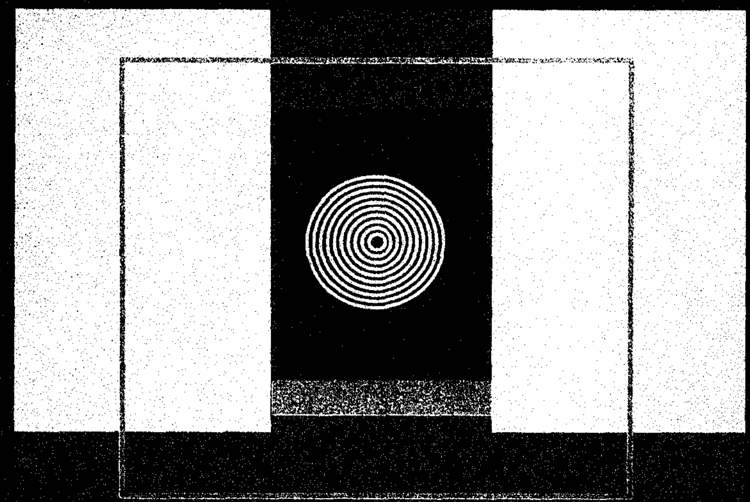
HOW TO THINK STRAIGHT ABOUT



KEITH E. STANOVICH

Henderson, Metropolitan State University; George Howard, University of Notre Dame; Barry Kendall; Bernie Koenig, Fanshawe College; Victor Koop, Goshen College; P. A. Lamal, University of North Carolina, Charlotte; Stephen Louisell, Kalamazoo Community College; Margaret Matlin, SUNY-Geneseo; Douglas Mook, University of Virginia; Timothy Moore, York University; Edward Morris, University of Kansas; Joseph E. Morrow, California State University at Sacramento; Michael O'Boyle, Iowa State University; Blaine Peden, University of Wisconsin, Eau Claire; John F. Pfister, Dartmouth College; Sam Rakover, University of Haifa; Richard Redding, Hahneman University; Michael Ross, University of Waterloo; John Ruscio, Elizabethtown College; Walter Sa, Grand Valley State University; Allen Salo, University of Maine at Presque Isle; Frank Schieber, University of South Dakota; Marjorie Semonick, University of Minnesota; David Share, University of Haifa; Jeffrey Sherman, Northwestern University; Linda Siegel, University of British Columbia; Norman Silverman, University of Illinois, Chicago; Frank Smoll, University of Washington; Paul Solomon, Williams College; Mike Stadler, University of Missouri; Maggie Toplak, York University; Larry Vandervert, Spokane Falls Community College; John Vokey, University of Lethbridge; Carol Wade, College of Marin; Marty Wall, University of Toronto; Barbara Wanchisen, Baldwin-Wallace College; Toni G. Wegner, University of Virginia; Edward Wisniewski, Northwestern University; Murray S. Work, California State University at Sacramento; and Edward Zuckerman, Guilford Press. The insights from many discussions about teaching methodology with Ted Landau, Larry Lilliston, and Dean Purcell, all of Oakland University, were incorporated into the book.

Robyn Macpherson is thanked for her diligent library and reference assistance, which was extremely helpful in meeting the deadlines for the eighth edition. Carol Kelley has bolstered my morale throughout the writing of this and several previous editions. Andrée Debruge is thanked for some valuable last minute assistance. I appreciate the valuable suggestions provided by the following reviewers for this edition: Steven Isonio, Golden West College; Cindy Sifonis, Oakland University; Michael Tagler, Nebraska Wesleyan University; and Chris Ward, Stonehill College. My editor at Allyn & Bacon, Susan Hartman, has provided guidance, enthusiasm, and support for the book, as has her editorial assistant, Therese Felser.

Finally, I wish to thank Paula J. Stanovich for more than just the emotional support that is routinely alluded to in acknowledgments. Her concern for all human beings, particularly those less fortunate, is an inspiration to all who know her. A view we both share is that all human beings should have the opportunity to utilize their full potential. This book attests to the fact that I have had such an opportunity. Paula works to speed the day when this opportunity will be fully extended to all individuals with disabilities.

1

Psychology Is Alive and Well (and Doing Fine Among the Sciences)

The Freud Problem

Stop 100 people on the street and ask them to name a psychologist, either living or dead. Record the responses. Of course, Dr. Phil, Wayne Dyer, and other "media psychologists" would certainly be named. If we leave out the media and pop psychologists, however, and consider only those who have made a recognized contribution to psychological knowledge, there would be no question about the outcome of this informal survey. Sigmund Freud would be the winner hands down. B. F. Skinner would probably finish a distant second. No other psychologist would get enough recognition even to bother about. Thus, Freud, along with the pop psychology presented in the media, largely defines psychology in the public mind.

The notoriety of Freud has greatly affected the general public's conceptions about the field of psychology and has contributed to many misunderstandings. For example, many introductory psychology students are surprised to learn that, if all the members of the American Psychological Association (APA) who were concerned with Freudian psychoanalysis were collected, they would make up less than 10 percent of the membership. In another major psychological association, the Association for Psychological Science, they would make up considerably less than 5 percent.

Modern psychology is not obsessed with the ideas of Sigmund Freud (as are the media and some humanities disciplines), nor is it largely defined by them. Freud's work is an extremely small part of the varied set of issues, data, and theories that are the concern of modern psychologists. This larger body of research and theory encompasses the work of five recent Nobel Prize winners (David Hubel, Daniel Kahneman, Herbert Simon, Roger Sperry, and Torsten Wiesel) and a former director of the National Science Foundation (Richard Atkinson), all of whom are virtually unknown to the public.

It is bad enough that Freud's importance to modern psychology is vastly exaggerated. What makes the situation worse is that Freud's methods of investigation are completely unrepresentative of how modern psychologists conduct their research (recall that Freud began his well-known work over a hundred years ago). In fact, the study of Freud's methods gives an utterly misleading impression of psychological research. For example, Freud did not use controlled experimentation, which, as we shall see in Chapter 6, is the most potent weapon in the modern psychologist's arsenal of methods. Freud thought that case studies could establish the truth or falsity of theories. We shall see in Chapter 4 why this idea is mistaken. Finally, a critical problem with Freud's work concerns the connection between theory and behavioral data. As we shall see in Chapter 2, for a theory to be considered scientific, the link between the theory and behavioral data must meet some minimal requirements. Freud's theories often do not meet these criteria (Crews, 1996, 1998; Hines, 2003; Macmillan, 1997; McCullough, 2001; Watters & Ofshe, 1999). To make a long story short, Freud built an elaborate theory on a database (case studies and introspection) that was not substantial enough to support it. Freud concentrated on building complicated theoretical structures, but he did not, as modern psychologists do, ensure that they would rest on a database of reliable, replicable behavioral relationships. In summary, familiarity with Freud's style of work can be a significant impediment to the understanding of modern psychology.

In this chapter, we shall deal with the Freud problem in two ways. First, when we illustrate the diversity of modern psychology, the rather minor position occupied by Freud will become clear (see Haggbloom et al., 2002; Robins, Gosling, & Craik, 1999, 2000). Second, we shall discuss what features are common to psychological investigations across a wide variety of domains. A passing knowledge of Freud's work has obscured from the general public what is the only unifying characteristic of modern psychology: the quest to understand behavior by using the methods of science.

The Diversity of Modern Psychology

There is, in fact, a great diversity of content and perspectives in modern psychology. This diversity drastically reduces the coherence of psychology as a discipline. Henry Gleitman (1981), winner of the American Psychological Foundation's Distinguished Teaching Award, characterized psychology as "a loosely federated intellectual empire that stretches from the domains of the biological sciences on one border to those of the social sciences on the other" (p. 774).

Understanding that psychology is composed of an incredibly wide and diverse set of investigations is critical to an appreciation of the nature of the discipline. Simply presenting some of the concrete indications of this diversity will illustrate the point. The American Psychological Association has 53 different divisions, each representing either a particular area of research and study or a particular area of practice (see Table 1.1). From the table, you can see the range of subjects studied by psychologists, the range of settings involved, and the different aspects of behavior studied. The other large organization of psychologists—the Association for Psychological Science—is just as diverse. Actually, Table 1.1 understates the diversity within the field of psychology because it gives the impression that each division is a specific specialty area. In fact, each of the 53 divisions listed in the table is a broad area of study that contains a wide variety of subdivisions! In short, it is difficult to exaggerate the diversity of the topics that fall within the field of psychology.

Implications of Diversity

Many people come to the study of psychology hoping to learn the one grand psychological theory that unifies and explains all aspects of human behavior and consciousness. Such hopes are often disappointed, because psychology contains not one grand theory, but many different theories, each covering a limited aspect of behavior (Benjamin, 2001; Griggs, Proctor, & Bujak-Johnson, 2002; Zechmeister & Zechmeister, 2000). The diversity of psychology guarantees that the task of theoretical unification will be immensely difficult. Indeed, many psychologists would argue that such a unification is impossible. Others, however, are searching for greater unification within the field (Henriques, 2003, 2004; Kenrick, 2001; Kimble, 1999; Sternberg, 2005). For example, the coherence of psychology as a discipline has increased over the last decade due to the theoretical efforts of evolutionary psychologists. These researchers have tried to bring unification to our conceptualization of human psychological processes by viewing them as mechanisms serving critical evolutionary functions such as kinship recognition, mate selection, cooperation, social exchange, and child rearing (Barrett, Dunbar, & Lycett, 2002; Bjorklund & Pellegrini, 2002; Buss, 2003; Cosmides & Tooby, 2000; Geary, 2005; Pinker, 2002).

Nevertheless, no matter what their position on the issue of the coherence of the subject matter of psychology, all psychologists agree that theoretical unification will be extremely difficult, if it is to occur at all. The lack of theoretical integration leads some critics of psychology to denigrate the scientific progress that psychology has made. Such criticism often arises from the mistaken notion that all true sciences must have a grand, unifying theory. It is a mistaken notion because many other sciences also lack a unifying

TABLE 1.1 ***Divisions of the American Psychological Association***

1.	General Psychology
2.	Teaching of Psychology
3.	Experimental Psychology
5.	Evaluation, Measurement, and Statistics
6.	Behavioral Neuroscience and Comparative Psychology
7.	Developmental Psychology
8.	Personality and Social Psychology
9.	Psychological Study of Social Issues
10.	Psychology of Aesthetics, Creativity, and the Arts
12.	Clinical Psychology
13.	Consulting Psychology
14.	Industrial and Organizational Psychology
15.	Educational Psychology
16.	School Psychology
17.	Counseling Psychology
18.	Psychologists in Public Service
19.	Military Psychology
20.	Adult Development and Aging
21.	Applied Experimental and Engineering Psychology
22.	Rehabilitation Psychology
23.	Consumer Psychology
24.	Theoretical and Philosophical Psychology
25.	Behavior Analysis
26.	History of Psychology
27.	Community Psychology
28.	Psychopharmacology and Substance Abuse
29.	Psychotherapy
30.	Psychological Hypnosis
31.	State Psychological Association Affairs
32.	Humanistic Psychology
33.	Mental Retardation and Developmental Disabilities
34.	Population and Environmental Psychology
35.	Psychology of Women
36.	Psychology of Religion
37.	Child, Youth, and Family Services
38.	Health Psychology
39.	Psychoanalysis
40.	Clinical Neuropsychology
41.	Psychology and Law
42.	Psychologists in Independent Practice
43.	Family Psychology
44.	Psychological Study of Lesbian, Gay, and Bisexual Issues
45.	Psychological Study of Ethnic Minority Issues
46.	Media Psychology
47.	Exercise and Sport Psychology
48.	Peace Psychology
49.	Group Psychology and Group Psychotherapy
50.	Addictions
51.	Psychological Study of Men and Masculinity
52.	International Psychology
53.	Clinical Child Psychology and Adolescent Psychology
54.	Pediatric Psychology
55.	Pharmacotherapy

Note: There is no Division 4 or 11.

TABLE 1.1 ***Divisions of the American Psychological Association (continued)***

conceptualization. Harvard psychologist William Estes (1979) has emphasized this point:

> The situation in which the experimental psychologists find themselves is not novel, to be sure, nor peculiar to psychology. Physics during the early twentieth century subdivided even at the level of undergraduate teaching into separate disciplines. Thus I was introduced to that science through separate university courses in mechanics, heat, optics, acoustics, and electricity. Similarly, chemistry has branched out, evidently irreversibly, into inorganic, organic, physical, and biochemical specialties, among which there may be no more communication than among some of the current subdisciplines of psychology. In both cases, unity has reemerged only at the level of abstract mathematical theory. Medicine has similarly fragmented into specialties, but is like psychology in that there has been no appearance of a new unity. (pp. 661–662)

In fact, psychology is so fragmented that it is not difficult to imagine a university disbanding its psychology department and integrating its members into other departments. Physiological psychologists could go into biology departments; many social psychologists could go into sociology departments; cognitive and perceptual psychologists could go into interdisciplinary departments of cognitive science; organizational and industrial psychologists could go into business schools; clinical and counseling psychologists could go into departments of social work, human resources, and education; developmental psychologists might go to departments of education, cognitive science, or human resources; and so on. Few psychologists would notice any difference in the intellectual interchange with their new colleagues. Actually, many would experience greater camaraderie with their new colleagues than with the old ones in "psychology." Modern psychology, in terms of content, simply does not hang together as a coherent set of topics. One must look to a more general level to find anything that unifies the discipline.

Once we acknowledge the implications of the social and historical factors that determine the structure of disciplines, we can recognize that it is illogical to demand that all fields be unified. Indeed, it has been suggested that the term *psychological studies,* rather than *psychology,* would more accurately reflect the diversity of the discipline. The use of this new term would also make it less surprising to the student that the different areas within the discipline have been characterized by vastly different rates of scientific progress. Some have made impressive progress in the explanation and prediction of behavior, while others have progressed much more slowly. The term *psychology* does not convey this state of affairs. Instead, it implies a coherence of subject matter that is simply not characteristic of the discipline.

