rather than to the efficacy of the drug itself. One study (see Bok, 1974) of patients with a variety of ailments indicated that about 35 percent reported satisfactory relief after receiving a placebo (see also Roberts et al., 1993). A recent study suggested that the placebo effect associated with taking the popular antidepressant drug Prozac is twice as large as the effect of the drug itself (Bower, 1996b). Placebo effects can be very powerful, so powerful that there have even been reports of people who have become addicted to placebo pills (Bok, 1974; Ernst & Abbot, 1999), needing more and more to maintain their state of health!

Of course, in actual research on drug therapies the placebo control is not a pill containing nothing but instead is one containing the best currently known agent for the condition. The issue isolated by the experimental comparison is whether the new drug is superior to the best one currently available.

Placebo effects are implicated in all types of psychological therapy (Wampold, et al., 1997). Many people with psychological problems of mild to moderate severity report improvement after receiving psychotherapy. However, controlled studies have demonstrated that some proportion of this recovery rate is due to a combination of placebo effects and the mere passage of time, often termed *spontaneous remission.* Most therapeutic treatments are some unknown combination of an active therapeutic component and a placebo effect. As Dodes (1997) notes, "Even serious diseases have periods of exacerbation and remission; arthritis and multiple sclerosis are prime examples. There are even cases of cancer inexplicably disappearing" (p. 45). He also cautions that a positive response to a placebo does not mean that a patient's problem was imaginary and warns that, contrary to popular belief, placebos can be harmful: "Placebo responses can 'teach' chronic illness by confirming and/or reinforcing the delusion of imagined disease. Patients can become dependent on nonscientific practitioners who employ placebo therapies" (Dodes, 1997, p. 45).

In studies of psychotherapy effectiveness, it is often difficult to determine exactly how to treat the placebo control group, but these complications should not concern us here. Instead, it is important to understand why

researchers are concerned about separating true therapeutic effects from placebo effects and spontaneous remission. A study of therapeutic effectiveness by Gordon Paul (1966, 1967) provides an example of what the outcomes of such investigations reveal. Paul studied groups of students who suffered from maladaptive anxiety in public-speaking situations. Of an experimental group that received desensitization therapy specific to their speech anxiety problem, 85 percent showed significant improvement. A placebo group received a pill that they were told was a potent tranquilizer but that was actually a bicarbonate capsule. Of this group, 50 percent displayed significant improvement. Of a third group that did not receive any therapy at all, 22 percent also displayed significant improvement. Thus, it appears that the spontaneous remission rate was 22 percent for this particular problem, that an additional 28 percent of the subjects displayed improvement due to generalized placebo effects (50 percent minus 22 percent), and that the desensitization therapy did have a specific effect over and above that of the placebo (85 percent compared to 50 percent).

Like the Paul study, other research has shown that psychotherapies do have a positive effect over and above what would be expected purely as the result of a placebo (Landman & Dawes, 1982; Lipsey & Wilson, 1993; Prioleau, Murdock, & Brody, 1983; Smith, Glass, & Miller, 1980). But experiments using placebo controls have demonstrated that merely citing the overall percentage of people who report improvement vastly overestimates the degree of improvement that is uniquely due to the particular treatment. Thus, the statements of a group of patients testifying that they were helped are not sufficient to justify belief in a particular therapy. Instead, it must be shown that the placebo effect has been accounted for. The same is true for drug therapies as well. Kirsch and Lynn (1999) discuss a meta-analysis (see Chapter 8) in which the effect size (a measure of the potency of a variable in standard deviation units) of the pretreatment to posttreatment in patients receiving antidepressant medication was 1.55. However, the effect size for those receiving a placebo control was 1.16 and the effect size for a completely untreated group was .37. Thus, the authors concluded that only about 25 percent of the response to the medication was "a true drug effect, another 25% may be due to the natural history of the condition, and 50% is an expectancy effect" (p. 505).

The problem here is that testimonials are just too easy to generate. Cornell University psychologist Thomas Gilovich (1991) noted that, "with the body so effective in healing itself, many who seek medical assistance will experience a positive outcome even if the doctor does nothing beneficial. Thus, even a worthless treatment can appear effective when the base-rate of success is so high" (p. 128). As Peter Medawar (1967) described the situation, "If a person a) is poorly, b) receives treatment intended to make him better, and c) gets better, then no power of reasoning known to medical science can convince him that it may not have been the treatment that restored his health" (p. 14). In short, placebo effects are potentially occurring whenever a thera-

peutic intervention is undertaken, *regardless of the efficacy of the intervention.* The problem is that placebo effects are so potent that, no matter how ludicrous the therapy one uses, if it is administered to a large group of people a few will be willing to give a testimonial to its efficacy (the early-morning whack-on-the-head therapy—use it every day and you'll feel better! Send $10.95 for your special, medically tested rubber hammer).

