

Chapter 3

Personality Research

Chapter Goals

- Familiarize the reader with key terms used in the field of psychological research
- Provide a comprehensive understanding of the reasons for the research and testing of the various theories of personality
- Provide in-depth knowledge of the several testing approaches used in research as well as their advantages and disadvantages

Chapter 2 discussed the progress of personality theory from the general intuitions of great thinkers and innovators to a more formal science. A closer examination of the history of personality theory should make clear why a scientific approach is essential. Before the emergence of psychology as a discipline, the characteristics that define people as individuals were sometimes thought to result from the activity of spirits and demons. So, too, they were sometimes explained as the results of astronomical configurations. The apparent placement of the stars within constellations and the movements of the moon and the planets within the solar system were (and by some still are) believed to ordain the personality of those born under them.

Many of these explanations were quite understandable during eras in which there were few established means of testing their truth. Thus, if something seemed credible or could be convincingly argued, it might very well be accepted as a truth. After all, why do many people today study their astrological signs and explain their personality in terms of their sun sign in the zodiac? Perhaps this is so because such explanations are simple and easily understood and because the benign character descriptions for each zodiacal sign are universally appealing. The seductive attraction of other theories has led to large followings for phrenology, **spiritualism**, **animism**, and numerous similar movements that have failed to earn scientific support. It is only through unbiased and systematic examination of received ideas and rejection of those not supported by objective evidence that we can come ever closer to an understanding of what is true. The study of personality, as in all other subfields within psychology, is now conducted and evaluated by the same methods used in all other sciences.

54 PERSONALITY THEORIES

Image 3.1



THE SCIENTIFIC METHOD

What is science? Most people, if asked that question, would reply that science is an advanced body of knowledge in a so-called hard or technical area like chemistry or physics. But one of the most important things for college students to understand is that science is not so much a collection of information as a method that leads to knowledge about reality. A science can be defined as any discipline in which the scientific method is used.

During the last few centuries, a number of individuals sought methods of drawing conclusions objectively and in ways that can be communicated to other scholars. A major figure in this movement was Francis Bacon (1561–1626), an English nobleman who began his career as a lawyer and member of Parliament during the reign of Queen Elizabeth I. Bacon advocated an experimental approach to the acquisition of knowledge and encouraged using qualitative and inductive methods. In his *Novum Organum* (1620/1994), Bacon proposed that scientific knowledge is best advanced

by an inductive rather than a deductive approach. According to Bacon, nothing should be accepted *a priori* but must be tested and examined. He suggested that the essential principles of any natural phenomenon can be understood if a sufficient number of observations are made to allow the induction of the underlying principles.

Contemporary science still uses the inductive method or **induction**. Let us clearly define what that method is and how it differs from **deduction**. Sherlock Holmes, the famous fictional detective created by Sir Arthur Conan Doyle, is typically described by his creator as using deduction. Actually, Holmes, or any wise detective for that matter, uses induction: building the case against a suspect by accumulating small facts to support it. In a short story called “The Five Orange Pips,” which was published in 1891, Conan Doyle has Holmes compare himself to a famous French zoologist who reconstructed extinct animals from fossilized bones: “As [Georges] Cuvier could correctly describe a whole animal by the contemplation of a single bone, so the observer who has thoroughly understood one link in a series of incidents should be able to state all the other ones, before and after.” (Doyle, 2002, p. 81)

A scientist is basically a detective. In contrast to deductive reasoning, which moves from general premises to specific conclusions, induction begins from small actual observations, using an accumulation of these to formulate some generalizations—generalizations that are useful only if they lead to testable predictions. Just as Holmes used such observations

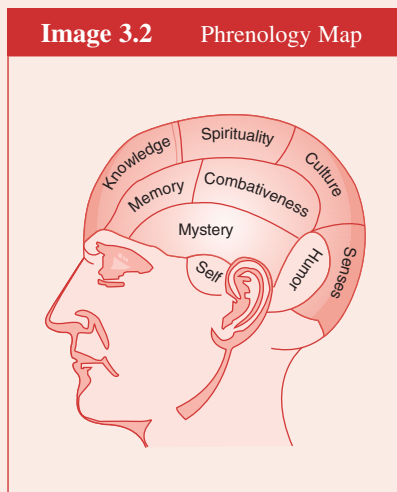
as the mud and scrape marks on his friend Watson's boots to infer that (a) Watson had been out in the rain that day and (b) had recently hired an incompetent servant who scratched the boot leather in attempting to remove the mud, a scientist might begin with some apparently unconnected phenomena and formulate a hypothesis to be tested by gathering further data.