If we wish to find any unity in the subject of psychology, we must not look for connections among the topics that psychologists study. We must instead address the methods that psychologists use to advance knowledge. This is the only place that we have any hope of finding common cause among psychologists. But here, in the domain of the methods psychologists use to advance knowledge, is where we also find some of the greatest misunderstandings of the discipline.

Unity in Science

Simply to say that psychology is concerned with human behavior does not distinguish it from other disciplines. Many other professional groups and disciplines—including economists, novelists, law, sociology, history, political science, anthropology, and literary studies—are, in part, concerned with human behavior. Psychology is not unique in this respect.

Practical applications do not establish any uniqueness for the discipline of psychology either. For example, many university students decide to major in psychology because they have the laudable goal of wanting to help people. But helping people is an applied part of an incredibly large number of fields, including social work, education, nursing, occupational therapy, physical therapy, police science, human resources, and speech therapy. Similarly, helping people by counseling them is an established part of the fields of education, social work, police work, nursing, pastoral work, occupational therapy, and many others. The goal of training applied specialists to help people by counseling them does not demand that we have a discipline called psychology.

The *only* two things that justify psychology as an independent discipline are that it studies the full range of human and nonhuman behavior with the techniques of science and that applications that derive from this knowledge are scientifically based. Were this not true, there would be no reason for psychology to exist.

Psychology is different from other behavioral fields in that it attempts to give the public two guarantees. One is that the conclusions about behav-

ior that it produces derive from scientific evidence. The second is that prac tical applications of psychology have been derived from and tested by sc entific methods. Does psychology ever fall short of these goals? Yes, quit often (Lilienfeld, Lynn, & Lohr, 2003; Loftus & Guyer, 2002; Lynn, Loftu Lilienfeld, & Lock, 2003; Mook, 2001; Watters & Ofshe, 1999). This book i about how we might better attain them. I will return in Chapter 12 to the issu of psychologists themselves undermining their own legitimacy by not mee ing appropriate scientific standards. But, in principle, these are the standard that justify psychology as an independent field. If psychology ever decide that these goals are not worth pursuing—that it does not wish to adhere t scientific standards—then it might as well fold its tent and let its various cor cerns devolve to other disciplines, as previously outlined, because it woul be a totally redundant field of intellectual inquiry.

Clearly, then, the first and most important step that anyone must tak in understanding psychology is to realize that its defining feature is that it i the data-based scientific study of behavior. Comprehending all of the impli cations of this fact will occupy us for the rest of this book, because it is th primary way that we develop the ability to think straight about psycholog Conversely, the primary way that people get confused in their thinking abou psychology is that they fail to realize that it is a scientific discipline. For ex ample, it is quite common to hear people outside the discipline voice the opin ion that psychology is not a science. Why is this a common occurrence?

Attempts to convince the public that psychology cannot be a scienc stem from a variety of sources. As will be discussed in later chapters, mucl confusion about the actual discipline of psychology is deliberately fostere by purveyors of bogus psychology. There has grown up in our society a con siderable industry of pseudoscientific belief systems that have a vested in terest in convincing the public that anything goes in psychology and that ther are no rational criteria for evaluating psychological claims. This is the per fect atmosphere in which to market such offers as "Lose weight through hyp nosis," "Develop your hidden psychic powers," and "Learn French while yo sleep," along with the many other parts of the multimillion-dollar self-hel industry that either are not based on scientific evidence or, in many cases are actually contradicted by much available evidence.

Another source of resistance to scientific psychology stems from the ten dency to oppose the expansion of science into areas where unquestioned au thorities and "common sense" have long reigned. History provides man examples of initial public resistance to the use of science rather than philo sophical speculation, theological edict, or folk wisdom to explain the nat ural world. Each science has gone through a phase of resistance to it development. Learned contemporaries of Galileo refused to look into his new telescope because the existence of the moons of Jupiter would have violate their philosophical and theological beliefs. For centuries, the understanding o human anatomy progressed only haltingly because of lay and ecclesiastical prohibitions of the dissection of human cadavers (the Christian view was that the interior of the body was "God's province"; see Grice, 2001). Charles Darwin was repeatedly denounced. Paul Broca's Society of Anthropology was opposed in France because knowledge about human beings was thought to be subversive to the state.

Each scientific step to greater knowledge about human beings has evoked opposition. This opposition eventually dissipated, however, when people came to realize that science does not defile humanity by its investigations but contributes to human fulfillment by widening the sphere of knowledge. Who now believes that astronomy's mapping of the galaxies and its intricate theories about the composition of distant stars destroy our wonder at the universe? Who would substitute the health care available in their community for that available before human cadavers were routinely dissected? An empirical attitude toward the stars or the human body has not diminished humanity. More recently, Darwin's evolutionary synthesis laid the foundation for startling advances in genetics and biology. Nevertheless, as we get closer to the nature of human beings and their origins, vestiges of opposition remain. In the United States, religious advocates continue to press for the teaching of creationism in the public schools, and surveys show that the scientific fact that humans evolved by natural selection is not accepted by a large portion (in some surveys, a majority) of the public (Lerner, 2005). If evolutionary biology, with its long and impressive record of scientific achievements, still engenders public opposition, is it any wonder that psychology, the most recent discipline to bring long-held beliefs about human beings under scientific scrutiny, currently provokes people to deny its validity?

What, Then, Is Science?

In order to understand what psychology is, we must understand what science is. We can begin by dealing with what science is not. In this way, we can rid ourselves of the vast majority of common misconceptions. First, science is not defined by subject matter. Any aspect of the universe is fair game for the development of a scientific discipline, including all aspects of human behavior. We cannot divide the universe into "scientific" and "nonscientific" topics. Although strong forces throughout history have tried to place human beings outside the sphere of scientific investigation, they have been unsuccessful, as we shall see. The reactions against psychology as a scientific discipline probably represent the modern remnants of this ancient struggle.

Science is also not defined by the use of particular experimental apparatus. It is not the test tube, the computer, the electronic equipment, or the investigator's white coat that defines science. (If this were the case, there would be no question at all about psychology's status, because psychology departments in all major universities are full of computers, chemicals, and electronic equipment of all types.) These are the trappings of science but are not its defining features. Science is, rather, a way of thinking about and observing the universe that leads to a deep understanding of its workings.

In the remainder of this chapter, we shall discuss three important and interrelated features that define science: (1) the use of systematic empiricism; (2) the production of public knowledge; and (3) the examination of solvable problems. Although we shall examine each feature separately, remember that the three connect to form a coherent general structure. (For a more detailed discussion of the general characteristics of a science, see the works of Bronowski, Cournaud, Medawar, Popper, Raymo, and Sagan listed in the references section of this book.)

Systematic Empiricism

If you look up the word *empiricism* in any dictionary, you will find that it means "the practice of relying on observation." Scientists find out about the world by examining it. The fact that this point may seem obvious to you is an indication of the spread of the scientific attitude in the past couple of centuries. In the past, it has not always seemed so obvious. Recall the refusal to look into Galileo's telescope. It was long thought that knowledge was best obtained through pure thought or through appeal to authority. Galileo claimed to have seen moons around the planet Jupiter. Another scholar, Francesco Sizi, attempted to refute Galileo, not with observations, but with the following argument:

> There are seven windows in the head, two nostrils, two ears, two eyes and a mouth; so in the heavens there are two favorable stars, two unpropitious, two luminaries, and Mercury alone undecided and indifferent. From which and many other similar phenomena of nature such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of planets is necessarily seven. . . . Besides, the Jews and other ancient nations, as well as modern Europeans, have adopted the division of the week into seven days, and have named them from the seven planets; now if we increase the number of planets, this whole system falls to the ground. . . . Moreover, the satellites are invisible to the naked eye and therefore can have no influence on the earth and therefore would be useless and therefore do not exist. (Holton & Roller, 1958, p. 160)

The point is not that the argument is laughably idiotic, but that it was seen as a suitable rebuttal to an actual observation! We laugh now because we have the benefit of hindsight. Three centuries of the demonstrated power of the empirical approach give us an edge on poor Sizi. Take away those years of empiricism, and many of us might have been there nodding our heads and urging him on. No, the empirical approach is not necessarily obvious, which is why we often have to teach it, even in a society that is dominated by science.

Empiricism pure and simple is not enough, however. Note that the heading for this section is "*Systematic* Empiricism." Observation is fine and necessary, but pure, unstructured observation of the natural world will not lead to scientific knowledge. Write down every observation you make from the time you get up in the morning to the time you go to bed on a given day. When you finish, you will have a great number of facts, but you will not have a greater understanding of the world. Scientific observation is termed *systematic* because it is structured so that the results of the observation reveal something about the underlying nature of the world. Scientific observations are usually theory driven; they test different explanations of the nature of the world. They are structured so that, depending on the outcome of the observation, some theories are supported and others rejected.

Publicly Verifiable Knowledge: Replication and Peer Review

Scientific knowledge is public in a special sense. By *public*, we, of course, do not mean that scientific observations are posted on community-center bulletin boards. Instead, we refer to the fact that scientific knowledge does not exist solely in the mind of a particular individual. In an important sense, scientific knowledge does not exist at all until it has been submitted to the scientific community for criticism and empirical testing by others. Knowledge that is considered "special"—the province of the thought processes of a particular individual, immune from scrutiny and criticism by others—can never have the status of scientific knowledge.

Science makes the idea of public verifiability concrete via the procedure of replication. In order to be considered in the realm of science, a finding must be presented to the scientific community in a way that enables other scientists to attempt the same experiment and obtain the same results. When this occurs, we say that the finding has been replicated. Scientists use replication to define the idea of public knowledge. Replication ensures that a particular finding is not due simply to the errors or biases of a particular investigator. In short, for a finding to be accepted by the scientific community, it must be possible for someone other than the original investigator to duplicate it. When a finding is presented in this way, it becomes public. It is no longer the sole possession of the original researcher; it is instead available for other investigators to extend, criticize, or apply in their own ways.

The poet John Donne told us that "no man is an island." In science, no researcher is an island. Each investigator is connected to the scientific community and its knowledge base. It is this interconnection that enables science to grow cumulatively. Researchers constantly build on previous knowledge in order to go beyond what is currently known. This process is possible only if previous knowledge is stated in such a way that any investigator can use it to build on.

By *publicly verifiable knowledge*, then, we mean findings presented to the scientific community in such a way that they can be replicated, criticized, or extended by anyone in the community. This is a most important criterion not only for scientists but also for the layperson, who, as a consumer, must evaluate scientific information presented in the media. As we shall see in Chapter 12, one important way to distinguish charlatans and practitioners of pseudoscience from legitimate scientists is that the former often bypass the normal channels of scientific publication and instead go straight to the media with their "findings." One ironclad criterion that will always work for the public when presented with scientific claims of uncertain validity is the question, Have the findings been published in a recognized scientific journal that uses some type of peer review procedure? The answer to this question will almost always separate pseudoscientific claims from the real thing.

Peer review is a procedure in which each paper submitted to a journal is critiqued by several scientists, who then submit their criticisms to an editor (usually a scientist with an extensive history of work in the specialty area covered by the journal), who decides whether the weight of opinion warrants publication of the paper, publication after further experimentation and statistical analysis, or rejection because the research is flawed or trivial. Most journals carry a statement of editorial policy in each issue, so it is easy to check whether a journal is peer reviewed.

Not all information in peer-reviewed scientific journals is necessarily correct, but at least it has met a criterion of peer criticism and scrutiny. It is a minimal criterion, not a stringent one, because most scientific disciplines publish dozens of different journals of varying quality. Most scientific ideas can get published somewhere in the legitimate literature if they meet some rudimentary standards. The idea that only a narrow range of data and theory can get published in science is false. This is an idea often suggested by purveyors of bogus remedies and therapies who try to convince the media and the public that they have been shut out of scientific outlets by a conspiracy of "orthodox science." But consider for a minute just how many legitimate outlets there are in a field like psychology. The publication *Psychological Abstracts* summarizes articles from over 1,000 different journals. Most of these journals are peer reviewed. Virtually all halfway legitimate theories and experiments can find their way into this vast array of publication outlets.

Again, I am not suggesting that all ideas published in the journals summarized in *Psychological Abstracts* are necessarily valid. On the contrary, I emphasized earlier that this is only a minimal criterion. However, the point is that the failure of an idea, a theory, a claim, or a therapy to have adequate documentation in the peer-reviewed literature of a scientific discipline is a very sure sign. Particularly when the lack of evidence is accompanied by a media campaign to publicize the claim, *it is a sure sign that the idea, theory, or therapy is bogus*. For example, in a famous Pennsylvania court case in 2005 regarding attempts to teach creationism in school biology classes, one of the witnesses advocating for intelligent design (a form of creationism) admitted that "he was unable to name any peer-reviewed research generated by intelligent design, though the movement has been around for more than a decade" (Talbot, 2005, p. 68).

The mechanisms of peer review vary somewhat from discipline to discipline, but the underlying rationale is the same. Peer review is one way (replication is another) that science institutionalizes the attitudes of objectivity and public criticism. Ideas and experimentation undergo a honing process in which they are submitted to other critical minds for evaluation. Ideas that survive this critical process have begun to meet the criterion of public verifiability. The peer review process is far from perfect, but it is really the only consumer protection that we have. To ignore it (or not to be aware of it) is to leave ourselves at the mercy of the multimillion-dollar pseudoscience industries that are so good at manipulating the media to their own ends (see Chapter 12). In subsequent chapters, we shall discuss in much more detail the high price we pay for ignoring the checks and balances inherent in the true scientific practice of psychology.