But we really should not joke about such a serious matter. Unwarranted reliance on testimonials and case study evidence may have disastrous consequences. Members of a research team that contributed to the modern conceptualization of Tourette syndrome as an organically based disorder (Shapiro et al., 1978; see Chapter 2) pointed out that inappropriate reliance on case study evidence helped to perpetuate the unfalsifiable psychoanalytic explanations of the syndrome that impeded true scientific progress in investigating the nature of the disorder:

> Unfortunately, these theories, invested with strong emotional feelings, then lead to specific recommendations for treatment.... Those investigators interested in studying psychological factors should be encouraged to do so. Judgments about the results of such investigations, however, should be based on established principles of scientific investigation, or, in our opinion, on a data oriented approach. *Data from a single case study or clinical impressions should be acknowledged as opinion or hypotheses for future investigation.* [italics added] (p. 363)

An editorial in the *New England Journal of Medicine* illustrates what practitioners in the medical sciences believe to be the place of the case study and the testimonial in medicine "If, for example, the *Journal* were to receive a paper describing a patient's recovery from cancer of the pancreas after he had ingested a rhubarb diet...we might publish a case report—not to announce a remedy, but, only to suggest a hypothesis that should be tested in a proper clinical trial. In contrast, anecdotes about alternative remedies (usually published in books and magazines for the public) have no such documentation and are considered insufficient in themselves as support for therapeutic claims" (Angell & Kassirer, 1998, pp. 839–840).

THE "VIVIDNESS" PROBLEM

It is fine to point out how the existence of placebo effects renders testimonials useless as evidence, but we must recognize another obstacle that prevents people from understanding that testimonials cannot be accepted as proof of a claim. Social and cognitive psychologists have studied what is termed the *vividness effect* in human memory and decision making (see Baron, 1998; Nisbett & Ross, 1980). When faced with a problem-solving or decision-making situation, people retrieve from memory the information that seems relevant

to the situation at hand. Thus, they are more likely to use the facts that are more accessible to solve a problem or make a decision. One factor that strongly affects accessibility is the vividness of information.

The problem is that there is nothing more vivid or compelling than sincere personal testimony that something has occurred or that something is true. The vividness of personal testimony often overshadows other information of much higher reliability. How often have we carefully collected information on different product brands before making a purchase, only to be dissuaded from our choice at the last minute by a chance recommendation of another product by a friend or an advertisement? Car purchases are a typical example (Nisbett & Ross, 1980). We may have read surveys of thousands of customers in *Consumer Reports* and decided on car X. After consulting the major automotive magazines and confirming that the experts also recommend car X, we feel secure in our decision—until, that is, we meet a friend at a party who knows a friend who knows a friend who bought an X and got a real lemon, spent hundreds on repairs, and would never buy another. Obviously, this single instance should not substantially affect our opinion, which is based on a survey of thousands of owners and the judgment of several experts. Yet how many of us could resist the temptation to overweight this evidence?

The auto purchase situation illustrates that the problems created by vivid testimonial evidence are not unique to psychology. For example, Stephen Budiansky (1984), Washington correspondent of the British science journal *Nature,* has summarized the situation in medicine with a statement that reinforces many of the points about science that we have discussed in previous chapters:

> Science eschews the personal. Although it is commonplace to ascribe this tendency to some fundamental coldness on the part of scientists, in fact it is really one of the great intellectual triumphs of the 20th century that scientists have learned to discount the experiences of individuals when searching for cause and effect in the natural world. The health sciences have had a particularly rough time of it; people tend to get sick for a variety of reasons and people more often than not get better no matter what "cure" is prescribed. The apparently undying popularity of quack remedies, invariably supported by testimonials from satisfied customers, is vivid proof of how hard we find it to look beyond personal experience. (p. 7)

Instances of how vividness affects people's opinions are not hard to find. Reporter Haynes Johnson (1991) wrote of how President Reagan came to recognize the severity of the AIDS problem:

> [Reagan] had not realized the seriousness of AIDS until July 1985, when he saw a news report disclosing that the actor Rock Hudson had died of the disease. This was more than five years after AIDS had been identified, thousands of Americans had been infected, and AIDS had been the subject of intense na-

> tional publicity. When Reagan saw the news report about Hudson's death, he asked [Brigadier] General Hutton [a former doctor] to tell him about the disease. (p. 454)

In short, the constant news reports and statistics on the increasing numbers of AIDS deaths had not attracted the president's attention, but the report of a single person who was known to him did. Similarly, writer Michael Lewis (1997) describes how conservative commentator George Will—a notorious opponent of government regulation—published a column calling for mandatory air bags after seeing a death in a car crash outside of his home.