Early scientists sought ways to study natural events without bias and to develop common sets of terms and systems of measurement for communicating their findings. One example was the development of the metric system, which began in 18th-century France when King Louis XVI charged a group of famous scientists to develop a universal system of measurement to replace the many different standards then in use in different parts of France. This process of standardization has led to a consensus among researchers in the many realms of science to examine events and understand their causes, effects, and even to predict their recurrence—as when astronomers learned to plot the courses of comets and predict the dates of their return.

Theories in personality psychology commonly offer an overarching explanation of the differences found among people, the degree to which they differ, and the sources of these differences. In seeking to answer these questions, researchers set forth propositions that can be tested against real-world events. Unless theories are testable and subjected to appropriate examination, the study of personality will not differ from the many pseudosciences that some people accept on faith. If you are taking the time to read a textbook, you most likely want something more than one individual's opinion, even if that person is reputed to have a first-rate intellect and deep insight. The following box contains an example of a hypothesis in personality theory, one that failed its tests.

Application of the Scientific Method to Personality Theory

In the mid-1790s, Franz Josef Gall (1758–1828) employed the consistently verified hypothesis that the brain is the seat of human personality. Gall combined this hypothesis with another proposal—that each personality characteristic is exclusively controlled by a specific and identifiable region of the brain. Consequently, he concluded that the skull overlying a particularly well-developed brain region would bulge in direct proportion to the strength of that characteristic. The graphic below illustrates this notion.



The drawing illustrates the theory of phrenology, which became widely employed as a clinical method throughout Europe in the late 18th and early 19th centuries. In the early 1800s, however, experimental data presented by Jean Pierre Flourens (1794–1867) (Klein, 1970, p. 668) showed that Gall's notion of localization of function was a gross oversimplification of brain functioning. Consequently, a theory with very good *face validity* fell into oblivion.

56 PERSONALITY THEORIES

Theoretical Research and Personality

Theories are essential for the growth of any field of study, provided that they are challenged, tested, and refined. Personality psychology is no exception to this rule. It is a field presently in flux, containing numerous theories—some new and others largely unchanged for a century. Irrespective of the pedigree of these theories, however, scientific personality psychology is based on philosophical **determinism**, which holds that all events are invariably caused by preceding events. This cause-and-effect relationship is governed by inviolable natural laws. The acceptance of this basic principle has profound consequences for the student of science. It means that anything one observes in the world can be traced to a preexisting cause that in turn can be understood on the basis of such invariant natural laws as those set forth in fields like mathematics and physics. Of course, human behavior is more complicated than the invariant world of chemicals and atoms. Predicting the behavior of people in a group, or even what one person will do in a given situation, is a far harder task than predicting what will happen when two chemicals are combined. It may, in any event, be a probabilistic matter, akin to tossing a die, but the essential concept is the same.

Determinism holds that events never occur as a result of spontaneous or magical causes. Moreover, determinists maintain that anything that does happen can ultimately be understood as a rule-governed event. For example, if a specific individual has an open and agreeable personality, we can look to preceding events or states to understand it. Such events may include the individual's genetic code, his or her parenting, or the behavior of peers. Only empirical study can be the basis of discovery of the causes of a person's behavior.

Once we accept the premise that human personality has identifiable causes and that these can be discovered through research, the next step is to use the procedures in Table 3.1 to develop an explanation of personality. Table 3.2 summarizes some of the more prominent theoretical approaches.

Reliability

If you are in college or graduate school, it is highly likely that you have taken some test of academic or intellectual ability. Such tests include the Scholastic Aptitude Test (SAT), the Graduate Record Examination (GRE), and the Law School Admission Test (LSAT). These tests are used to predict your performance in future educational endeavors. According to the publishers of these tests, the higher your score, the more likely you are to perform well in the school that you are competing to enter. If a test predicts students' later performance with a high degree of accuracy, it will be considered valid. But before we can even consider such a measure of legitimacy, we would need to know whether the test yields consistent results.

Let us suppose, for example, that you took the SAT and received a score of 1400. Suppose further that you felt that this result understated your academic aptitude and took the test again the next month, only to receive a score of 900 on the second try. We suspect that your reaction would be one of disappointment and disbelief. Wouldn't you wonder if there might be some quirk in the design of the test rather than a tremendous drop in your actual abilities?