Empirically Solvable Problems: Scientists' Search for Testable Theories

Science deals with solvable, or specifiable, problems. This means that the types of questions that scientists address are potentially answerable by means of currently available empirical techniques. If a problem is not solvable or a theory is not testable by the empirical techniques that scientists have at hand, then scientists will not attack it. For example, the question "Will 3-year-old children given structured language stimulation during day care be ready for reading instruction at an earlier age than children not given such extra stimulation?" represents a scientific problem. It is answerable by currently available empirical methods. The question "Are human beings inherently good or inherently evil?" is not an empirical question and, thus, is simply not in the realm of science. Likewise, the question "What is the meaning of life?" is not an empirical question and so is outside the realm of science.

Science advances by positing theories to account for particular phenomena in the world, by deriving predictions from these theories, by testing the predictions empirically, and by modifying the theories based on the tests (the sequence is typically theory → prediction → test → modification). So what a scientist often means by the term *solvable problem* is "*testable* theory." What makes a theory testable? The theory must have specific implications for observable events in the natural world; this is what is meant by *empirically testable.* This criterion of testability is often termed the *falsifiability criterion,* and it is the subject of Chapter 2.

By saying that scientists tackle empirically solvable problems, we do not mean to imply that different classes of problems are inherently solvable or unsolvable and that this division is fixed forever. Quite the contrary: Some problems that are currently unsolvable may become solvable as theory and empirical techniques become more sophisticated. For example, decades ago historians would not have believed that the controversial issue of whether Thomas Jefferson fathered a child by his slave Sally Hemings was an empirically solvable question. Yet by 1998 this problem had become solvable through advances in genetic technology, and a paper was published in the journal *Nature* (Foster et al., 1998) indicating that it was highly probable that Jefferson was the father of Eston Hemings Jefferson.

This is how science in general has developed and how new sciences have come into existence. There is always ample room for disagreement about what is currently solvable. Scientists themselves often disagree on this point as it relates to current problems of ambiguous status. Thus, although all scientists agree on the solvability criterion, they may disagree on its specific applications. Nobel laureate Peter Medawar titled one of his books *The Art of the Soluble* (1967) to illustrate that part of the creativity involved in science is finding the problem on the furthest edge of the frontier of human knowledge that will yield to empirical techniques.

Psychology itself provides many good examples of the development from the unsolvable to the solvable. There are many questions (such as "How does a child acquire the language of his or her parents?" "Why do we forget things we once knew?" "How does being in a group change a person's behavior and thinking?") that had been the subjects of speculation for centuries before anyone recognized that they could be addressed by empirical means. As this recognition slowly developed, psychology coalesced as a collection of problems concerning behavior in a variety of domains. Psychological issues gradually became separated from philosophy, and a separate empirical discipline evolved.

Cognitive psychologist Steven Pinker (1997) discusses how ignorance can be divided into *problems* and *mysteries.* In the case of problems, we know that an answer is possible and what that answer might look like even though we might not actually have the answer yet. In the case of mysteries, we can't even conceive of what an answer might look like. Using this terminology, we can see that science is a process that turns mysteries into problems. In fact, Pinker (1997) noted that he wrote his book *How the Mind Works* "because dozens of mysteries of the mind, from mental images to romantic love, have recently been upgraded to problems" (p. ix).

Psychology and Folk Wisdom: The Problem With "Common Sense"

We all have implicit models of behavior that govern our interactions and our thoughts about ourselves and other people. Indeed, some social, personality, and cognitive psychologists study the nature of these implicit psychological theories. Rarely do we state our theories clearly and logically. Instead, we usually become aware of them only when attention is drawn to them or when we find them challenged in some way. Actually, our personal models of behavior are not really coherent in the way that an actual theory would have to be. Instead, we carry around a ragbag of general principles, homilies, and clichés about human behavior that we draw on when we feel that we need an explanation. The problem with this commonsense knowledge about behavior is that much of it contradicts itself and is, therefore, unfalsifiable (the principle of falsifiability is the topic of the next chapter).

Often a person uses some folk proverb to explain a behavioral event even though, on an earlier occasion, this same person used a directly contradictory folk proverb to explain the same type of event. For example, most of us have heard or said, "look before you leap." Now there's a useful, straightforward bit of behavioral advice—except that I vaguely remember admonishing on occasion, "he who hesitates is lost." And "absence makes the heart grow fonder" is a pretty clear prediction of an emotional reaction to environmental events. But then what about "out of sight, out of mind"? And if "haste makes waste," why does "time wait for no man"? How could the saying "two heads are better than one" not be true? Except that "too many cooks spoil the broth." If I think "it's better to be safe than sorry," why do I also believe "nothing ventured, nothing gained"? And if "opposites attract," why do "birds of a feather flock together"? I have counseled many students "never to put off until tomorrow what you can do today." But I hope my last advisee has never heard me say this because I just told him, "cross that bridge when you come to it."

The enormous appeal of clichés like these is that, taken together as implicit "explanations" of behavior, they cannot be refuted. No matter what happens, one of these explanations will be cited to cover it. No wonder we all think we are such excellent judges of human behavior and personality. We have an explanation for anything and everything that happens.

So sometimes our implicit psychological theories can't be refuted. We will see in the next chapter why this inability to be refuted makes such theories not very useful. However, a further problem occurs even in cases in which our folk beliefs do have some specificity, that is, even when they are empirically testable. The problem is that psychological research has shown that, when many common cultural beliefs about behavior are subjected to empirical test, they turn out to be false. Take, for example, common beliefs about the value of work experience to teenage high-school students. Most adult Americans think that if teenagers work while they attend school this is a good thing because (1) they will be earning money to help pay for their further education and for family expenses; (2) they will develop a "work ethic" that will make them more responsible employees in later life; (3) they will develop

a greater respect for our economy; and (4) they will be more motivated students because they are integrated into the economy.

Developmental psychologists have conducted extensive studies on the effect of working on the behavior, attitudes, and educational achievement of high-school students (Steinberg, Brown, & Dornbusch, 1996). They have found that virtually every one of our cultural beliefs about teenage employment is false. Only an extremely small amount of the money that teenagers earn goes to help with family expenses or to further their education. The vast majority of earnings are spent on luxury items that convey social status or for which television advertising has created a "need." Working while in high school has harmful effects on the students' education and educational experiences. And most interesting of all, work experience makes teenagers more cynical and less respectful of work and its value in our economy. In one study, working teenagers were more likely than nonworking teenagers to endorse statements like "People who work harder at their jobs than they have to are a little bit crazy" and "There's no such thing as a company that cares about its employees" (Greenberger & Steinberg, 1986). Finally, from a review of the research in this area, Greenberger and Steinberg concluded that "Working appears to promote, rather than deter, some forms of delinquent behavior" (p. 6). It appears that we have developed a substantial cultural myth about the value of work to young people. The common rhetoric about "building character" and "learning the value of money" is false. These clichés are folktales of the type that anthropologists are fond of studying in underdeveloped countries—tales that make us feel good and that justify current cultural practices but that have no basis in reality.

It is not difficult to generate other instances of folk beliefs (or "common sense") that are wrong. Take, for example, the idea that children who excel academically or who read a lot are not socially or physically adept. This idea still circulates in our society even though it is utterly false. There is voluminous evidence that, contrary to "commonsense" folk belief, readers and academically inclined individuals are more physically robust and are more socially involved than are people who do not read (Gage & Berliner, 1984, pp. 18–19; Zill & Winglee, 1990). For example, children high in scholastic achievement are more likely to be accepted by their peers than children low in achievement. People who read a lot are more likely to play sports, jog, camp, hike, and do car repair than are people who do not read very much.

Many of our folk beliefs about behavior arise and take on a life of their own. For example, throughout the 1990s the folk belief developed in our society and in schools that low self-esteem was a cause of aggression. But empirical investigations indicated that there was no connection between aggression and low self-esteem. If anything, the opposite appeared to be the case—aggression is more often associated with high self-esteem (Baumeister, Bushman, & Campbell, 2000). Likewise, an extremely popular hypothesis in the 1990s was that school achievement problems were the result of low self-esteem. In fact, it turns out that the relationship between self-esteem and school achievement is more likely to be in the opposite direction from that assumed by educators and parents. It is superior accomplishment in school (and in other aspects of life) that leads to high self-esteem and not the reverse (Baumeister, Campbell, Krueger, & Vohs, 2003; Stout, 2000).

Radford (1999) has discussed the folk myth that we use only 10 percent of our brainpower. Despite having absolutely no basis in cognitive neuroscience (see Beyerstein, 1999; Higbee & Clay, 1998), this one has been around for decades and has taken on the status of what has been termed a "psychofact." Radford quotes columnist Robert Samuelson's definition of a psychofact as "a belief that, though not supported by hard evidence, is taken as real because its constant repetition changes the way we experience life" (p. 53).

Folk beliefs are not always immune to evidence. Sometimes, when the contradictory evidence becomes too widely known, folk psychology ("common sense") does change. For example, years ago, one widely held cliché about children was "Early ripe, early rot" (Fancher, 1985, p. 141). The cliché reflected the belief that childhood precocity was associated with adult abnormality, a belief sustained by many anecdotes about childhood prodigies who came to ruin in later life. In this case, the psychological evidence documenting the inaccuracy of the cliché has been absorbed into the general culture, and you will almost never hear this bit of folk "wisdom" anymore.

This last example also carries a warning by reminding us to beware of today's "common sense"—because it is not difficult to show that yesterday's common sense has often turned into today's nonsense. After all, common sense is what "everybody knows," right? Right. Well, everybody knows that women shouldn't be able to vote, right? Everybody knows that African Americans shouldn't be taught to read, right? Everybody knows that individuals with disabilities should be institutionalized out of the sight of society, right? In fact, 150 years ago, all of these beliefs were what "everybody knew." Of course, we now recognize this common sense of the past as nonsense—as beliefs based on totally unverified assumptions. But in these examples we can see the critical role that psychology plays vis-à-vis common sense. Psychology tests the empirical basis of the assumptions that common sense rests on. Sometimes the assumptions do not hold up when tested, as we saw in many of the previous examples. From the examples discussed—and many more could be cited—we can see that psychology's role as the empirical tester of much folk wisdom often brings it into conflict with many widely held cultural beliefs. Psychology is often the bearer of the "bad tidings" that comfortable folk beliefs do not stand up to the cold light of day. Perhaps it is not surprising that many people would like not only to ignore the message but also to do away with the messenger.

Psychology as a Young Science

There has always been opposition to an empirically based psychology. Just 100 years ago, Cambridge University refused to establish a psychophysics laboratory because the study of such a topic would "insult religion by putting the human soul on a pair of scales" (Hearst, 1979, p. 7). Psychology's battle to establish its problems as empirically solvable has only recently been won. But as the science progresses, psychologists will address more and more issues that are the subject of strongly held beliefs about human beings because many of these problems are empirically testable. Psychologists now study topics such as the development of moral reasoning, the psychology of romantic love, the nature of racial prejudice, and the psychological and social determinants of religious beliefs. Studies of childhood sexual activity have recently incited much controversy (Hagen, 2001; Rind, Tromovitch, & Bauserman, 2001). Some people object to empirical investigation in these areas (Hunt, 1999); yet there has been scientific progress in each one of them.

Levin and O'Donnell (2000) discuss how opposition to some psychological research is based on what they claim is a "need *not* to know." They describe a school board where parents were given the option of having their child educated in K–2 multi-aged classrooms or in their usual age-graded classrooms. The school board disparaged their teachers' suggestion for a research study on the issue because they thought that if the research study showed one or the other method to be more effective, parents would force them to switch to this type of instruction completely. As Levin and O'Donnell (2000) note, "the school board simply did not want to know!" (p. 66). Thus, we should be aware that psychological research is often denigrated not because people think it is bad but because they desire to avoid the implications of the information that it might produce.

Psychology is often in a no-win situation as a discipline. On one hand, some people object to calling psychology a science and deny that psychologists can establish empirical facts about behavior. On the other hand, there are those who object to the investigation of certain areas of human behavior because they fear that facts uncovered by psychology might threaten their beliefs. Skinnerian psychologists regularly deal with these contradictory criticisms. For instance, critics have argued that the laws of reinforcement formulated by behaviorists do not apply to human behavior. At the same time, other critics are concerned that the laws will be used for the rigid and inhumane control of people. Thus, the behaviorists are faced with some critics who deny that their laws can be applied and others who charge that their laws can be applied too easily!

Examples such as this arise because the relatively new science of psychology has just begun to uncover facts about aspects of behavior that have previously escaped study. The relative youth of psychology as a science partially explains why many people are confused about the discipline. Nevertheless, during the past four decades, psychology has become firmly established in the interconnecting structure of knowledge that we call science. Failure to appreciate this fact is the source of almost all of the confused thinking about psychology that you will encounter.

Summary

Psychology is an immensely diverse discipline covering a range of subjects that are not always tied together by common concepts. Instead, what unifies the discipline is that it uses scientific methods to understand behavior. The scientific method is not a strict set of rules; instead it is defined by some very general principles. Three of the most important are that (1) science employs methods of systematic empiricism; (2) it aims for knowledge that is publicly verifiable; and (3) it seeks problems that are empirically solvable and that yield testable theories (the subject of the next chapter). The structured and controlled observations that define systematic empiricism are the subject of several later chapters of this book. Science renders knowledge public by procedures such as peer review and mechanisms such as replication.

Psychology is a young science and, thus, is often in conflict with so-called folk wisdom. This conflict is typical of all new sciences, but understanding it helps to explain some of the hostility directed toward psychology as a discipline. This characteristic of questioning common wisdom also makes psychology an exciting field. Many people are drawn to the discipline because it holds out the possibility of actually testing "common sense" that has been accepted without question for centuries.

2

Falsifiability

How to Foil Little Green Men in the Head

In 1793, a severe epidemic of yellow fever struck Philadelphia. One of tl
leading doctors in the city at the time was Benjamin Rush, a signer of tl
Declaration of Independence. During the outbreak, Rush was one of the fe
physicians who were available to treat literally thousands of yellow fev
cases. Rush adhered to a theory of medicine that dictated that illnesses a
companied by fever should be treated by vigorous bloodletting (the remov
of blood from the body either by using an instrument such as a lancet or b
the application of leeches). He administered this treatment to many patient
including himself when he came down with the illness. Critics charged th
his treatments were more dangerous than the disease. However, followin
the epidemic, Rush became even more confident of the effectiveness of h
treatment, even though several of his patients had died. Why?