Imagine that you saw the following headline one Friday morning in your newspaper: "Jumbo Jet Crash: 413 Killed." Goodness, you might think, what a horrible accident. What a terrible thing to happen. Imagine though, that the following Thursday you got up and your newspaper said, "Another Jumbo Jet Disaster: 442 Die." My God, you might think. Not another disaster! How horrible. What in the world is wrong with our air traffic system. And then imagine—please imagine as best you can—getting up the following Friday and seeing in the paper: "Third Tragic Airline Crash: 431 Dead." Not only you but the nation would be beside itself. A federal investigation would be demanded. Flights would be grounded. Commissions would be appointed. Massive lawsuits would be filed. *Newsweek* and *Time* would run cover stories. It would be the lead item on television news programs for several days. Television documentaries would explore the issue. The uproar would be tremendous.

But this is not an imaginary problem. It is real. A jumbo jet *does* crash every week. Well, not one jet, but a lot of little jets. Well, not little jets really, but little transportation devices. These devices are called passenger automobiles. And approximately 457 people die in them *each week* in the United States (23,800 people each year—37,500 if we count trucks and motorcycles; National Safety Council, 1990), enough to fill a jumbo jet.

A jumbo jet's worth of people die in passenger cars on our nation's highways every week, *yet we pay no attention*. The "Jumbo Jet's Worth of People Who Die" are not presented to us in a vivid way by the media. Hence, the 457 people who die *each week* in passenger cars (plus the additional 263 who die *each week* in trucks and on motorcycles) have no vividness for us. We don't talk about them at the dinner table as we do when a jet goes down and kills a lot of people. We do not debate the safety and necessity of car travel as we would the safety of the air traffic system if a jumbo jet crashed every week killing 400 people each time. The 457 are not on the news because they are distributed all over the country and thus are a statistical abstraction to most of us. The media do not vividly present to us these 457 deaths because they do not happen in the same place. Instead, the media present to us (occasionally) a number (e.g., 457 per week). This *should* be enough to get us thinking, but it is not. Driving automobiles is an extremely dangerous activity, however

it is measured (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; National Safety Council, 1990), yet there has never been a national debate about its risk relative to the benefits involved. Is this an acceptable toll for a suburban lifestyle that demands a lot of driving? We never ask the question because no problem is recognized. No problem is recognized because the cost is not presented to us in a vivid way, as is the cost of airline crashes.

Think of the absurdity of the following example. A friend drives you 20 miles to the airport where you are getting on a plane for a trip of about 750 miles. Your friend is likely to say, "Have a safe trip," as you part. This parting comment turns out to be sadly ironic, because your friend is *three times more likely to die in a car accident on the 20-mile trip back home than you are on your flight of 750 miles* (National Safety Council, 1990). It is the vividness problem that accounts for the apparent irrationality of person A's wishing person B safety, when it is person A who is in more danger.

Misleading personal judgments based on the vividness of media-presented images are widespread. One study (MacDonald, 1990; see also Cole, 1998) surveyed parents to see which risks to their children worried them the most. Parents turned out to be most worried about their children's being abducted, an event with a probability of 1 in 700,000. In contrast, the probability of their child's being killed in a car crash, which the parents worried about much less, is well over *100 times more likely* than their being abducted (Paulos, 1988). Of course, the fears of abduction are mostly a media-created worry. The results actually suggest that, largely because of perceptions skewed by vividness effects, "American parents seem to worry about events that are least likely to happen" (MacDonald, 1990). One of the researchers lamented that this focus on worries that are "currently fashionable" misdirects the attention of parents and leads them to "ignore areas where they could have more impact, like school performance, television viewing habits, drug use and car safety" (MacDonald, 1990).

Writer Katherine Dunn (1993) recounted the fear she felt as a parent in the mid-1980s when stories of abducted children were sweeping the country. Rumors were spread that as many as 70,000 children had been snatched in malls and used by pornographers and/or tortured by strangers. Dunn reported tracking down the actual evidence by calling the FBI. It turned out that the number 70,000 referred not to abductions but to the number of runaways and children involved in custody disputes. Regarding children being abducted by strangers—which was the heart of the rumor sweeping the country—the FBI had recorded seven such cases nationwide. It was obviously not the number of these cases that had prompted such parental fears, but the vividness of the descriptions of the harm to the children. The vastly greater danger to any child in the passenger seat of an automobile (even when buckled up) was simply not as vivid.