Table 3.1 The Scientific Method

1. An event or object is observed that provokes interest and is thought to require further study or explanation.
2. A model is then formulated that either explains the observation or connects it to observations made by other scientists. This model is called a hypothesis, although made prior to a formal study; it is usually the product of a deep understanding in the applicable field of study.
3. The hypothesis is then employed to predict or create similar events in experimental settings or in observations in natural settings.
4. If the hypothesis turns out to be inaccurate in predicting experimental outcomes or worldly events, it is modified or replaced by a new hypothesis.
5. If the hypothesis proves consistently accurate, it is accepted and then becomes a theory, which is a set of connected propositions to explain a class of phenomena.

Table 3.2 Theoretical Approaches to Personality

<i>Theoretical Approach</i>	<i>Foundation</i>	<i>Typical Research Method</i>
Analytic	Personality is a function of dynamic and unconscious forces that control most human thought and behavior.	Idiographic case studies based on clinical encounters.
Trait	Personality comprises several discrete and enduring characteristics.	Statistical/correlational and psychometric research.
Social learning	Personality develops out of directly reinforced, observed, or socially encouraged behaviors.	Observation, quasi-experiments, and true experiments
Cognitive behavioral	Personality is a function of both thinking and behavior, reinforced through life events and interactions with others. Tendencies toward reinforcement can be biologically or genetically determined.	Observation, quasi-experiments, and true experiments.
Genetic/biological	Personality is driven by genetic endowments and the resulting physiological processes	Epidemiology studies, quasi- and true experiments; statistical analyses.

58 PERSONALITY THEORIES

If you were a bit research-oriented, you might check the published data on the SAT's reliability. You would likely find that it has a **reliability coefficient** of approximately .91. This number means that there is a .91 correlation between the scores people obtain on the first administration of the test and those obtained on a subsequent administration. We can conclude that although the test did not seem reliable to you, it is indeed a reliable instrument. On the other hand, if you had discovered that the test had a low reliability coefficient, that fact would indicate that the test is a poor measure of your ability.

Reliability in research or personality measurement refers to the dependability of the measure. A reliable test will yield very similar results in repeated administrations except in the face of basic changes that might have taken place between measurements. Logically, any test that yields unreliable results cannot be considered a valid measure of anything it claims to measure. For a test to serve as a valid measure of personality or any other attribute, however, assessing reliability is just the first step in evaluating the test.

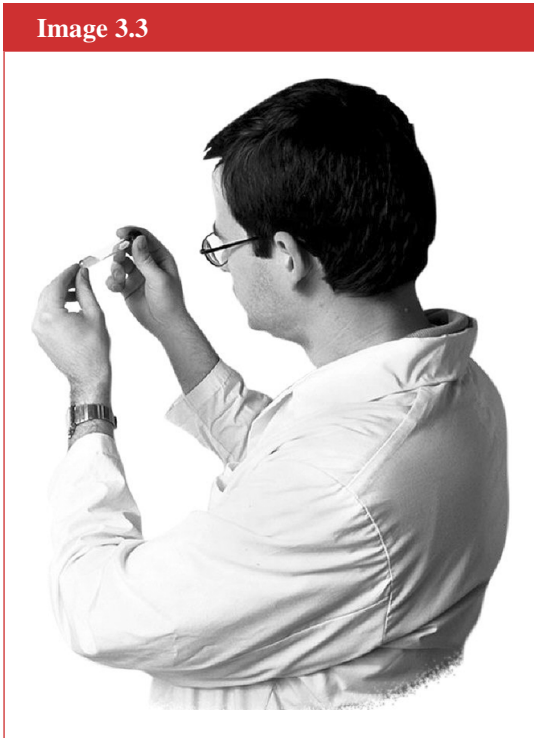
Validity

If we were to provide a self-administered test that measures the quality of your soul, for example, we would run into some significant problems, not the least of which would be that we would not be clear as to exactly what we are measuring. The test may very well give consistent results. That is, you might take the test several times and always come out with the result that you possess a soul of the highest caliber. But we would still be left with the vexing problem of defining the soul.