One writer summarized Rush's attitude this way: "Convinced of tl
correctness of his theory of medicine and lacking a means for the systemat
study of treatment outcome, he attributed each new instance of improveme
to the efficacy of his treatment and each new death that occurred despite
to the severity of the disease" (Eisenberg, 1977, p. 1106). In other words,
the patient got better, this improvement was taken as proof that bloodlettin
worked. If instead the patient died, Rush interpreted this to mean that the pa
tient had been too ill for *any* treatment to work. We now know that Rush
critics were right: His treatments were as dangerous as the disease. In th
chapter, we will discuss how Rush went wrong. His error illustrates one c
the most important principles of scientific thinking, one that is particularl
useful in evaluating psychological claims.

In this chapter, we focus in more detail on the third general characteristic of science that we discussed in Chapter 1: Scientists deal with solvable problems. What scientists most often mean by a *solvable problem* is a "testable theory." The way scientists make sure they are dealing with testable theories is by ensuring that their theories are falsifiable, that is, that they have implications for actual events in the natural world. We will see why what is termed the *falsifiability criterion* is so important in psychology.

Theories and the Falsifiability Criterion

Benjamin Rush fell into a fatal trap when assessing the outcome of his treatment. His method of evaluating the evidence made it impossible to conclude that his treatment did not work. If the recovery of a patient meant confirmation of his treatment (and, hence, his theory of medicine), then it only seems fair that the death of a patient should have meant disconfirmation. Instead, he rationalized away these disconfirmations. By interpreting the evidence as he did, Rush violated one of the most important rules regarding the construction and testing of theories in science: He made it impossible to falsify his theory.

Scientific theories must always be stated in such a way that the predictions derived from them could potentially be shown to be false. Thus, the methods of evaluating new evidence relevant to a particular theory must always include the possibility that the data will falsify the theory. This principle is often termed the *falsifiability criterion*, and its importance in scientific progress has been most forcefully articulated by Karl Popper, a philosopher of science whose writings are read widely by working scientists (Magee, 1985).

The falsifiability criterion states that, for a theory to be useful, the predictions drawn from it must be specific. The theory must go out on a limb, so to speak, because in telling us what *should* happen, the theory must also imply that certain things will *not* happen. If these latter things *do* happen, then we have a clear signal that something is wrong with the theory: It may need to be modified, or we may need to look for an entirely new theory. Either way, we shall end up with a theory that is nearer to the truth. In contrast, if a theory does not rule out any possible observations, then the theory can never be changed, and we are frozen into our current way of thinking, with no possibility of progress. Thus, a successful theory is not one that accounts for every possible outcome because such a theory robs itself of any predictive power.

Because we shall often refer to the evaluation of theories in the remainder of this book, we must clear up one common misconception surrounding the term *theory*. The misconception is reflected in the commonly used phrase "Oh, it's only a theory." This phrase captures what *laypeople* often mean when they use the term *theory*: an unverified hypothesis, a mere guess, a hunch. It

implies that one theory is as good as another. This is most definitely *not* the way the term *theory* is used in science. When scientists refer to theories, they do not mean unverified guesses.

A theory in science is an interrelated set of concepts that is used to explain a body of data and to make predictions about the results of future experiments. *Hypotheses* are specific predictions that are derived from theories (which are more general and comprehensive). Currently viable theories are those that have had many of their hypotheses confirmed. The theoretical structures of such theories are, thus, consistent with a large number of observations. However, when the database begins to contradict the hypotheses derived from a theory, scientists begin trying to construct a new theory (or, more often, simply make adjustments in the previous theory) that will provide a better interpretation of the data. Thus, the theories that are under scientific discussion are those that have been verified to some extent and that do not make many predictions that are contradicted by the available data. They are not mere guesses or hunches.

The difference between the layperson's and the scientist's use of the term *theory* has often been exploited by some religious fundamentalists who want creationism taught in the public schools (Forrest & Gross, 2004; Scott, 2005; Talbot, 2005). Their argument often is "After all, evolution is only a theory." This statement is intended to suggest the *layperson's* usage of the term *theory* to mean "only a guess." However, the theory of evolution by natural selection is not a theory in the layperson's sense (to the contrary, in the layperson's sense it would be called a fact; see Randall, 2005). Instead, it is a theory in the *scientific* sense. It is a conceptual structure that is supported by a large and varied set of data (Maynard Smith, 1998; Ridley, 1996, 1999; Scott, 2005). It is not a mere guess, equal to any other guess. Instead, it interlocks with knowledge in a host of other disciplines, including geology, physics, chemistry, and all aspects of biology. The distinguished biologist Theodosius Dobzhansky (1973) made this point by titling a well-known article, "Nothing in Biology Makes Sense Except in the Light of Evolution."

The Theory of Knocking Rhythms

A hypothetical example will show how the falsifiability criterion works. A student knocks at my door. A colleague in my office with me has a theory that makes predictions about the rhythms that different types of people use to knock. Before I open the door, my colleague predicts that the person behind it is a female. I open the door and, indeed, the student is a female. Later I tell my colleague that I am impressed, but only mildly so because he had a 50 percent chance of being correct even without his "theory of knocking rhythms." He says he can do better. Another knock comes. My colleague tells me it is a male under 22 years old. I open the door to find a male student whom I know to be just out of high school. I comment that I am somewhat impressed because our university has a considerable number of students over the age of 22. Yet I still maintain that, of course, young males are quite common on campus. Thinking me hard to please, my colleague proposes one last test. After the next knock, my colleague predicts, "Female, 30 years old, 5 feet 2 inches tall, carrying a book and a purse in the left hand and knocking with the right." After opening the door and confirming the prediction completely, I have quite a different response. I say that, assuming my colleague did not play a trick and arrange for these people to appear at my door, I am now in fact extremely impressed.

Why the difference in my reactions? Why do my friend's three predictions yield three different responses, ranging from "So what?" to "Wow!"? The answer has to do with the specificity and precision of the predictions. The more specific predictions made a greater impact when they were confirmed. Notice, however, that the specificity varied directly with the falsifiability. The more specific and precise the prediction was, the more potential observations there were that could have falsified it. For example, there are a lot of people who are *not* 30-year-old females and 5 feet 2 inches tall. Note that implicitly, by my varied reactions, I signaled that I would be more impressed by a theory that made predictions that maximized the number of events that should *not* occur.

Good theories, then, make predictions that expose themselves to falsification. Bad theories do not put themselves in jeopardy in this way. They make predictions that are so general that they are almost bound to be true (for example, the next person to knock on my door will be less than 100 years old) or are phrased in such a way that they are completely protected from falsification (as in the Benjamin Rush example). In fact, a theory can be so protected from falsifiability that it is simply no longer considered scientific at all. Indeed, it was philosopher Karl Popper's attempt to define the criteria that separate science from nonscience that led him to emphasize the importance of the falsifiability principle. There is a direct link here to psychology and to our discussion of Freud in Chapter 1.

Freud and Falsifiability

In the early decades of this century, Popper was searching for the underlying reasons that some scientific theories seem to lead to advances in knowledge and others lead to intellectual stagnation (Hacohen, 2000). Einstein's general relativity theory, for example, led to startlingly new observations (for instance, that the light from a distant star bends when it passes near the sun) precisely because its predictions were structured so that many possible events could have contradicted them and, thus, falsified the theory.

Popper reasoned that this is not true of stagnant theories and pointed to Freudian psychoanalysis as an example. Freudian theory uses a complicated conceptual structure that explains human behavior after the fact but does not predict things in advance. It can explain everything, but Popper argued, it is precisely this property that makes it scientifically useless. It makes no specific predictions. Adherents of psychoanalytic theory spend much time and effort in getting the theory to explain every known human event, from individual quirks of behavior to large-scale social phenomena, but their success in making the theory a rich source of after-the-fact explanation robs it of any scientific utility. Freudian psychoanalytic theory currently plays a much larger role as a spur to the literary imagination than as a theory in contemporary psychology (Robins, Gosling, & Craik, 1999, 2000). Its demise within psychology can be traced in part to its failure to satisfy the falsifiability criterion.

But the existence of such unfalsifiable theories does real damage. As one critic has argued, "Incorrect but widely dispersed ideas about the mind inevitably end by causing social damage. Thanks to the once imposing prestige of psychoanalysis, people harboring diseases or genetic conditions have deferred effective treatment while scouring their infantile past for the sources of their trouble" (Crews, 1993, p. 65). Take, for example, the history of Gilles de la Tourette syndrome. This is a disorder characterized by physical tics and twitches that may involve any part of the body, as well as vocal symptoms such as grunts and barks, echolalia (involuntary repetition of the words of others), and coprolalia (compulsive repetition of obscene words). Tourette syndrome is an organically based disorder of the central nervous system and is now successfully treated with drug therapies (Bower, 1990, 1996a). Throughout history, individuals with Tourette syndrome have been persecuted—earlier as witches by religious authorities, and in more modern times by being subjected to exorcisms (Hines, 2003). Importantly, understanding of the cause and treatment of the disorder was considerably hampered from 1921 to 1955, when explanations and treatments for Tourette syndrome were dominated by psychoanalytic conceptualizations (see Kushner, 1999). Author after author presented unfalsifiable psychoanalytic explanations for the syndrome. The resulting array of vague explanations created a conceptual sludge that obscured the true nature of the syndrome and probably impeded scientific progress toward an accurate understanding of it. For example, according to one author,

> [Tourette syndrome is] a classic example of the retrogressive effect of psychoanalysis on the investigation of brain disease. La Tourette had attributed the disease to a degenerative process of the brain. After Freud's theories became fashionable in the early decades of the present century, attention in such conditions was deflected from the brain. . . . The consequence of this retrograde movement was that patients tended to be referred to psychiatrists (usually of a psychoanalytic persuasion) rather than to neurologists, so that physical examinations and investigation were not performed. (Thornton, 1986, p. 210)

Shapiro et al. (1978) described one psychoanalyst who thought that his patient was "reluctant to give up the tic because it became a source of erotic pleasure to her and an expression of unconscious sexual strivings." Another considered the tics "stereotyped equivalents of onanism. . . . The libido connected with the genital sensation was displaced into other parts of the body." A third considered the tic a "conversion symptom at the anal-sadistic level." A fourth thought that a person with Tourette syndrome had a "compulsive character, as well as a narcissistic orientation" and that the patient's tics "represent[ed] an affective syndrome, a defense against the intended affect." The summary by Shapiro et al. (1978) of the resulting theoretical situation demonstrates quite well the harmful effects of ignoring the falsifiability criterion:

> Psychoanalytic theorizing of this kind in effect leaves no base untouched. Tics are a conversion symptom but not hysterical, anal but also erotic, volitional but also compulsive, organic but also dynamic in origin. . . . These psychological labels, diagnoses, and treatments were unfortunately imposed on patients and their families, usually with little humility, considerable dogmatism, and with much harm. . . . These papers, because of their subsequent widespread influence, had a calamitous effect on the understanding and treatment of this syndrome. (pp. 39–42, 50, 63)

Progress in the treatment and understanding of Tourette syndrome began to occur only when researchers recognized that the psychoanalytic "explanations" were useless. These explanations were enticing because they seemed to explain things. In fact, they explained everything—after the fact. However, the explanations they provided created only the illusion of understanding. By attempting to explain everything after the fact, they barred the door to any advance. Progress occurs only when a theory does not predict *everything* but instead makes specific predictions that tell us—in advance—something specific about the world. The predictions derived from such a theory may be wrong, of course, but this is a strength, not a weakness.

The Little Green Men

It is not difficult to recognize unfalsifiable conceptualizations when one is detached from the subject matter and particularly when one has the benefit of historical hindsight (as in the Benjamin Rush example). It is also easy to detect unfalsifiable conceptualizations when the instance is obviously concocted. For example, it is a little known fact that I have discovered the underlying brain mechanism that controls behavior. You will soon be reading about this discovery (in the *National Enquirer*, available at your local supermarkets). In the left hemisphere of the brain, near the language areas, reside two tiny green men. They have the power to control the electrochemical processes taking place in many areas of the brain. And, well, to make a long story short, they basically control everything. There is one difficulty, however. The little green men have the ability to detect any intrusion into the brain (surgery, X-rays, etc.), and when they do sense such an intrusion, they tend to disappear. (I forgot to mention that they have the power to become invisible.)

I have no doubt insulted your intelligence by using an example more suitable for elementary school students. I have obviously concocted this example so that my hypothesis about the little green men could never be shown to be wrong. However, consider this. As a lecturer and public speaker on psychological topics, I am often confronted by people who ask me why I have not lectured on all the startling new discoveries in extrasensory perception (ESP) and parapsychology that have been made in the past few years. I have to inform these questioners that most of what they have heard about these subjects has undoubtedly come from the general media rather than from scientifically respectable sources. In fact, some scientists have looked at these claims and have not been able to replicate the findings. I remind the audience that replication of a finding is critical to its acceptance as an established scientific fact and that this is particularly true in the case of results that contradict either previous data or established theory.

I further admit that many scientists have lost patience with ESP research. Although one reason is undoubtedly that the area is tainted by fraud, charlatanism, and media exploitation, perhaps the most important reason for scientific disenchantment is the existence of what Martin Gardner (1972) years ago called the catch-22 of ESP research.