The previous anecdote calls to mind science writer K. C. Cole's (1998) description of the ridiculous image of a woman driving down the street with

a young child romping in the front seat, arriving at a shopping mall, and then getting out and grabbing the child's hand very tightly as she worries about child kidnappers. Cole discusses some of the reasons why people misassess risk. One factor involved is that people exaggerate risks that are perceived to be beyond their control. This is one reason why airline accidents—with people strapped in seats and dependent on the skills of others, on the performance of technology, and on the weather—seem so unacceptable to people. Cole (1998) reports on a poll taken after a TWA crash in which a large majority of people were willing to pay $50 more for a round-trip airline ticket if it increased safety. Yet the same people resist safety features in automobiles that would provide a much greater increment in safety at a much lower cost.

Writer Peter Boyer (1999) describes how similar misperceptions of risk are fostered by the gun lobby in the United States which tries to keep the public focused on vivid cases of "intruders" coming through the doors of households. The not-so-subtle subtext here is that one lowers one's risk by having a gun to protect yourself. Boyer (1999) points to the irony that the gun industry tries to focus attention on "guns in the hands of bad people" when the actual statistics show that the real problem is "guns in the hands of *good* people." Criminals do not account for most gun deaths in this country. There are actually more suicides with guns than there are homicides with guns. Most gun deaths are unintentional shootings and suicides—which is why research indicates that bringing a gun into a home actually *increases* family risk.

The Overwhelming Impact of the Single Case

Psychologists have extensively studied the tendency for people's judgments to be dominated by a single, salient example when more accurate information is available. Hamill, Wilson, and Nisbett (1980) showed subjects a taped interview with a prison guard. Some subjects viewed interviews with a guard whose responses and manner suggested that he was a truly humane individual. Others viewed an interview with an extremely inhumane and disagreeable guard. The interviews modified opinions about guards in a positive or negative direction, depending on which interview had been viewed. More interesting was the fact that half the subjects received information indicating that the interview they had witnessed was a part of a large study of prison guards and that the guard they had seen was highly typical of all guards in the prison system. The other subjects were told that the guard they had viewed was highly atypical of all guards and that his behavior and opinions were very extreme, either positively or negatively, depending on the interview. The information about whether the interview they had witnessed was typical or not had no effect on the subjects' opinions about prison guards. Knowledge of the statistical reliability of the interview was overwhelmed by the reactions to the interview itself.

Wilson and Brekke (1994) demonstrated how insidious the vividness problem is and also how it influences actual consumer behavior. They investigated how people were influenced by two different types of information about two different brands (brand A and brand B) of condom. One type of information was a survey and analysis in *Consumer Reports* magazine, and the other was the opinions of two university students about their preferences for condom brands. First, Wilson and Brekke surveyed a group of subjects on which type of information they would want to be influenced by. Over 85 percent of the subjects said that they would *want* to be more influenced by the *Consumer Reports* article than by the opinions of the two students. A similar group of subjects were then recruited for a study in which they were told that they would be given, free of charge, some condoms of their own choosing. The subjects were told that they could consult either or both of two types of information: a survey and analysis in *Consumer Reports* magazine and the opinions of two university students about their preferences. Even though less than 15 percent of a similar group of subjects *wanted* to be influenced by the opinions of the two students, 77 percent of the subjects requested both types of information. Apparently the subjects could not resist seeing the testimonials even though they did not believe that they should be affected by them. And they were indeed affected by them. When the subjects chose to see both types of information and the recommendations of the two sources of information differed, 31 percent of the subjects chose the brand of condom recommended in the student testimonials over the brand recommended by *Consumer Reports*.

Another example of how people respond differently to vivid anecdotal information comes from the media coverage of the Vietnam War in the mid to late 1960s. As the war dragged on and the death toll of Americans killed continued without an end in sight, the media took to reporting the weekly number of American service personnel who had been killed that week. Week after week, the figure varied between 200 and 300, and the public, seemingly, became quite accustomed to this report. However, one week a major magazine published a spread, running on for several pages, of the individual pictures of those persons who had died in the previous week. The public was now looking, concretely, at the approximately 250 individual lives that had been lost in a typical week. The result was a major outcry against the toll that the war was taking. The 250 pictures had an effect that the weekly numbers had not had. But we, as a society, must overcome this tendency not to believe numbers—to have to *see* everything. Most of the complex influences on our society are abstractions that are accurately captured only by numbers. Until the public learns to treat abstractions as seriously as images, public opinion will be as fickle as the latest image to flicker across the screen.