Depending on your religious background and convictions, you may or may not accept the existence of a soul. And even if you do, your conception of *soul* may be radically different from that of someone accepting a different set of religious teachings. For a test of a human soul to be valid, a researcher must have a clear and precise definition of it. In addition, the test must yield some outcome that can meaningfully rate, categorize, or measure a soul. Thus, even if we could construct an operational definition of *soul*, we would still have the problem of measuring it in a meaningful way.

Validity in personality research refers to the same conceptual issue. To perform a valid measurement of personality, we are required to define precisely what we are measuring. Are we measuring a component trait? Or are we measuring the entire person? Does our

Image 3.3



measure of personality make some assumption about the nature of personality that other researchers may simply reject? These and related questions must be answered before moving on to the next step in developing a method to measure the construct we are studying. Once this step is completed, we must then find the appropriate individuals to serve as test subjects for the method.

For example, suppose researchers wish to ascertain if overcoming feelings of inferiority is a key factor in the development of a mature personality. They then construct an experiment in which a random sample of teenagers is divided into two groups based on their scores on a test that measures feelings of inferiority. Those who express a low level of inferiority feelings are placed in one group, and those with a high level of such feelings are placed in the second group. The participants are contacted 5 years later and retested to produce three groups.

The first group consists of young adults who scored high on both the first and second administration of the examination. The second group consists of those who were shown to

VIGNETTE

Buddy, as he liked to be called, lived up to his name. He worked hard at being everyone's friend; he loved to be liked. The youngest of three brothers, Buddy grew up in an ethnic neighborhood in the suburb of a large city. His mother was an elementary school teacher and his father a postal worker. The family was close-knit; moreover, all of the boys were good students and accepted into college. In fact, both of Buddy's brothers became doctoral-level professionals.

Buddy himself earned two master's degrees; however, these degrees were in fields that offered few prospects for faculty teaching positions. After earning an M.A. in Asian history, Buddy referred to himself as a historian, but he took a job as a corporate messenger that belied that appellation. Despite all exhortations by his brothers to get a teaching credential or some other training that would enable him to enter a more remunerative profession, Buddy went back to school for an M.A. in philosophy.

Buddy was somewhat diminutive, and his boyish demeanor made him seem even smaller. In fact, he would often come across as a prematurely aged child, always joking and allowing others to make him the object of ridicule. On several occasions, he was warned by friends in his workplace that he was sacrificing all chances to move up the company ladder by playing the clown. One friend even told him that he was liked by all but respected by no one and that he would always be a messenger if he did not stop his self-effacing silliness. Despite Buddy's high intellect, he could not absorb and learn from these well-intended admonitions.

In addition to Buddy's joking around, and in some contrast to it, he spent a significant portion of his spare time doing others' work. Although Buddy's job at the

(Continued)

60 PERSONALITY THEORIES

(Continued)

company was considered low-level, he was bright enough to perform some higher level tasks for his superiors. Many of them exploited his ability and openly took credit for his work. Buddy never complained or objected; he was satisfied if they condescended to pat him on the head or offer some superficial compliment.

Outside the workplace, Buddy's self-defeating nature was most evident in his relationships with women. His typical relationship, which usually lasted only a few weeks, consisted of a series of expensive dates in which the woman would become increasingly critical of him. He would then respond to her criticisms with gifts or some other bit of submissive behavior. The woman would lose interest in him and leave him for another man.

Buddy seemed to be completely oblivious to the pattern of self-defeat that was evident to everyone else who knew him. He acted as though he sought out rejection and humiliation; this was not the case, however, as each additional rejection became increasingly painful to him. Unfortunately, with each blow to his self-image, he would become even more ingratiating and subservient to the people in his life. This behavior perpetuated Buddy's vicious circle of self-effacement: from contempt on the part of others to rejection to lowered self-image and back to self-effacement.

This pattern characterized most of Buddy's adult life. His self-defeating style exemplifies the way in which a personality can be both ineffective and opaque to the individual possessing it. Personality usually describes a pattern of behavior that is adaptive for an individual, but in some cases, the pattern can be quite destructive. In such cases, it is difficult for the afflicted individual to be aware of the interaction between his personality characteristics and his unhappy life, and the person will often fail to see the relationship when it is pointed out. This blindness was the situation with Buddy, who not only could not perceive his pattern of self-defeat but also seemingly did all he could to reinforce it.

Vignette Questions

1. Why do you think Buddy is the way he is: childhood, inborn events, or social learning?
2. Do you agree that Buddy is unaware of his self-defeating behavior?
3. Why or why not?

have low levels of inferiority feelings on both administrations. The third group, however, comprises those young people who overcame their former high level of inferiority feelings and now score low on the test.

