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

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

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

Reliability and Validity

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

Consider how a layperson might talk about whether something was eliable 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 top at 7:20 A.M. One week the bus arrives at 7:20, 7:21, 7:20, 7:19, and 7:20, espectively. 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 vould 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 ave their *erg* and *joule*, psychology has its *dissonance* and *encoding*, words that re not actually coined but are uncommon enough to prevent confusion.

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

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

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

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