But it is not only the public that is plagued by the vividness problem. Experienced clinical practitioners in both psychology and medicine struggle all the time with the tendency to have their judgment clouded by the overwhelming impact of the single case. Writer Francine Russo (1999) describes

the dilemma of Willie Anderson, an oncologist at the University of Virginia. Anderson is an advocate of controlled experimentation and routinely enrolls his patients in controlled clinical trials, but he still struggles with his own reactions to single, salient cases that have an emotional impact on his decisions. Despite his scientific orientation, he admits that "when it's real people looking you in the eye, you get wrapped up in their hopes and your hopes for their hopes, and it's *hard*" (p. 36). But Anderson knows that sometime the best thing for his patients is to ignore the "real person looking you in the eye" and go with what the best evidence says. And the best evidence comes from a controlled clinical trial (described in Chapter 6) not from the emotional reaction to that person looking you in the eye.

What to Do About the Vividness Problem

The vividness problem is a difficulty we all face when evaluating evidence. And in an environment saturated with media images, it is becoming increasingly difficult for society not to be dominated by the images and instead to solve its problems based on valid evidence. Writer Barry Glassner (1999) describes an all too familiar example. On an Oprah Winfrey program in 1995, information was being presented on a surgical intervention (which will here remain unnamed so as not to contribute further to a vividness effect) that had caused some controversy because some people had claimed to have been injured by it. Evidence was being presented that studies from the Mayo Clinic, Harvard, and the University of Michigan had shown no overall danger from the procedure—at which point a woman claiming to have been injured jumped up from the audience and shouted "We are the evidence. The study is us sitting here!" (Glassner, 1999, p. 164). Which do you think the television audience of millions remembered better—the study from the Mayo Clinic or the woman screaming that she had been injured?

Even though we all are prone to overestimating the value of testimonial and other single-case evidence, we can become more self-aware and more conscious of when our opinions are being overwhelmed by personal testimony or particularly vivid single cases. A column by writer Remar Sutton (1987) illustrates quite well how becoming more aware of these influences can help. He wrote about the beginning of his attempt at a major weight loss by describing how "a diving buddy" had recommended a popular book, *Fit for Life*. Sutton described seeing the authors on a talk show and mentioned that they "appeared awfully sincere, sounded authoritative, and attacked some traditional nutritional thinking." He recalled that "every time a critic attacked them, they rebutted effectively *with their own experiences* [italics added]. All of that, *and* the talk show host said he lost weight and felt better with their plan, too." In short, Sutton admitted in his column that he had become interested in the ideas in this book through his hearing of personal experiences and testimonials.

But fortunately Sutton had acquired the critical thinking skills that this book is trying to teach. He researched the information given in the book, looking for confirming evidence—not on talk shows and in popular magazines but in the peer-reviewed scientific literature (see Chapters 1 and 12). He concluded that "too many ideas presented in the book were simply not backed up by long-term controlled scientific research that was statistically valid. Indeed, most of these ideas were counter to statistically valid research." Furthermore, in the course of his research, Sutton found out that the two authors had received their nutrition certificates from a "school" in Austin, Texas, that was nothing more than a post office box!

Sutton had the insight to realize that he had been close to being "sucked in" by the quackery in the book and to ask himself, "So why did I nearly fall for the *Fit for Life* approach?" His answer provides a good review of several of the pitfalls described previously. He admitted that he "trusted [his] diving friend's recommendation based on his personal experience." He had also liked the way the authors "presented *themselves* as much as their information" and that he "believed that television...and large numbers of book sales made the message of *Fit for Life* legitimate."

Sutton's conclusion sums up the vividness issue nicely:

> Like it or not, personal testimonials and the sincerity of a person's presentation do not necessarily make any product or opinion accurate. Carefully designed, properly controlled, replicable studies which prove statistically valid over the long term are the only assurance any of us have that a diet or medical treatment ...can benefit us.

Sutton ended by stating that he was still not hostile to new or innovative approaches to dieting, and he urged proponents of new approaches to write him. But he reminded them, "I'll be happy to present the opposite point of view on diet and fitness as long as the proponents of these views can point me to the reliable scientific studies that provide statistical validity for that position. Please do write, but don't forget those bibliographies."

In summary, the problems created by reliance on testimonial evidence are ever-present. The vividness of such evidence often eclipses more reliable information and obscures understanding. Psychology instructors worry that merely pointing out the logical fallacies of reliance on testimonial evidence is not enough to provide a deep understanding of the pitfalls of this type of data. What else can be done? Is there any other way to get this concept across to people? Fortunately there is an alternative—an alternative somewhat different from the academic approach. The essence of this approach is to fight vividness with vividness. To hoist testimonials by their own petard! To let testimonials devour themselves with their own absurdity. A practitioner of this approach is the one, the only, the indubitable Amazing Randi!