Recall that the hypothesis being tested states that people who overcome feelings of inferiority will develop the highest levels of maturity. To test this hypothesis, the researchers

must find out whether the participants in Group 3 are significantly more mature than those in Groups 1 and 2.

Let us assume that in order to make this determination, the researchers administered the Washington University Sentence Completion Test (WUSCT) to measure ego development (Loevinger, 1998). They found that the participants who had overcome feelings of inferiority had statistically higher levels of ego development, which the researchers posit as synonymous with emotional maturity. They have proven their hypothesis—or have they?

We can readily find several problems with the validity of the outcomes in this hypothetical experiment. First, the researchers seem to have a problem with the **internal validity** of the experiment. In experiments with high internal validity, almost all **confounding** factors are controlled, so that any change in the **dependent variable** can be attributed to the changes in the **independent variable**. The dependent variable in this experiment is feelings of inferiority. The researchers identified participants who showed reductions in this measure, but they did not control all the variables that could have produced the change. Moreover, they concluded that changes in feelings of inferiority *caused* differences in the subjects' level of maturity.

The researchers may very well have additional problems with their **construct validity**. Their outcome measure—the measure of maturity in this case—is determined by using the WUSCT, a test of ego development. Using this particular instrument implies that ego development is synonymous with maturity. This equivalence is true only if we have a precisely defined measure of ego development and if this measure has been demonstrated to be directly proportional with maturity.

Finally, there also seems to be a problem with **external validity** in this experiment. An experiment with satisfactory external validity can be applied in settings outside the original experimental area. If our experiment has a high level of external validity, we can reasonably conclude that all people everywhere can be expected to be more mature if they have overcome feelings of inferiority. But an obvious problem with this study lies in the selection of the participants, all of whom were university students. Are the individuals we select for our study truly representative of all people we are seeking to understand? These researchers, like many others, are inclined to study the groups most likely or available to participate in research—such as college psychology majors. Of course, there have been many important and elegant studies in all subdisciplines within psychology that recruited students as participants. We still face the troublesome concern, however, as to whether we can legitimately conclude that measures of personality among college students apply to people old enough to be their parents, or to people who never completed high school, or to people from other countries.

Replication and Verification

Irrespective of the specific theory of personality under consideration, relevant research must meet some basic criteria to be scientifically sound. One of the most essential of these criteria is **replicability**. This criterion requires presenting a study examining a principle of personality in a way that allows other researchers to repeat the study. Suppose, for example, a theorist concludes that stinginess and a need to control others results from restrictive parental toilet training. The theorist bases this conclusion on comments about childhood

62 PERSONALITY THEORIES

memories made by his clients during therapy sessions. He presents as further evidence that clients who gain insight into the origin of their controlling or stingy behavior during therapy showed marked reduction in compulsive or anal-retentive acts or attitudes.

We now have a theory or partial theory of personality based on idiographic or case study research. Can other personality psychologists accept this theory as fact? Or must they test it? Obviously, before scientists accept any new theory as an explanation of real-world observations, they must show that it is something other than a chance outcome, a biased interpretation of data, an error, or even a fraud or deception.

Next, given the requirement of testing this theory about the effects of toilet training on personality, how should we go about proving it? At best, it would be very difficult, at worst impossible. To replicate and confirm this therapist's findings, other researchers would have to answer several questions. Did this particular therapist record in detail every exchange in therapy sessions that led to his conclusions? Did he explain whether he combined observations from various clients or how much consistency he found across his client population? Was he able to confirm the accuracy of their reporting of their past? And was he able to correlate the magnitude of restrictive parental behavior among his clients? What was his **outcome measure**—that is, how did he quantify or set his standards for determining personality change or improvement? Did he discuss the adjustments he made for observed individuals who did not conform to what he considered typical conduct? Even if this therapist-researcher did document these and related questions, it would still be very difficult and time consuming for other psychologists seeking to confirm his findings.