It works as follows: A "believer" (someone who accepts the existence of ESP phenomena before beginning an investigation) claims to have demonstrated ESP in the laboratory. A "skeptic" (someone who doubts the existence of ESP) is brought in to confirm the phenomena. Often, after observing the experimental situation, the skeptic calls for more controls (controls of the type we will discuss in Chapter 6), and though these are sometimes resisted, well-intentioned believers often agree to them. When the controls are instituted, the phenomena cannot be demonstrated (see Alcock, 1990; Hines, 2003; Humphrey, 1996; Hyman, 1992, 1996; Kelly, 2005; Marks, 2001; Milton & Wiseman, 1999). The skeptic, who correctly interprets this failure as an indication that the original demonstration was due to inadequate experimental control and, thus, cannot be accepted, is often shocked to find that the believer does not regard the original demonstration as invalid. Instead, the believer invokes the catch-22 of ESP: Psychic powers, the believer maintains, are subtle, delicate, and easily disturbed. The "negative vibes" of the skeptic were probably responsible for the disruption of the "psi powers." The believer thinks that the powers will undoubtedly return when the negative aura of the skeptic is removed.

This way of interpreting failures to demonstrate ESP in experiments is logically analogous to my story about the little green men. ESP operates just as the little green men do. It's there as long as you don't intrude to look at it carefully. When you do, it disappears. If we accept this explanation, it will be impossible to demonstrate the phenomenon to any skeptical observers. It appears only to believers. Of course, this position is unacceptable in science. We do not have the magnetism physicists and the nonmagnetism physicists (those for whom magnetism does and does not "work"). Interpreting ESP experiments in this way makes the hypothesis of ESP unfalsifiable just as the hypothesis of the little green men is. Interpreting the outcomes in this way puts it outside the realm of science.

Not All Confirmations Are Equal

The principle of falsifiability has important implications for the way we view the confirmation of a theory. Many people think that a good scientific theory is one that has been confirmed repeatedly. They assume that the amount of confirming evidence is critical in the evaluation of a theory. But falsifiability implies that the number of times a theory has been confirmed is not the critical element. The reason is that, as our example of the "theory of knocking rhythms" illustrated, not all confirmations are equal. Confirmations are more or less impressive depending on the extent to which the prediction exposes itself to potential disconfirmation. One confirmation of a highly specific, potentially falsifiable prediction (for instance, a female, 30 years old, 5 feet 2 inches tall, carrying a book and a purse in the left hand and knocking with the right) has a greater impact than the confirmation of 20 different predictions that are all virtually unfalsifiable (for instance, a person less than 100 years old).

Thus, we must look not only at the *quantity* of the confirming evidence, but also at the *quality* of the confirming instances. Using the falsifiability criterion as a tool to evaluate evidence will help the research consumer resist the allure of the nonscientific, all-explaining theory that inevitably hinders the search for deeper understanding. Indeed, such theoretical dead ends are often tempting precisely because they can never be falsified. They are islands of stability in the chaotic modern world.

Popper often made the point that "the secret of the enormous psychological appeal of these [unfalsifiable] theories lay in their ability to explain everything. To know in advance that whatever happens you will be able to understand it gives you not only a sense of intellectual mastery but, even more important, an emotional sense of secure orientation in the world" (Magee, 1985, p. 43). However, the attainment of such security is not the goal of science, because such security would be purchased at the cost of intellectual stagnation. Science is a mechanism for continually challenging previously held beliefs by subjecting them to empirical tests in such a way that they can be shown to be wrong. This characteristic often puts science—particularly psychology—in conflict with so-called folk wisdom or common sense (as we discussed in Chapter 1).

Falsifiability and Folk Wisdom

Psychology is a threat to the comfort that folk wisdom provides because, as a science, it cannot be content with explanations that cannot be refuted. The goal of psychology is the empirical testing of alternative behavioral theories in order to rule out some of them. Aspects of folk wisdom that are explicitly stated and that do stand up to empirical testing are, of course, welcomed, and many have been incorporated into psychological theory. However, psychology does not seek the comfort of explanatory systems that account for everything after the fact but predict nothing in advance. It does not accept systems of folk wisdom that are designed never to be changed and that end up being passed on from generation to generation. It is self-defeating to try to hide this fact from students or the public. Unfortunately, some psychology instructors and popularizers are aware that psychology's threat to folk wisdom disturbs some people, and they sometimes seek to soothe such feelings by sending a false underlying message that implies, "You'll learn some interesting things, but don't worry—psychology won't challenge things you believe in strongly." This is a mistake, and it contributes to confusion both about what science is and about what psychology is.

Science *seeks* conceptual change. Scientists try to describe the world as it really is, as opposed to what our prior beliefs dictate it should be. The dangerous trend in modern thought is the idea that people must be shielded from the nature of the world—that a veil of ignorance is necessary to protect a public unequipped to deal with the truth. Psychology is like other sciences in rejecting the idea that people need to be shielded from the truth. Biologist Michael Ghiselin (1989) argued further that we *all* lose when knowledge is not widespread:

> We are better off if we have healthy neighbors, and it would be utter folly to monopolize the supply of medicine in order to be more healthy than they are. So too with knowledge. Our neighbor's ignorance is as bad for us as his ill health, and may indeed be the cause of it. Industry and all the rest of us benefit from a supply of skilled labor. We rely upon others for their skill and expertise. (p. 192)

Psychologists feel, like Ghiselin, that we all lose when we are surrounded by others who hold incorrect views of human behavior. Our world is shaped by public attitudes toward education, crime, health, industrial productivity, child care, and many other critical issues. If these attitudes are the products of incorrect theories of behavior, then we are all harmed.

The Freedom to Admit a Mistake

Scientists have found that one of the most liberating and useful implications of the falsifiability principle is that, in science, making a mistake is not a sin. Falsified hypotheses provide information that scientists use to adjust their theories so that these theories accord more closely with the data. Philosopher Daniel Dennett (1995) has said that the essence of science is "making mistakes in public" (p. 380). By the process of continually adjusting theory when data do not accord with it, scientists eventually arrive at theories that better reflect the nature of the world.

In fact, our way of operating in everyday life might be greatly improved if we could use the falsifiability principle on a personal level. This is why the word *liberating* was used in the opening sentence of this section. It has a personal connotation that was specifically intended, for the ideas developed here have implications beyond science. We would have many fewer social and personal problems if we could only understand that, when our beliefs are contradicted by evidence in the world, it is better to adjust our beliefs than to deny the evidence and cling tenaciously to dysfunctional ideas. Physicist J. Robert Oppenheimer argued,

> There's a point in anybody's train of reasoning when he looks back and says, "I didn't see this straight." People in other walks of life need the ability to say without shame, "I was wrong about that." Science is a method of having this happen all the time. You notice a conflict or some oddity about a number of things you've been thinking about for a long time. It's the shock that may cause you to think another thought. That is the opposite of the worldly person's endless web of rationalization to vindicate an initial error. (Dos Passos, 1964, pp. 150–151)

How many times have you been in a hot argument with someone when right in the middle—perhaps just as you were giving a heated reply and defending your point of view—you realized that you were wrong about some critical fact or piece of evidence? What did you do? Did you back down and admit to the other person that you had assumed something that wasn't true and that the other person's interpretation now seemed more correct to you? Probably not. If you are like most of us, you engaged in an "endless web of rationalization to vindicate an initial error." You tried to extricate yourself from the argument without admitting defeat. The last thing you would have done was admit that you were wrong. Thus, both you and your partner in the argument became a little more confused about which beliefs more closely tracked the truth. If refutations never become public (as they do in science), if both true and false beliefs are defended with equal vehemence, and if the correct feedback about the effects of argument is not given (as in this example), there is no mechanism for getting beliefs more reliably in sync with reality. This is why so much of our private and public discourse is confused and why the science of psychology is a more reliable guide to the causes of behavior than is so-called common sense.

Many scientists have attested to the importance of understanding that making errors in the course of science is normal and that the real danger to scientific progress is our natural human tendency to avoid exposing our beliefs

to situations in which they might be shown to be wrong. Nobel Prize winner Peter Medawar (1979) wrote,

> Though faulty hypotheses are excusable on the grounds that they will be superseded in due course by acceptable ones, they can do grave harm to those who hold them because scientists who fall deeply in love with their hypotheses are proportionately unwilling to take no as an experimental answer. Sometimes instead of exposing a hypothesis to a cruelly critical test, they caper around it, testing only subsidiary implications, or follow up sidelines that have an indirect bearing on the hypothesis without exposing it to the risk of refutation. . . . I cannot give any scientist of any age better advice than this: *the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not.* (p. 39; italics in original)

Many of the most renowned scientists in psychology have followed Medawar's advice. In an article on the career of noted experimental psychologist Robert Crowder, one of his colleagues, Mahzarin Banaji, is quoted as saying that "he is the least defensive scientist I know. If you found a way to show that his theory was wobbly or that his experimental finding was limited or flawed, Bob would beam with pleasure and plan the demise of his theory with you" (Azar, 1999, p. 18). Azar (1999) describes how Crowder developed a theory of one component of memory called precategorical acoustic storage and then carefully designed the studies that falsified his own model. Finally, in an eloquent statement about the attitudes that led to Darwin's monumental contributions to science, evolutionary psychologist John Tooby (2002) said that "Darwin went farther than his contemporaries because he was less bound by the compulsion to make the universe conform to his predilections" (p. 12). Philosopher Jonathan Adler (1998) puts it another way: "What truly marks an open-minded person is the willingness to follow where evidence leads. The open-minded person is willing to defer to impartial investigations rather than to his own predilections. . . . Scientific method is attunement to the world, not to ourselves" (p. 44).

But the falsifying attitude doesn't always have to characterize each and every scientist for *science* to work. Jacob Bronowski (1973, 1977) often argued in his many writings that the unique power of science to reveal knowledge about the world does *not* arise because scientists are uniquely virtuous (that they are completely objective; that they are never biased in interpreting findings, etc.) but instead it arises because fallible scientists are immersed in a process of checks and balances—in a process in which other scientists are always there to criticize and to root out the errors of their peers. Philosopher Daniel Dennett (2000) has made the same point by arguing that it is not necessary for every scientist to display the objectivity of Robert Crowder. Like Bronowski, Dennett stresses that "scientists take themselves to be just as weak and fallible as anybody else, but recognizing these very sources of error in themselves and in the groups to which they belong, they have devised elaborate systems to tie their own hands, forcibly preventing their frailties and prejudices from infecting their results" (p. 42). More humorously, psychologist Ray Nickerson (1998) makes the related point that the vanities of scientists are actually put to use by the scientific process, by noting that it is "not so much the critical attitude that individual scientists have taken with respect to their own ideas that has given science its success . . . but more the fact that individual scientists have been highly motivated to demonstrate that hypotheses that are held by some other scientists are false" (p. 32). These authors suggest that the strength of scientific knowledge comes not because scientists are virtuous, but from the social process where scientists constantly cross-check each others' knowledge and conclusions.

Thoughts Are Cheap

Our earlier discussion of the idea of testing folk wisdom leads us to another interesting corollary of the falsifiability principle: Thoughts are cheap. To be specific, what we mean here is that certain *kinds* of thoughts are cheap. Biologist and science writer Stephen J. Gould (1987) illustrated this point:

> Fifteen years of monthly columns have brought me an enormous correspondence from nonprofessionals about all aspects of science. . . . I have found that one common misconception surpasses all others. People will write, telling me that they have developed a revolutionary theory, one that will expand the boundaries of science. These theories, usually described in several pages of single-spaced typescript, are speculations about the deepest ultimate questions we can ask—what is the nature of life? the origin of the universe? the beginning of time? But thoughts are cheap. Any person of intelligence can devise his half dozen before breakfast. Scientists can also spin out ideas about ultimates. We don't (or, rather, we confine them to our private thoughts) because we cannot devise ways to test them, to decide whether they are right or wrong. What good to science is a lovely idea that cannot, as a matter of principle, ever be affirmed or denied? (p. 18)

The answer to Gould's last question is "No good at all." The type of thoughts that Gould is saying are cheap are those that we referred to earlier in our discussion of Karl Popper's views: grand theories that are so global, complicated, and "fuzzy" that they can be used to explain anything—theories constructed more for emotional support because they are not meant to be changed or discarded. Gould was telling us that such theories are useless for scientific purposes, however comforting they may be. Science is a creative endeavor, but the creativity involves getting conceptual structures to fit the confines of empirical data. This is tough. These types of thoughts—those that explain the world as it actually is—are not cheap. Probably this is why good scientific theories are so hard to come by and why unfalsifiable pseudoscientific belief systems proliferate everywhere—the latter are vastly easier to construct.

Theories in science make contact with the world. They are falsifiable. They make specific predictions. Actually coming up with the theories that are truly scientific explanations is a difficult task. However, understanding the general logic by which science works is *not* so difficult. Indeed, there are books about the logic of scientific thinking that have been written for children (Kramer, 1987; Swanson, 2001, 2004).

Errors in Science: Getting Closer to the Truth

In the context of explaining the principle of falsifiability, we have outlined a simple model of scientific progress. Theories are put forth and hypotheses are derived from them. The hypotheses are tested by a variety of techniques that we shall discuss in the remainder of this book. If the hypotheses are confirmed by the experiments, then the theory receives some degree of corroboration. If the hypotheses are falsified by the experiments, then the theory must be altered in some way, or it must be discarded for a better theory.

Of course, saying that knowledge in science is tentative and that hypotheses derived from theories are potentially false does not mean that everything is up for grabs. There are many relationships in science that have been confirmed so many times that they are termed *laws* because it is extremely doubtful that they will be overturned by future experimentation. It is highly unlikely that we shall find one day that blood does not circulate or that the earth does not orbit the sun. These mundane facts are not the type of hypotheses that we have been talking about. They are of no interest to scientists precisely because they are so well established. Scientists are interested only in those aspects of nature that are on the fringes of what is known: that is, those things that are not so well confirmed that there is no doubt about them.

This aspect of scientific practice—that scientists gravitate to those problems on the fringes of what is known and ignore things that are well confirmed (so-called laws)—is very confusing to the general public. It seems that scientists are always emphasizing what they don't know rather than what is known. This is true, and there is a very good reason for it. To advance knowledge, scientists must be at the outer limits of what is known. Of course, this is precisely where things are uncertain. But science advances by a process of trying to reduce the uncertainty at the limits of knowledge. This can often make scientists look "uncertain" to the public. But this perception is deceiving. Scientists are only uncertain at the fringes of knowledge—where our understanding is currently being advanced. Scientists are *not* uncertain about the many facts that have been well established by replicable research.