The Amazing Randi: Fighting Fire with Fire

James Randi is a magician and jack-of-all trades who was given a MacArthur Foundation "genius" grant. For many years, he has been trying to teach the public some basic skills of critical thinking. The Amazing Randi (his stage name) has done this by exposing the fraud and charlatanism surrounding claims of "psychic" abilities. Although he has uncovered many magicians and conjurors masquerading as psychics, he is best known for exposing the trickery of Uri Geller, the psychic superstar of the 1970s. Bursting on the scene with his grand claims of psychic powers, Geller captivated the media to an extraordinary degree. He was featured in newspapers, on television shows, and in major news magazines on several continents. Randi detected and exposed the common and sometimes embarrassingly simple magic tricks that Geller used to perform his psychic "feats," which included bending keys and spoons, and starting watches—mundane fare for a good magician. Since the Geller exposé, Randi has continued to use his considerable talents in the service of the public's right to know the truth in spite of itself by exposing the fallacies behind ESP, biorhythms, psychic surgery, extraterrestrials, levitation, and other pseudosciences, all marvelously detailed with great humor in his book *Flim-Flam* (1980; see also his book *The Faith Healers*, 1987, in which he exposed several bogus religious "healers" as frauds.)

One of Randi's minor diversions consists of demonstrating how easy it is to garner testimonial evidence for any preposterous event or vacuous claim. His technique is to let people be swallowed up in a trap set by their own testimonials. Randi makes much use of that fascinating American cultural institution, the talk show, often appearing as a guest in the guise of someone other than himself. On a New York show a few years ago, he informed the audience that, while driving through New Jersey earlier in the day, he had seen a formation of orange V-shaped objects flying overhead in a northerly direction. Within seconds, as Randi put it, "the station switchboard lit up like an electronic Christmas tree." Witness after witness called in to confirm this remarkable sighting. Unfortunately for them, the "sighting" was only a product of Randi's imagination. Callers provided many details that Randi had "omitted," including the fact that there had been more than one pass of the "saucers." This little scam illustrates how completely unreliable are individual reports that "something happened."

In Winnipeg, Canada, Randi appeared on a radio show as an "astrographologist." A week earlier, listeners had been told to send in their handwriting samples and birth dates. Three were chosen and were contacted so that Randi could assess the "readings" of their personalities. He was hugely successful, receiving ratings from the listeners of 9, 10, and 10 on a 1-to-10 scale of accuracy. Randi did eventually reveal the secret of his method to the radio listeners. He had read, word for word, the "readings" that astrologer Sydney Omarr had given to three members of the audience of a recent *Merv Griffin Show.*

On a different radio show, Randi demonstrated the basis for the popularity of another pseudoscience: biorhythms (Hines, 1998). One listener agreed to keep a day-by-day diary and compare it with a two-month biorhythm chart that had been prepared especially for her. Two months later, the woman called back to inform the audience that biorhythms should be taken very seriously because her chart was more than 90 percent accurate. Randi had to inform her of the silly mistake made by his secretary, who had sent someone else's chart to her, rather than her own. However, the woman did agree to evaluate the correct chart, which would be mailed to her right away, and to call back. A couple of days later, the woman called back, relieved. Her own chart was just as accurate—in fact, even *more* accurate. On the next show, however, it was discovered that, whoops, another error had been made. The woman had been sent Randi's secretary's chart, rather than her own!

Randi's biorhythm and astrographologist scams are actually examples of a phenomenon that has been termed the *P. T. Barnum* effect (Barnum, the famous carnival and circus operator, coined the statement "There's a sucker born every minute"). This effect has been extensively studied by psychologists (e.g., Dickson & Kelly, 1985), who have found that the vast majority of individuals will endorse generalized personality summaries as accurate and specific descriptions of themselves. The Barnum effect makes it easier to generate testimonials and, of course, shows why they are worthless. There are certain sets of statements and phrases that most people see as applicable to themselves (many of these phrases have been studied by psychologists; see, e.g., Dickson & Kelly, 1985; Hyman, 1981; Marks & Kammann, 1980). Anyone can feed them to a "client" as individualized psychological "analysis." The client is usually very impressed by the individualized accuracy of the "personality reading," not knowing that the same reading is being given to everyone. Of course, the Barnum effect is the basis of belief in the accuracy of palm readers and astrologists.

What Randi is trying to do in these little scams is to teach people a lesson about the worthlessness of testimonial evidence. He consistently demonstrates how easy it is to generate testimonials in favor of just about *any* bogus claim (Randi, 1983). For this reason, presenting a testimonial in support of a particular claim is meaningless. Only evidence from controlled observations (to be described in Chapter 6) is sufficient to actually *test* a claim.

TESTIMONIALS OPEN THE DOOR TO PSEUDOSCIENCE

It is sometimes claimed that pseudosciences like the ones we have just discussed are simply a way to have a little fun, that they really do no harm. After all, why should we care? Isn't it just a case of a few people engaging in wishful thinking and a few others making a couple of bucks out of them?