▣ APPROACHES TO THE STUDY OF PERSONALITY

True Experiments

A **true experiment** is one in which the researcher controls all possible variables that can influence the outcome of the experiment. In a true experiment, the researcher wants to examine the results of changing one or more environmental factors to determine the effect these variables have on the measure being studied. All true experiments follow the same basic procedures. First, the subjects in a true experiment are randomly assigned to either the experimental or the control group. In this application, *random* assignment means that each participant has an equal chance of being assigned to either group. The randomness of the assignment minimizes the chances that the subjects in the experimental group are significantly different from those in the control group. It follows that the groups being compared do not differ in any meaningful or systematic way. Then any differences found after the experiment can be attributed to the treatment or **intervention** that the experiment sought to study.

Some questions, however, do not allow for random assignment of research subjects; thus the research they generate will of necessity be only quasi-experimental. As examples of **quasi-experiments**, let us assume that we are interested in the differences between blind and sighted people, young or old, males or females. Obviously, subjects cannot be randomly assigned to a blind or sighted group or to a group of men only or women only.

To give a specific example of a quasi-experiment, we can review an experiment (von Mayrhauser, 1989) in which the researchers wanted to study the phenomenon that people with

introverted personalities tend to have central nervous systems (CNS) that are more easily aroused. Greater CNS arousal would then lead to a potentially uncomfortable level of mental stimulation when the person confronts ordinary life events, which would result in **introversion** as an effort to escape overstimulation. In contrast, **extroverts** with lower levels of baseline stimulation would be prone to seek stimulation and would need far greater levels of external activity to reach a comfortable level of stimulation. One result of this behavioral dichotomy is that extroverts tend to have shorter reaction times to events when compared to people classified as introverts. The researchers cite evidence that a **neurotransmitter** called **dopamine** (DA) seems to play a role in some of these differences between introverted and extroverted people.

The researchers further noted that the glutamate neurotransmitter system in the brain acts in concert with the dopaminergic system, but they found no evidence that glutamate plays a role in the extraversion/introversion dichotomy. They then set out to test the hypothesis that glutamate activity in the brain does indeed play a role in the dichotomy. Their experiment involved dividing 48 randomly selected and demographically similar participants into introversion and extraversion groups based on their scores on the Eysenck Personality Questionnaire, revised extraversion scale. Both groups were given a drug called **memantine**, developed in the 1990s to treat Alzheimer's disease. Memantine blocks the receptors in the brain to which glutamate binds.

This experiment was performed in a **double-blind** fashion. Double blinding means that neither the subjects nor the researchers who rated their performance knew who received the active drug and who received a **placebo**, an inactive substance given to the subject as if it were an active medication. The participants were given a series of reaction time tasks both with the drug and with a placebo. The results revealed that introverts taking the real drug had statistically significant longer reaction times than did the extroverts. These results provided meaningful evidence of the role of glutamate receptors in the personality dimension of introversion-extraversion.

The memantine test is an example of a true experiment. The researchers made reasonable efforts to control for every variable that might lead to a change in the dependent variable. When these precautions have been taken, virtually all changes in the dependent variable can be attributed to the independent variables that the researchers controlled.

Quasi-experiments

The prefix *quasi* means “approximately” or “as it were,” so that a **quasi-experiment** is a research project that is virtually but not quite a true experiment. Quasi-experiments generally lack some of the precise control of confounding factors that true experiments provide. Quasi-experiments are employed when researchers cannot randomly allocate subjects to control or treatment groups because random assignment is either functionally or ethically impossible. As the control over any variable that might alter the outcome of an experiment decreases, however, the chance increases that some factor other than the independent variable will intrude or confound and produce an observed change in the dependent variable. The interfering factor is what is meant by the term *confound*. Quasi-experiments are the most common type of experiment used in psychological research (Mark & Reichardt, 2004). They are commonly used in non-laboratory settings in the study of personality and in clinical psychology.

64 PERSONALITY THEORIES

The independent variable in quasi-experiments will often be a participant variable such as age, socioeconomic status, or a personality attribute. The independent variable also can be an environmental variable, such as a natural disaster, economic trend, or social change. Or the independent variable may involve a time variable, such as aging. Because each of these types of variables is inextricably linked with other factors, the experimenter always has other possible explanations for the conclusions he or she draws from the experimental results.

For example, let us suppose that one researcher studies achievement among people who score high on extraversion and discovers that they tend to earn higher incomes than individuals who score low on extraversion. Can we conclude that extraversion is the direct cause of higher incomes in those who possess this trait? Suppose extraverts are outgoing because they tend to be perceived as more physically attractive or because they come from more stable and affluent families. Suppose further that being physically attractive and coming from affluent homes often leads to high achievement. If these two suppositions are indeed the case, then these attributes lead to both extraversion and achievement, and the apparent connection between extraversion and income is illusory. Affluence and attractiveness are uncontrolled variables that are confounds in this study.