It should also be emphasized that, when scientists talk about falsifying a theory based on observation and about replacing an old, falsified theory with a new one, they do not mean that all the previous facts that established the old theory are thrown out (we shall talk about this at great length in Chapter 8). Quite the contrary, the new theory must explain all of the facts that the old theory could explain plus the new facts that the old theory couldn't explain. So the falsification of a theory does not mean that scientists have to go back to square one. Science writer Isaac Asimov illustrated the process of theory revision very well in an essay titled "The Relativity of Wrong" (1989), in which he wrote about how we have refined our notions of the earth's shape. First, he warned us not to think that the ancient belief in a flat earth was stupid. On a plain (where the first civilizations with writing developed), the earth looks pretty flat, and Asimov urged us to consider what a quantitative comparison of different theories would reveal. First, we could express the different theories in terms of how much curvature per mile they hypothesized. The flat-earth theory would say that the curvature is 0 degrees per mile. This theory is wrong, as we know. But in one sense, it is close. As Asimov (1989) wrote,

> About a century after Aristotle, the Greek philosopher Eratosthenes noted that the sun cast a shadow of different lengths at different latitudes (all the shadows would be the same length if the earth's surface were flat). From the difference in shadow length, he calculated the size of the earthly sphere and it turned out to be 25,000 miles circumference. The curvature of such a sphere is about 0.000126 degrees per mile, a quantity very close to 0 per mile, as you can see. . . . The tiny difference between 0 and 0.000126 accounts for the fact that it took so long to pass from the flat earth to the spherical earth. Mind you, even a tiny difference, such as that between 0 and 0.000126, can be extremely important. The difference mounts up. The earth cannot be mapped over large areas with any accuracy at all if the difference isn't taken into account and if the earth isn't considered a sphere rather than a flat surface. (pp. 39–40)

But science, of course, did not stop with the theory that the earth was spherical. As we discussed earlier, scientists are always trying to refine their theories as much as possible and to test the limits of current knowledge. For example, Newton's theories of gravitation predicted that the earth should not be perfectly spherical, and indeed this prediction has been confirmed. It turns out that the earth bulges a little at the equator and that it is a little flat at the poles. It is something called an *oblate spheroid*. The diameter of the earth from North Pole to South Pole is 7,900 miles, and the equatorial diameter is 7,927 miles. The curvature of the earth is not constant (as in a perfect sphere); instead, it varies slightly from 7.973 inches to 8.027 inches to the mile. As Asimov (1989) noted, "The correction in going from spherical to oblate spheroidal is much smaller than going from flat to spherical. Therefore, although the notion of the earth as a sphere is wrong, strictly speaking, it is not as wrong as the notion of the earth as flat" (p. 41). Asimov's example of the shape of the earth illustrates for us the context in which scientists use such terms as *mistake, error,* or *falsified.* Such terms do not mean that the theory being tested is wrong in every respect, only that it is incomplete. So when scientists emphasize that knowledge is tentative and may be altered by future findings, they are referring to a situation such as this example. When scientists believed that the earth was a sphere, they realized that in detail, this theory might someday need to be altered. However, the alteration from spherical to oblate spheroidal preserves the "roughly correct" notion that the earth is a sphere. We do not expect to wake up one day and find that it is a cube.

Clinical psychologist Scott Lilienfeld (2005) contextualizes Asimov's point for the psychology student:

> When explaining to students that scientific knowledge is inherently tentative and open to revision, some students may mistakenly conclude that genuine knowledge is impossible. This view, which is popular in certain postmodernist circles, neglects to distinguish knowledge claims that are more certain from those that are less certain. Although absolute certainly is probably unattainable in science, some scientific claims, such as Darwin's theory of natural selection, have been extremely well corroborated, whereas others, such as the theory underpinning astrological horoscopes, have been convincingly refuted. Still others, such as cognitive dissonance theory, are scientifically controversial. Hence, there is a continuum of confidence in scientific claims; some have acquired virtual factual status whereas others have been resoundingly falsified. The fact that methodological skepticism does not yield completely certain answers to scientific questions—and that such answers could in principle be overturned by new evidence—does not imply that knowledge is impossible, only that this knowledge is provisional. (p. 49)

Summary

What scientists most often mean by a *solvable problem* is a *testable theory.* The definition of a testable theory is a very specific one in science: It means that the theory is potentially falsifiable. If a theory is not falsifiable, then it has no implications for actual events in the natural world and, hence, is useless. Psychology has been plagued by unfalsifiable theories, and that is one reason why progress in the discipline has been slow.

Good theories are those that make specific predictions, and such theories are highly falsifiable. The confirmation of a specific prediction provides more support for the theory from which it was derived than the confirmation of a prediction that was not precise. In short, one implication of the falsifiability criterion is that all confirmations of theories are not equal. Theories that receive confirmation from highly falsifiable, highly specific predictions are to be preferred. Even when predictions are not confirmed (i.e., when they are falsified), this falsification is useful to theory development. A falsified prediction indicates that a theory must either be discarded or altered so that it can account for the discrepant data pattern. Thus, it is by theory adjustment caused by falsified predictions that sciences such as psychology get closer to the truth.

3

Operationism and Essentialism

"But, Doctor, What Does It Really Mean?"

Do physicists really know what gravity is? I mean *really*. What is the real *meaning* of the term *gravity?* What is the underlying essence of it? What does it ultimately mean even to speak of gravity? When you get down to rock bottom, what is it all about?

Questions such as these reflect a view of science that philosopher Karl Popper called *essentialism*. This is the idea that the only good scientific theories are those that give ultimate explanations of phenomena in terms of their underlying essences or their essential properties. People who hold this view usually also believe that any theory that gives less than an ultimate explanation of a phenomenon is useless. It does not reflect the true underlying situation, the essence of the way the world is. In this chapter, we shall discuss why science does not answer essentialist questions and why, instead, science advances by developing *operational definitions* of concepts.

Why Scientists Are Not Essentialists

Scientists, in fact, do not claim to acquire the type of knowledge that the essentialist seeks. The proper answer to the preceding questions is that physicists do *not* know what gravity is in this sense. Science does not attempt to answer "ultimate" questions about the universe. Peter Medawar (1984) wrote,

> [There exist] questions that science cannot answer and that no conceivable advance of science would empower it to answer. These are the questions that children ask—the "ultimate questions." . . . I have in mind such questions as: How did everything begin? What are we all here for? What is the point of living? (p. 66)
>
> [However,] the failure of science to answer questions about first and last things does not in any way entail the acceptance of answers of other kinds; nor can it be taken for granted that because these questions can be put they can be answered. So far as our understanding goes, they cannot. (p. 60)
>
> [Finally, however,] there is no limit upon the ability of science to answer the kind of questions that science can answer. . . . Nothing can impede or halt the advancement of scientific learning except a moral ailment such as the failure of nerve. (p. 86)

One reason that scientists are suspicious of claims that some person, theory, or belief system provides absolute knowledge about ultimate questions is that scientists consider questions about "ultimates" to be unanswerable. Scientists do not claim to produce perfect knowledge; the unique strength of science is not that it is an error-free process, but that it provides a way of eliminating the errors that are part of our knowledge base. Furthermore, claims of perfect or absolute knowledge tend to choke off inquiry. Because a free and open pursuit of knowledge is a prerequisite for scientific activity, scientists are always skeptical of claims that the ultimate answer has been found.

Essentialists Like to Argue About the Meaning of Words

A common indication of the essentialist attitude is an obsessive concern about defining the meaning of terms and concepts before the search for knowledge about them begins. "But we must first define our terms" is a frequent essentialist slogan. "What does that theoretical concept really *mean?*" The idea seems to be that, before a word can be used as a concept in a theory, we must have a complete and unambiguous understanding of all the underlying language problems involved in its usage. In fact, this is exactly the opposite of the way scientists work. Before they begin to investigate the physical world, physicists do not engage in debates about how to use the word *energy* or whether the word *particle* really captures the essence of what we mean when we talk about the fundamental constituents of matter.

The meaning of a concept in science is determined *after* extensive investigation of the phenomena the term relates to, not before such an investigation. The refinement of conceptual terms comes from the interplay of data and theory that is inherent in the scientific process, not from debates on language usage. Essentialism leads us into endless argument about words, and

many scientists believe that such language games distract us from matters of substance. For example, concerning the question "What is the true meaning of the word *life?*" two biologists answer "There is no true meaning. There is a usage that serves the purposes of working biologists well enough, and it is not the subject of altercation or dispute" (Medawar & Medawar, 1983, pp. 66–67). In short, the explanation of phenomena, not the analysis of language, is the goal of the scientist. The key to progress in all the sciences has been to abandon essentialism and to adopt operationism, our topic of inquiry in this chapter. Nowhere is this more evident than in psychology.

Operationists Link Concepts to Observable Events

Where, then, does the meaning of concepts in science come from if not from discussions about language? What are the criteria for the appropriate use of a scientific concept? To answer these questions, we must discuss operationism, an idea that is crucial to the construction of theory in science and one that is especially important for evaluating theoretical claims in psychology.

Although there are different forms of operationism, it is most useful for the consumer of scientific information to think of it in the most general way. *Operationism* is simply the idea that concepts in scientific theories must in some way be grounded in, or linked to, observable events that can be measured. Linking the concept to an observable event is the operational definition of the concept and makes the concept public. The operational definition removes the concept from the feelings and intuitions of a particular individual and allows it to be tested by anyone who can carry out the measurable operations.

For example, defining the concept *hunger* as "that gnawing feeling I get in my stomach" is not an operational definition because it is related to the personal experience of a "gnawing feeling" and, thus, is not accessible to other observers. In contrast, definitions that involve some measurable period of food deprivation or some physiological index such as blood sugar levels are operational because they involve observable measurements that anyone can carry out. Similarly, psychologists cannot be content with a definition of *anxiety*, for example, as "that uncomfortable, tense feeling I get at times" but must define the concept by a number of operations such as questionnaires and physiological measurements. The former definition is tied to a personal interpretation of bodily states and is not replicable by others. The latter puts the concept in the public realm of science.

It is important to realize that a concept in science is defined by a *set* of operations, not by just a single behavioral event or task. Instead, several slightly different tasks and behavioral events are used to converge on a concept (we will talk more about the idea of converging operations in Chapter 8). For example, educational psychologists define a concept such as *reading ability* in terms of performance on a standardized instrument such as the

Woodcock Reading Mastery Tests (Woodcock, 1998). The total reading ability score on the Woodcock Reading Mastery instrument comprises indicators of performance on a number of different subtests that test slightly different skills but are all related to reading; for example, reading a passage and thinking of an appropriate word to fill in a blank in the passage, coming up with a synonym for a word, pronouncing a difficult word correctly in isolation, and several others. Collectively, performance on all of these tasks defines the concept *reading ability.*

Operational definitions force us to think carefully and empirically—in terms of observations in the real world—about how we want to define a concept. Imagine trying to define operationally something as seemingly conceptually simple as typing ability. Imagine you need to do this because you want to compare two different methods of teaching typing. Think of all the decisions you would have to make. You would want to measure typing speed, of course. But over how long a passage? A passage of only 100 words would seem too short, and a passage of 10,000 words would seem to long. But exactly how long then? How long does speed have to be sustained to match how we best conceive of the theoretical construct *typing ability?* And what kind of material has to be typed? Should it include numbers and formulas and odd spacing? And how are we going to deal with errors? It seems that both time and errors should come into play when measuring typing ability, but exactly what should the formula be that brings the two together? Do we want time and errors to be equally weighted, or is one somewhat more important than the other? The need for an operational definition would force you to think carefully about all of these things; it would make you think very thoroughly about how you conceptualize typing ability.

Reliability and Validity

For an operational definition of a concept to be useful, it must display both reliability and validity. *Reliability* refers to the consistency of a measuring instrument—whether you would arrive at the same measurement if you assessed the same concept multiple times. The scientific concept of reliability is easy to understand because it is very similar to its layperson's definition and very like one of its dictionary definitions: "an attribute of any system that consistently produces the same results."

Consider how a layperson might talk about whether something was reliable or not. Imagine a New Jersey commuter catching the bus to work in Manhattan each morning. The bus is scheduled to arrive at the commuter's stop at 7:20 A.M. One week the bus arrives at 7:20, 7:21, 7:20, 7:19, and 7:20, respectively. We would say that the bus was pretty reliable that week. If the next week the bus arrived at 7:35, 7:10, 7:45, 7:55, and 7:05, respectively, we would say that the bus was very unreliable that week.

The reliability of an operational definition in science is assessed in much the same way. If the measure of a concept yields similar numbers for multiple measurements of the same concept, we say that the measuring device displays high reliability. If we measured the same person's intelligence with different forms of an IQ test on Monday, Wednesday, and Friday of the same week and got scores of 110, 109, and, 110, we would say that that particular IQ test seems to be very reliable. In contrast, if the three scores were 89, 130, and 105, we would say that that particular IQ test does not seem to display high reliability. There are specific statistical techniques for assessing the reliability of different types of measuring instruments, and these are discussed in all standard introductory methodology textbooks.

But remember that reliability is only about consistency and nothing else. Reliability alone is not enough for an operational definition to be adequate. Reliability is necessary but not sufficient. To be a good operational definition of a concept, the operations assessed must also be a *valid* measure of that concept. The term *construct validity* refers to whether a measuring instrument (operational definition) is measuring what it is supposed to be measuring. In his methodology textbook, professor Paul Cozby (2006) gives us a humorous example of reliability without validity. Imagine you are about to get your intelligence assessed. The examiner tells you to stick out your foot and clamps on a measuring device like those at the shoestore and reads out a number. You would, of course, think that this was a joke. But note that this measuring instrument would display many of the types of reliability that are discussed in methodology textbooks. It would give virtually the same readings on Monday, Wednesday, and Friday (what is termed *test-retest reliability*) and it would give the same reading no matter who used it (what is termed *interrater reliability*).