A complete examination of the problem reveals, however, that the harm done to society by the prevalence of pseudosciences is more widespread than is generally believed. First, belief in pseudoscience leads to personal tragedies. Consider the case of Rhea Sullins, the 7-year-old daughter of the president of the American Natural Hygiene Society, whose tenets include the substitution of the consumption of fruits and vegetables for the use of drugs and conventional medical treatment. When Rhea became ill, "her father put her on a water-only fast for 18 days and then on a diet of fruit juice for 17 more" (Gilovich, 1991, p. 6). Sadly, Rhea died of malnutrition. Examples like this one of Rhea are unfortunately not hard to find (see Bulgatz, 1992; Kaminer, 1999; Shermer, 1997).

In a complex, technological society, the influence of pseudoscience can be propagated by decisions that affect thousands of other people. That is, you may be affected by pseudoscientific beliefs even if you do not share those beliefs. For example, major banks and several Fortune 500 companies employ graphologists for personnel decisions (Sutherland, 1992) even though voluminous evidence indicates that graphology is useless for this purpose (Ben-Shakhar, Bar-Hillel, Blui, Ben-Abba, & Flug, 1989; Neter & Ben-Shakhar, 1989). To the extent that pseudodiagnostic graphological cues lead employers to ignore more valid criteria, both economic inefficiency and personal injustice are the result. How would you like to lose your chance for a job that you really want because you have a particular little "loop" in your handwriting? Or alternatively, how would you feel if you were denied a job because a "psychic" saw a disturbance in your "aura"? In fact, this is actually happening to some people. Some corporations are paying for "psychic readings" of individuals who are candidates for hiring. For example, Susan King is a so-called clairvoyant whom companies pay to give them "readings" to aid in personnel decisions. She claims that she doesn't even need to meet the applicant—that she can work from photos and first names—although "some clients call her in to observe short-listed candidates during final interviews and even at cocktail parties afterwards" (Kershaw, 1991). In tough and competitive economic times, is this how you want your employment opportunities to be determined?

Unfortunately, these examples are not rare (see Shermer, 1997; Sutherland, 1992). We are all affected in numerous ways when pseudoscientific beliefs permeate society—even if we do not subscribe to the beliefs. For example, police departments hire psychics to help with investigations (Marshall, 1980) even though research has shown that this practice has no effectiveness (Hines, 1988; Reiser, Ludwig, Saxe, & Wagner, 1979; Rowe, 1993). Programmers at the ABC television network hired a Hollywood psychic to help make decisions about the content of the most influential communications technology in our society (Auletta, 1992, p. 114). And most astonishingly, an astrologer was employed in the Reagan White House to advise on the "timing of presidential speeches, appearances, meetings with heads of state, airplane travel schedules, and even discussion topics" (Johnson, 1991, p. 454).

Pseudosciences such as astrology are now large industries, involving newspaper columns, radio shows, book publishing, the Internet, magazine articles, and other means of dissemination. The leading horoscope magazine has a circulation of one-quarter million, larger than that of many legitimate science magazines (Bastedo, 1981). The House of Representatives Select Committee on Aging has estimated that the amount wasted on medical quackery nationally now approaches $10 billion each year. The committee's report stated that:

> quackery now invades nearly every aspect of our lives, and, at points, attracts adherents with near religious zeal. The easy to recognize and somewhat comic figure of pitchmen of the past has been supplanted by sophisticated and shrewd hucksters with a keen sense of the vulnerability of potential customers, the limitations of the law and the profitability of exploiting both. Quackery has become big business.... The modern quack's prime target is the senior citizen. His purview is chronic and incurable diseases. His approach is the illustrated brochure, the supermarket press, television commercials, "testimonial" ads, phony foundations providing "scientific" information and storefront clinics. (U.S. Congress, 1984, pp. 3–4)

In short, pseudoscience is a multimillion-dollar business, and the incomes of thousands of individuals depend on its public acceptance.

The topic of pseudoscience in the area of medicine is useful to consider, because medical associations have been more aggressive than psychology in attacking pseudoscience and dissociating legitimate medical practice from the illegitimate. Consider the guidelines published by the Arthritis Foundation and cited by the House Committee on Aging for spotting the unscrupulous promoter:

1. He may offer a "special" or "secret" formula or device for "curing" arthritis.
2. He advertises. He uses "case histories" and testimonials from satisfied "patients."
3. He may promise (or imply) a quick or easy cure.
4. He may claim to know the cause of arthritis and talk about "cleansing" your body of "poisons" and "pepping up" your health. He may say surgery, X rays and drugs prescribed by a physician are unnecessary.
5. He may accuse the "medical establishment" of deliberately thwarting progress, or of persecuting him...but he doesn't allow his method to be tested in tried and proven ways. (U.S. Congress, 1984, p. 12)