The greater the number of uncontrolled variables in a quasi-experiment, the higher the likelihood that the researchers will fail to draw a cause-and-effect relationship from their study. In fact, true experiments have the great advantage of being the only research method that does define cause-and-effect relationships. In quasi-experiments, the determination of cause and effect becomes more uncertain. Such nonexperimental methods as correlational studies cannot determine cause and effect.

Correlational Studies

Correlational studies are studies that aim to identify relationships between variables. They can yield three kinds of results: no relationship, positive correlation, and negative correlation. In a positive correlation, as one variable rises in value, the other rises as well. In a negative correlation, as one variable rises, the other falls. Correlation is measured by using a coefficient that ranges between 1 (perfect positive correlation) to 0 (variables not related at all) to -1 (perfect negative correlation).

The major limitation of correlational studies is that a correlation does not establish a causal relationship. For example, let us consider two variables, (A) high self-esteem and (B) academic achievement, which have a positive correlation. Does (A) lead to (B)? Or is it the other way around? Or are there some other hidden factors that produce both (A) and (B)?

As it turned out in real life, two carefully conducted independent studies found that there is no causal relationship between these two factors. They are correlated because both of them are linked to other factors, such as intelligence and the social status of the subjects' families. When these factors were controlled, the correlation between academic achievement and self-esteem disappeared. You must be very careful about not falling into the trap of assuming that a correlation implies a causal relationship; and you must recognize that in advertising, politics, or the media, people with an agenda may intentionally misuse correlations to "prove" their point. In men, for example, the degree of hair loss is positively correlated with length of marriage. Does marriage (or a certain number of wedding anniversaries) cause hair loss? No—age and an inherited tendency to male pattern baldness "causes" both variables.

To give another example, a statistically significant positive correlation has been found between stork populations and birth rates in Europe (Matthews, 2000). This finding might be taken as clear evidence for scientific storkism, a theory of human reproduction that should be taught in schools alongside the notion that sexual intercourse is necessary to make babies. Well, does it not? In a more serious example, does the completion of a college degree lead to a higher income for its possessor? Or does a student's socioeconomic background lead to both successful completion of a college education and a higher income? To study cause-and-effect relationships, we must hold socioeconomic factors and all other factors constant. That can be done in true experiments but not in correlational studies.

Correlational studies may suggest that variables influence one another, but they are never proof of causality. Without an explicit awareness of this difference between the two types of studies, none of us are immune from being misled by apparently scientific results. Even highly paid expert analysts of football games have a hard time understanding that running the ball a lot does not lead to wins even though the two are correlated. Rather, wins and attempts at rushing are correlated because teams in the lead typically run in order to run out the clock.

Case and Epidemiological Studies

When researchers are trying to document inductive reasoning and demonstrate by example behavioral phenomena, whether of people, animals, businesses, or systems of some kind, they often use a **case study** approach. In general, a case study is a relatively short and factual account of a sequence of events chosen to illustrate a broader principle.

Here we present a case study of a client treated by one of the authors:

Anita suffered from severe anxiety and depression that persisted despite high dosages of tranquilizers. She was the mother of a young son and was married to a man nearly twice her age who was afflicted with a life-threatening illness. Anita had left her husband on two occasions for what proved to be the illusory love of other men but returned in response to family pressure to attend to her dying husband, the father of her son.

In the initial stages of therapy Anita focused on stress at work and anxiety about her husband's illness. After more than 2 months of therapy, she revealed that she suffered from panic attacks that had begun when she was an adolescent. These episodes had been sufficiently severe for her school counselor to arrange for her to be placed in a school for emotionally handicapped students. This placement did not serve as a stabilizing influence in her life; rather, it led to sexual acting out. Despite becoming a social outcast, Anita continued this behavior.

This pattern of sexual acting out continued into her adult life. In Anita's complaints about the men in her life, it became clear that she selected men who were manifestly abusive, antisocial, or polygynous (having sex with many women). Despite her apparent high intellect and completion of advanced degrees, she initially denied seeing any danger in dating married men or men who had several other girlfriends. She also denied seeing the obvious self-harm inherent in her pattern of serial sexual relationships. This denial led to her becoming emotionally attached to men who would drop her after brief sexual relationships. Remarkably, she failed to see this pattern until it was pointed out in therapy. Anita said she enjoyed the sordid and demeaning aspects of her behavior and did not think beyond the initially positive reaction of her boyfriend of the moment.