The problem with the shoe device as a measure of intelligence is not reliability (which it has) but validity. It is not a good measure of the concept it purports to measure (intelligence). One way we would know that it is not a valid measure of intelligence is that we would find that it does not relate to many other variables that we would expect a measure of intelligence to relate to. Measures from the shoe instrument would not relate to academic success; they would not relate to neurophysiological measures of brain functioning; they would not relate to job success; and they would not relate to measures of the efficiency of information processing developed by cognitive psychologists. In contrast, actual measures of intelligence relate to all of these things (Deary, 2000; Geary, 2005; Lubinski, 2004). Actual measures of intelligence in psychology have validity as well as reliability, whereas the shoesize measure of intelligence has reliability without validity.

You might be wondering about another combination of affairs at this point, so let me recapitulate where we are. In operational definitions, we are looking for both reliability and validity, so high reliability and high validity are sought. We have just discussed the shoe-size IQ test in order to demonstrate that high reliability and low validity get us nowhere. A third case, low reliability and low validity, is so obviously useless that it is not worth discussing. But you might be wondering about the fourth and last possible combination: What if something has high validity and low reliability? The answer is that, like its converse case of low validity and high reliability (the shoe-size example), this state of affairs gets you nowhere. And, actually, it is more accurate to say that this state of affairs is impossible—because you cannot claim to be measuring validly if you cannot measure reliably.

Direct and Indirect Operational Definitions

The link between concepts and observable operations varies greatly in its degree of directness or indirectness. Few scientific concepts are defined almost entirely by observable operations in the real world. Most concepts are defined more indirectly. For example, the use of some concepts is determined by both a set of operations and the particular concept's relationship to other theoretical constructs. Finally, there are concepts that are not directly defined by observable operations but linked to other concepts that are. These are sometimes called latent constructs, and they are common in psychology.

For example, much research has been done on the so-called type A behavior pattern because it has been linked to the incidence of coronary heart disease (Austin & Deary, 2002; Curtis & O'Keefe, 2002; Matthews, 2005; Smith, 2003; Suls & Bunde, 2005). We will discuss the type A behavior pattern in more detail in Chapter 8. The important point to illustrate here, however, is that the type A behavior pattern is actually defined by a set of subordinate concepts: a strong desire to compete, a potential for hostility, time-urgent behavior, an intense drive to accomplish goals, and several others. However, each one of these defining features of the type A behavior pattern (a strong desire to compete, etc.) is *itself* a concept in need of operational definition. Indeed, considerable effort has been expended in operationally defining each one. The important point for our present discussion is that the concept of the type A behavior pattern is a complex concept that is not directly defined by operations. Instead, it is linked with other concepts, which, in turn, have operational definitions. The type A behavior pattern provides an example of a concept with an indirect operational definition. Although theoretical concepts differ in how closely they are linked to observations, all concepts acquire their meaning partially through their link to such observations.

Scientific Concepts Evolve

It is important to realize that the definition of a scientific concept is not fixed but constantly changing as the observations that apply to the concept are enriched. If the original operational definition of a concept turns out to be theoretically unfruitful, it will be abandoned in favor of an alternative set of defining operations. Thus, concepts in science are continually evolving and can increase in abstractness as the knowledge concerning them increases. For example, at one time the electron was thought of as a tiny ball of negative charge circling the nucleus of an atom. Now it is viewed as a probability density function having wavelike properties in certain experimental situations.

In psychology, the development of the concept of intelligence provides a similar example. At first, the concept had only a strict operational definition: Intelligence is what is measured by tests of mental functioning. As empirical evidence accumulated relating intelligence to scholastic achievement, learning, brain injury, neurophysiology, and other behavioral and biological variables, the concept was both enriched and refined (Deary, 2000, 2001; Geary, 2005; Lubinski, 2004; Sternberg, 2000; Sternberg & Grigorenko, 2002; Sternberg & Kaufman, 1998; Unsworth & Engle, 2005). It now appears that intelligence is best conceptualized as a higher-order construct defined by several more specific information-processing operations. These hypothesized processes, in turn, have more direct operational definitions stated in terms of measurable performance.

The concepts in theories of human memory have likewise evolved. Psychologists now rarely use global concepts like *remembering* or *forgetting;* instead, they test the properties of more specifically defined memory subprocesses, such as short-term acoustic memory, iconic storage, semantic memory, and episodic memory. The older concepts of remembering and forgetting have been elaborated with more specifically operationalized concepts.

Thus, the usage of theoretical terms evolves from scientific activity rather than from debates about the meaning of words. This is one of the most salient differences between the operational attitude of science and the essentialist quest for absolute definition. Neurologist Norman Geschwind (1985) characterized this difference as follows: "I think that one of the things you learn in the history of medicine is that many people think that the way to study a problem is to define the problem and then study it. That turns out again and again to be wrong because you discover the only way to define the problem properly is to know the answer" (p. 15).

Philosopher Paul Churchland (1988) emphasized the idea that concepts in science derive meaning not from language definitions but from observations and other concepts to which they are related:

> To fully understand the expression "electric field" is to be familiar with the network of theoretical principles in which that expression appears. Collectively, they tell us what an electric field is and what it does. This case is typical. Theoretical terms do not, in general, get their meanings from single, explicit definitions stating conditions necessary and sufficient for their application. They are implicitly defined by the network of principles that embed them. (p. 56)

As scientific concepts evolve, they often become enmeshed in several different theoretical systems and acquire alternative operational definitions. There is not necessarily anything wrong with the concept when this happens. For example, many believe that psychology is discredited by the fact that many of its important theoretical constructs, such as intelligence, are operationalized and conceptualized in more than one way (Sternberg, 2000). But such a situation is not unique to psychology, and it is not a matter for despair or hand-wringing. In fact, it is a relatively common occurrence in science. Heat, for example, is conceptualized in terms of thermodynamic theory and in terms of kinetic theory. Physics is not scandalized by this state of affairs. Consider the electron. Many of its properties are explained by its being conceptualized as a wave. Other properties, however, are better handled if it is viewed as a particle. The existence of these alternative conceptualizations has tempted no one to suggest that physics be abandoned.

Operational Definitions in Psychology

Many people understand the necessity of operationism when they think about physics or chemistry. They understand that if scientists are going to talk about a particular type of chemical reaction, or about energy, or about magnetism, they must have a way of measuring these things. Unfortunately, when people think and talk about psychology, they often fail to recognize the need for operationism. Why is it not equally obvious that psychological terms must be operationally defined, either directly or indirectly, in order to be useful explanatory constructs in scientific theories?

One reason is what has been termed the *preexisting-bias problem* in psychology. We alluded to this problem in Chapter 1. People do not come to the study of geology with emotionally held beliefs about the nature of rocks. The situation in psychology is very different. We all have intuitive theories of personality and human behavior because we have been "explaining" behavior to ourselves all our lives. All our personal psychological theories contain theoretical concepts (for example, *smart, aggressive, anxiety*). Thus, it is only natural to ask why we have to accept some other definition. Although this attitude seems reasonable on the surface, it is a complete bar to any scientific progress in understanding human behavior and is the cause of much public confusion about psychology.

One of the greatest sources of misunderstanding and one of the biggest impediments to the accurate presentation of psychological findings in the media is the fact that many technical concepts in psychology are designated by words used in everyday language. This everyday usage opens the door to a wide range of misconceptions. The layperson seldom realizes that when psychologists use words such as *intelligence, anxiety, aggression,* and *attachment* as theoretical constructs, they do not necessarily mean the same thing that the general public does.

The nature of this difference should be apparent from the previous discussion of operationism. When terms such as *intelligence* and *anxiety* are used in psychological theories, their direct or indirect operational definitions determine their correct usage. These definitions are often highly technical, usually fairly specific, and often different from popular usage in many ways. For example, when hearing the phrase "the first principal component of the factor analysis of a large sampling of cognitive tasks," many people will not recognize it as part of the operational definition of the term *intelligence.*

Similarly, in lay usage, the term *depression* has come to mean something like "feeling down in the dumps." In contrast, the technical definition of major depressive disorder takes up over a dozen pages in the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994) and means something quite different from being "down in the dumps." A clinical psychologist's depression is not the same as the layperson's depression (Hollon, Thase, & Markowitz, 2002). Other sciences also have this problem, although perhaps in a less severe form than psychology. Recall the previous discussion of the concept *life.* As Medawar and Medawar (1983) pointed out, "The trouble is that 'life,' like many other technical terms in science, has been pirated from the vernacular and is used in scientific contexts far removed from those that might arise in common speech" (p. 66).

Physicist Lisa Randall (2005) discusses how this problem obscures the understanding of physics by the public. She points out that the term *relativity* in Einstein's theory has been taken by the public to imply that "there are no absolutes because everything is relative" when in fact the theory says just the opposite! Randall points out that actually Einstein's theory implies that "although the measurements any observer makes depend on his coordinates and reference frame, the physical phenomena he measures have an invariant description that transcends that observer's particular coordinates. Einstein's theory of relativity is really about finding an invariant description of physical phenomena. Indeed, Einstein agreed with the suggestion that his theory would have been better named 'invariantentheorie.' But the term 'relativity' was already too entrenched at the time for him to change" (p. 13).

Randall goes on to point out that even in physics "ambiguous word choices are the source of some misunderstandings. Scientists often employ colloquial terminology, which they then assign a specific meaning that is impossible to fathom without proper training" (p. 13). And the same is true in psychology. When the psychologist and the layperson use the same word to mean different things, they often misinterpret each other. Such confusion would be less prevalent if new words had been coined to represent psychological constructs. On occasion such words have been coined. Just as physicists have their *erg* and *joule*, psychology has its *dissonance* and *encoding*, words that are not actually coined but are uncommon enough to prevent confusion.

"But," the layperson may object, "why do psychologists inflict this on us? New jargon, highly technical definitions, uncommon uses of words. Why do we need them? Why is my idea of 'intelligence' not an acceptable idea to talk about?"

Here we see exemplified a critical misunderstanding of psychological research—a misunderstanding that is often reflected in media reports of psychological research. A national newspaper report on the 1996 meeting of the American Psychological Association (Immen, 1996) is headlined "Could You Repeat That in Klingon?" and refers to "psychologists speaking a language all their own." The article ridicules the following title of a paper delivered at the conference: "Interpreting WJ-R and KAIT Joint Factor Analyses from Gf-Gc Theory." Although the reporter states that he would "not even dare to speculate about the true meaning" of the title, almost all properly trained psychologists would recognize the title as referring to developments in intelligence test theory. And this is as it should be. Gf-Gc theory is a technical development in intelligence theory. There is no reason for the reporter to have heard of this concept—just as one would not expect the reporter to know the details of the latest elementary particle to be identified by physicists. Somehow, however, the reporter's (quite understandable) ignorance of the technical terminology is seen as reflecting negatively on modern psychology.

We come here to the crux of the problem. The first step in resolving it is to emphasize a point from our earlier discussion: Operationism is not unique to psychology. It is characteristic of all sciences. Most of the time, we accept it readily, recognizing its obvious nature. If a scientist is investigating radioactivity, we take it for granted that he or she must have some observable way of measuring the phenomenon—a method that another investigator could use to obtain the same results. This method is what makes possible the public nature of science, one of its defining features. Two different scientists agree on the same operational definition so that it is possible for one to replicate the other's results. However, what seems obvious in other contexts is sometimes not so clear when we think about psychology. The necessity for operational definitions of concepts like *intelligence* and *anxiety* is often not recognized because we use these terms all the time, and, after all, don't we all just "know" what these things mean?

The answer is "No, we don't"—not in the sense that a scientist has to know, that is, in a public sense. A scientist must "know" what intelligence means by being able to define, precisely, how another laboratory could measure it in exactly the same way and be led to the same conclusions about the concept. This is vastly different—in terms of explicitness and precision—than the vague verbal connotations that are needed in order to achieve casual understanding in general conversation.

Operationism as a Humanizing Force

The problem with relying on what we all just "know" is the same problem that plagues all intuitive (that is, nonempirical) systems of belief. What you "know" about something may not be quite the same as what Jim "knows" or what Jane "knows." How do we decide who is right? You may say, "Well, I feel strongly about this, so strongly that I *know* I'm right." But what if Jim, who thinks somewhat differently, feels even more strongly than you do? And then there's Jane, who thinks differently from you or Jim, claiming that she must be right because she feels *even more* strongly than Jim does.

This simple parody is meant only to illustrate a fundamental aspect of scientific knowledge, one that has been a major humanizing force in human history: In science, the truth of a knowledge claim is not determined by the strength of belief of the individual putting forth the claim. The problem with all intuitively based systems of belief is that they have no mechanism for deciding among conflicting claims. When everyone knows intuitively, but the intuitive claims conflict, how do we decide who is right? Sadly, history shows that the result of such conflicts is usually a power struggle.

Some people mistakenly claim that an operational approach to psychology dehumanizes people and that instead we should base our views of human beings on intuition. Psychologist Donald Broadbent (1973) argued that the truly humane position is one that bases theoretical views of human beings on observable behavior rather than on the intuition of the theorizer:

> We can tell nothing of other people except by seeing what they do or say in particular circumstances. . . . The empirical method is a way of reconciling differences. If one rejects it, the only way of dealing with a disagreement is by emotional polemic. (p. 206)

Thus, the humanizing force in science is that of making knowledge claims public so that conflicting ideas can be tested in a way that is acceptable to all disputants. Recall the concept of replication from Chapter 1. This allows a selection among theories to take place by peaceful mechanisms that we all agree on in advance. The public nature of science rests critically on the idea of operationism. By operationally defining concepts, we put them in the public realm, where they can be criticized, tested, improved, or perhaps rejected.