This list could also serve as a guide for spotting fraudulent psychological treatments and claims. Note, of course, point 2, which is the focus of this chapter. But also note that points 1 and 5 illustrate the importance of something discussed earlier: Science is public. In addition to putting forth testimonials as "proof," the practitioners of pseudoscience often try to circumvent the public verifiability criterion of science by charging that there is a conspir-

acy to suppress their "knowledge." They use this as justification for going straight to the media with their "findings" rather than submitting their work to the normal scientific publication processes. The ploy is usually more successful in the area of psychology than anywhere else because the media often show less respect for the normal scientific mechanisms of psychology than they do for those of other sciences. It is important to keep this bias in mind (it will receive an extended discussion in Chapter 12). Unverified claims that the media would never think of reporting if they occurred in the physical sciences are seen as legitimate topics of psychological reporting because journalists have been convinced by the purveyors of pseudoscience that "anything goes" in psychology. Thus, the consumer must be aware that television and the print media will publicize virtually any outlandish claim in the area of psychology if they think there is an audience for it, no matter how much the claim is contradicted by the available evidence. The public ends up the loser.

Pseudoscience hurts society in other ways. For example, there is what economists call the *opportunity cost:* people waste time and money that they could have spent on something else, and they therefore lose opportunities. Physicist Philip Morrison (1983) argued:

> The danger of pseudoscience is not to science, but to society.... It is primarily an opportunity cost, in the language of economics. Here we have all these people interested in the exciting and wonderful and amazing properties of the world who could be learning something real. Instead they are hearing from Uri Geller. They've turned themselves over to dependence on charlatans. They could be learning something different. (p. 56)

But there are costs more specific than the general learning-opportunity cost. For example, claims of miracle cures raise false hopes that can cause psychological damage when they are dashed. One of the most despicable examples in my files on this subject is an article from a grocery store tabloid entitled "Psychic Shows the Blind How to See Using ESP." People can fail to take advantage of the real knowledge available to them because they become involved in pseudosciences. Proponents of psychic surgery implicitly encourage people to spend money on bogus cures and to ignore traditional "nonpsychic" medical procedures that may help them (Angell & Kassirer, 1998). Reporter Richard Brenneman (1990) described how a bogus psychic surgeon had bilked cancer victims out of their money and had also delayed their eventual treatment by legitimate medical means. Similarly, at a civic meeting on health care fraud, the president of the Michigan Council Against Health Fraud displayed a brochure for a phony cancer cure which said that "this product cannot be used with any other cancer therapy" ("If It Sounds," 1990). The damage done in cases like this is simply incalculable.

Physicians are increasingly concerned about the spread of medical quackery on the Internet and its real health costs. Dr. Max Coppes was

prompted to write a letter to the *New England Journal of Medicine* warning of the real human costs of pseudoscience in medicine (Scott, 1999). He described the case of a 9-year-old girl who, after cancer surgery, had a 50 percent chance of three more years of life if she had undergone chemotherapy. Instead, her parents found an unproven treatment that utilized shark cartilage and opted for that instead. The young girl was dead in four months.

Thus, the spread of pseudoscience is quite costly. And nothing fosters the spread of pseudosciences more than confusion about what type of evidence does and does not justify belief in a claim about a phenomenon. By providing readily available support for virtually any claim and by the impact that they have when used, testimonials open the door to the development of and the belief in pseudosciences. There is no more important rule for the consumer of psychological information than to beware of them. In the next several chapters, we shall see what type of evidence *is* required to justify claims.

SUMMARY

Case study and testimonial evidence is useful in psychology (and other sciences) in the very earliest stages of an investigation, when it is important to find interesting phenomena and variables. As useful as case study evidence is in the early, pretheoretical, stages of scientific investigation, it is virtually useless in the later stages, when theories are being put to specific test. In psychology, one of the main reasons that case studies and testimonial evidence are useless for theory testing is the existence of the placebo effect. The placebo effect is the tendency of people to report that any treatment has helped them, regardless of whether the treatment had a real therapeutic element. The existence of placebo effects makes it impossible to prove the effectiveness of a psychological (or medical) treatment by producing testimonials to its effectiveness. The reason is that the placebo effect guarantees that no matter *what* the treatment, it will be possible to produce testimonial evidence to its effectiveness.

Despite the uselessness of testimonial evidence in theory testing, psychological research has indicated that such evidence is often weighted quite heavily by people because of the vividness effect: People overweight evidence that is more vivid and hence more retrievable from memory. One thing that is particularly vivid for most people is testimonial evidence. The result is an overreliance on such evidence in the justification of specific psychological claims. In fact, testimonial and case study evidence cannot be used to justify general theoretical claims.