Psychological concepts cannot rest on someone's personal definition, which may be uncommon, idiosyncratic, or vague. For this reason, psychology must reject all personal definitions of concepts (just as physics, for example, rejects personal definitions of energy and meteorology rejects personal definitions of what a cloud is) and must insist on publicly accessible concepts defined by operations that anyone with proper training and facilities can perform. In rejecting personal definitions, psychology is not shutting out the layperson but is opening up the field—as all sciences do—to the quest for a common, publicly accessible knowledge that all can share.

Such publicly accessible knowledge is available to solve human problems only when concepts have become grounded in operational definitions and are not the focus of essentialist arguments about the meaning of words. For example, Monk (1990) describes how during World War II the concept of *wound shock* had become problematic in medicine. Some physicians identified the condition based on an abnormally high concentration of red blood cells thought to be due to a leakage of plasma from the blood into tissue. Others identified wound shock on the basis of low blood pressure, skin pallor, and rapid pulse. In other words, operational definitions of the concept were inconsistent (and even idiosyncratic) and, thus, one physician by the name of Grant working for the British Medical Research Council recommended "that the very concept of 'wound shock' should be abandoned and that detailed observations of casualties should be made without using the term. . . . The lack of a common basis of diagnosis renders it impossible to assess the efficacy of the various methods of treatment adopted" (Monk, 1990, pp. 445–446). In other words, the concept was doing more harm than good because it did not have a definition that was common enough so that it could be considered public knowledge (i.e., generally shared and agreed upon).

Sometimes the changes in the meaning of concepts in science will put scientific understanding of a concept in conflict with the nonspecialist's understanding. Farber and Churchland (1995) discuss such a situation surrounding the concept of fire. The classical concept was used "to classify not only burning carbon-stuffs, but also activity on the sun and various stars (actually nuclear fusion), lightning (actually electrically induced incandescence), the northern lights (actually spectral emission), and the flash of fireflies (actually phosphorescence). In our modern conceptual scheme, since none of these things involves oxidation, none belongs to the same class as wood fires. Moreover, some processes that turned out to belong to the oxidation class—rusting, tarnishing, and metabolism—were not originally considered to share anything with burning, since felt heat was taken to be an essential feature of this class" (p. 1296). In short, the principle of oxidation uniting the phenomena of a campfire and rusting—and separating them from the phenomenon of lightning—may be a sign of progress to a scientist, but it can be confusing and disorienting to the layperson.

Essentialist Questions and the Misunderstanding of Psychology

Another reason many people seem to abandon the idea of operationism when they approach psychology is that they seek essentialist answers to certain human problems. Whether the cause is psychology's relatively recent separation from philosophy or the public's more limited understanding of psychology than of other sciences is unclear. In a sense, however, it does not matter. The net result is the same. Psychology is expected to provide absolute answers to complex questions in a way that other sciences are not.

Recall the questions at the beginning of this chapter: What is the real meaning of the word *gravity?* What is the underlying essence of it? What does it ultimately mean even to speak of gravity? Most people would recognize that these questions require knowledge of the ultimate, underlying nature of a phenomenon and that current theories in physics cannot provide answers to questions of this type. Anyone familiar with popular writing about the progress of physical science in the last few centuries will recognize that gravity is a theoretical construct of great complexity and that its conceptual and operational relationships have been in constant flux.

However, substitute the word *intelligence* for the word *gravity* in each of the preceding questions and, suddenly, a miracle occurs. Now the questions are imbued with great meaning. They seem natural and meaningful. They literally beg for an ultimate answer. When the psychologist gives the same answer as the physicist—that intelligence is a complex concept that derives meaning from the operations used to measure it and from its theoretical relationships to other constructs—he or she is belittled and accused of avoiding the real issues.

One problem facing psychology, then, is that the public demands answers to essentialist questions that it does not routinely demand of other sciences. These demands often underlie many of the attempts to disparage the progress that has been made in the field. Although these demands do not hinder the field itself—because psychologists, like other scientists, ignore demands for essentialist answers and simply go about their work—they are an obstacle to the public's understanding of psychology. The public becomes confused when an uninformed critic claims that there has been no progress in psychology. The fact that this claim so frequently goes unchallenged reflects the unfortunate truth of the major premise of this book: Public knowledge of what scientific achievement within psychology would actually mean is distressingly meager. When examined closely, such criticisms usually boil down to the contention that psychology has not yet provided the ultimate answer to any of its questions. To this charge, psychology readily pleads guilty—as do all the other sciences.

Some may find it discomforting to learn that no science, including psychology, can give answers to essentialist questions. Holton and Roller (1958) discussed the uneasiness that the layperson may feel when told that physicists cannot answer essentialist questions. They discuss the phenomenon of radioactive decay in which the number of atoms of a radioactive element that have decayed can be related to time via an exponential mathematical function. The function, however, does not explain why radioactive decay occurs. The solution to this problem will again probably involve a mathematical function, but it again will not answer the layperson's question of what radioactive decay really is. Holton and Roller tell us that "we must try to make our peace with the limitations of modern science; it does not claim to find out 'what things really are'" (pp. 219–220). As science writer Robert Wright (1988) explained,

> There was something bothersome about Isaac Newton's theory of gravitation. . . . How, after all, could "action at a distance" be realized? . . . Newton sidestepped such questions. . . . Ever since Newton, physics has followed his example. . . . Physicists make no attempt to explain why things obey laws of electromagnetism or of gravitation. (p. 61)

Likewise, those who seek essentialist answers to questions concerning human nature are destined to be disappointed if they are looking to psychology. Psychology is not a religion. It is a broad field that seeks a scientific understanding of all aspects of behavior. Therefore, psychology's current explanations are temporary theoretical constructs that account for behavior better than alternative explanations. These constructs will certainly be superseded in the future by superior theoretical conceptualizations that are closer to the truth.

Operationism and the Phrasing of Psychological Questions

The idea of an operational definition can be a very useful tool in evaluating the falsifiability of a psychological theory. The presence of concepts that are not directly or indirectly grounded in observable operations is an important clue to recognizing a nonfalsifiable theory. These concepts are usually intended to rescue such a theory from disconfirmation after the data have been collected. Thus, the presence of loose concepts—those for which the theorist cannot provide direct or indirect operational links—should be viewed with suspicion.

A principle that scientists term *parsimony* is relevant here. The principle of parsimony dictates that when two theories have the same explanatory power, the simpler theory (the one involving fewer concepts and conceptual relationships) is preferred. The reason is that the theory with fewer conceptual relationships will likely be the more falsifiable of the two in future tests.

A strong grasp of the principle of operationism will also aid in the recognition of problems or questions that are scientifically meaningless. For example, I have in my files a wire service article, from United Press International, entitled "Do Animals Think?" The article describes recent experimentation in animal behavior. There is nothing wrong with the research described in the article, but it is clear that the title is merely a teaser. The question in the title is scientifically meaningless unless some operational criteria are specified

for the term *think*, and none is given in the article. A similar problem concerns the many newspaper articles that have asked, "Can computers think?" Without some operational criteria, this question is also scientifically meaningless, even though it is infinitely useful as grist for cocktail party conversation.

Actually it is instructive to observe people debating this last question because such a debate provides an opportunity to witness concretely the preexisting-bias problem in psychology that we discussed earlier. Most people are strongly biased toward not wanting a computer to be able to think. Why? For a variety of reasons, the layperson's concept *think* has become so intertwined with the concept *human* that many people have an emotional reaction against the idea of nonhuman things thinking (for example, computers or extraterrestrial life forms that look nothing like the humans on our planet).

However, despite their strong feelings against the idea of thinking computers, most people have not thought about the issue very carefully and are at a loss to come up with a definition of thinking that would include most humans (babies, for example) and exclude all computers. It is sometimes humorous to hear the criteria that people who are unfamiliar with current work in artificial intelligence come up with, for they invariably choose something that computers can do. For example, many people propose the criterion "ability to learn from experience," only to be told that some robots and artificial intelligence systems have fulfilled this criterion (Churchland, 1995; Clark, 2001; McCorduck, 2004; Pfeifer & Scheier, 1999). The strength of preexisting bias can be observed in this situation. Is the person's response "Oh, I didn't know. Well, since the criterion for thinking that I put forth is met by some computers, I will have to conclude that at least those computers think"? Usually this intellectually honest response is not the one that is given. More commonly, the person begins groping around for another criterion in the hope that computers cannot meet it.

Usually the second choice is something like "creativity" ("coming up with something that people judge as useful that no person has thought of before"—we will ignore the question of whether most *humans* would meet this criterion). When told that most experts agree that computers have fulfilled this criterion (Boden, 2003; Pfeifer & Scheier, 1999), the person still does not admit the possibility of thinking machines. Often the person abandons the attempt to derive an operational definition at this point and instead attempts to argue that computers could not possibly think because "humans built them and programmed them; they only follow their programs."

Although this argument is one of the oldest objections to thinking machines (McCorduck, 2004; Robinson, 1992; Woolley, 2000), it is actually fallacious. Preexisting bias prevents many people from recognizing that it is totally irrelevant to the question at issue. Almost everyone would agree that thinking is a process taking place in the natural world. Now notice that we do not invoke the "origins" argument for other processes. Consider the process of heating food. Consider the question, "Do ovens heat?" Do we say, "Ovens don't really heat, because ovens are built by *people*. Therefore, it only makes sense to say that *people* heat. Ovens don't really heat"? Or what about lifting? Do cranes lift? Is our answer "Cranes don't really lift because cranes are built by *people*. Therefore, it only makes sense to say that *people* lift. Cranes don't really lift"? Of course not. The origin of something is totally irrelevant to its ability to carry out a particular process. The process of thinking is just the same. Whether or not an entity thinks is independent of the origins of the entity.

The failure to think rationally about the possibility of thinking machines was one reason that the noted computer scientist Alan Turing developed his famous test of whether computers think. What is important to our discussion is that the test Turing devised is an *operational* test. Turing began his famous article "Computing Machinery and Intelligence" (1950) by writing, "I propose to consider the question 'Can machines think?'" Not wanting discussion of the issue to descend to the usual circular cocktail-party chatter or endless essentialist arguments about what we mean by *think*, Turing proposed a strict operational test of whether a computer could think. His proposal was that it would be reasonable to grant a computer thinking powers if it could carry on an intelligent conversation.

The creativity in the Turing proposal was that he put forth a way to operationalize the question while at the same time guarding against the preexisting-bias problem. Turing strictly specified the logistics of the test of whether the computer could carry on an intelligent conversation. It was not to be done by having a tester interact with the computer via keyboard and screen and then having the tester judge whether the computer had, in fact, carried on an intelligent conversation. Turing did not propose this type of test because he was concerned about the preexisting-bias problem. Turing was sure that, once the person sat down before a computer, keyboard, and screen—something obviously a *machine*—the person would deny it thinking capabilities no matter what it did. Therefore, Turing proposed a test that controlled for the irrelevant external characteristics of the thinking device. His well-known proposal was to have the tester engage in conversation via two keyboards—one connected to a computer and the other to a human, both out of sight—and then to decide which was which. If the tester could not identify the human with greater than chance accuracy, then one reasonable inference was that the conversational abilities—the operational definition of thinking—of the computer were equal to those of a human.

Turing's key insight was the "same insight that inspires the practice among symphony orchestras of conducting auditions with an opaque screen between the jury and the musician. What matters in a musician, obviously, is musical ability and only musical ability: such features as sex, hair length, skin color, and weight are strictly irrelevant. . . . Turing recognized that people might be similarly biased in their judgments of intelligence by whether the contestant had soft skin, warm blood, facial features, hands and eyes—which are obviously not themselves essential components of intelligence" (Dennett, 1998, p. 5). Turing's test teaches us the necessity of operational definitions if we are to discuss psychological concepts rationally; that is, in a principled way rather than merely as a reflection of our own biases about the question at issue.

The intellectual style revealed when we observe people discussing the issue of artificial intelligence illustrates well the difference between scientific and nonscientific styles of thinking. The scientific approach is to develop an operational definition that seems reasonable and then to see what conclusions about thinking, computers, and humans it leads to. In contrast, preexisting bias dominates the thinking of most people. They have already arrived at certain conclusions and are not interested in what is actually known about the relative contrasts between computer and human performance. Instead, with minds made up, they spend their intellectual energies in a desperate juggling of words designed to protect their prior beliefs from change. What we see, then, is a combination of preexisting bias and nonoperational essentialist attitudes that fuel the assumption that people "just know" what thinking "really" is without any necessity of operational criteria. Such attitudes are what make most people's intuitive psychological theories unfalsifiable and, hence, useless. These very attitudes illustrate precisely why we need the *science* of psychology!

Summary

Operational definitions are definitions of concepts stated in terms of observable operations that can be measured. One of the main ways that we ensure that theories are falsifiable is by making certain that the key concepts in theories have operational definitions stated in terms of well-replicated behavioral observations. Operational definitions are one major mechanism that makes scientific knowledge publicly verifiable. Such definitions are in the public domain so that the theoretical concepts that they define are testable by all—unlike "intuitive," nonempirical definitions that are the special possession of particular individuals and not open to testing by everyone.

Because psychology employs terms from common discourse, such as *intelligence* and *anxiety*, and because many people have preexisting notions about what these terms mean, the necessity of operationally defining these terms is often not recognized. Psychology is like all other sciences in requiring operational definitions of its terms. However, people often demand answers to essentialist questions (questions about the absolute, underlying nature of a concept) of psychology that they do not demand of other sciences. No science provides such answers to ultimate questions. Instead, psychology, like other sciences, seeks continually to refine its operational definitions so that the concepts in theories more accurately reflect the way the world actually is.

4

Testimonials and Case Study Evidence

Placebo Effects and the Amazing Randi

Cut to the *Oprah Winfrey Show,* one of the most popular television talk shows of the last decade. Today's guest is Dr. Alfred Pontificate, director of the Oedipus Institute of Human Potential. Oprah attempts to elicit 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