66 PERSONALITY THEORIES

At a later point in treatment, she indicated that she was aware of the double standard that most men have regarding sexual behavior but failed to see its significance in her own life. Specifically, most of the men in her life both encouraged and enjoyed her licentious behavior but then despised her for it. This dynamic created a vicious circle of rejection, increased self-contempt, a higher degree of arousal from participating in sexually degrading activities, and further rejection. In addition, most of Anita's nonsexual friendships were unstable, typically lasting a few months at most. This pattern was replicated in her career, in which she would leave her job impulsively after a brief period of employment.

When Anita's behavior was explored with the therapist, she admitted the existence of a painful conflict—she had grown to resent and distrust men, yet was strongly drawn to them sexually. She seemed to operate according to the tacit belief that because men and the act of sex would ultimately degrade her, she might as well accept degradation as part of sex. This persistent seeking of self-defeating relationships accelerated after her husband died. Anita reopened two relationships with married men, both of whom had promised to leave their wives for her. Neither of these men took any action to pursue his relationship with her beyond meeting her for furtive sexual encounters. She would become repeatedly enraged after each tired excuse they offered for not leaving their wives, yet she would accept their next call.

The underlying reason for Anita's behavior was clarified, however, when the topic of child abuse arose by chance during a therapy session. In a nonchalant fashion, she began to describe her mother's bizarre methods of disciplining her as a child. Anita was frequently spanked, beaten, and sexually violated throughout her childhood, although at unpredictable intervals. Discussions of this period of her life often precipitated periods of acute anxiety.

Anita was treated with Rational Emotive/Cognitive Behavioral Therapy, which addressed her failure to adhere to any regimen in her life—including work attendance, work performance, sexual self-restraint, and the maintenance of stable relationships. The central premise of her irrational behavior was her belief in the inevitability of failure and rejection by significant people in her life. Anita was also given help in her ability to cope with life stresses and taught that she had considerable control over her reactions to adverse events.

Anita's relationship problems were addressed in therapy by helping her examine her choices of men, and how these misguided choices were related to her poor self-concept. She was guided through behavioral changes in both her work and social life. In addition, she was helped to understand that satisfying her desire to debase herself would lead only to despair and anxiety when she was not sexually aroused. She came to understand how these behavioral vicious circles operated in her life and gradually improved.

Case studies like the example of Anita are widely used in the study of phenomena that by their nature cannot be reproduced in a controlled laboratory setting. The unit of analysis in a case study may be an individual, a group, or a complex social entity like a corporation or non-profit organization. Like the other research strategies discussed in this chapter, case studies have their advantages and disadvantages.

The advantages of case studies are considerable. They include giving researchers opportunities to study unusual or uncommon behaviors or conditions and providing a rich source of new hypotheses. A case study can uncover detailed individual experiences that may provide very powerful insights into real human experience and behavior. It can also generate rich and valid information that has not been altered by the subject's exposure to questionnaires,

psychometric evaluations, or other tests. Using Anita's case as an example, the reader can see that such a study has high external ecological validity because the reported behavior comes from a real-life situation rather than the artificial environment of a psychology laboratory. And unlike the limited time frame of an experiment, in a case study, the therapist can follow the client for an extended period, during which he or she can build up a detailed picture of the client's behavior and monitor signs of progress. In addition, a case study often provides opportunities to meet with family members to garner further information related to the client. In the case of Anita, the therapist was able to meet with her son and with one of her current paramours. Last, the case study method provides opportunities to help subjects, as case studies are usually based on persons in psychological treatment.

The disadvantages of case studies include the possibility of bias on the observer's part. This bias includes the fact that the therapist will be using his or her preferred approach to therapy. Case studies being conducted by a therapist of one school will often neglect competing theoretical explanations of the observed behavior. For example, a psychiatrist might use a medical model to explain the patient's behavior whereas a therapist trained in family systems therapy might look for explanations in the patient's relationships with others. In addition, the data presented by the client are also subjective, which means that they are related from the client's own perspective rather than from the viewpoints of others involved in the client's life.

A second disadvantage of the case study approach is the difficulty of drawing generalizations from one instance—that is, observations of and conclusions about the psychological situation of one individual might not be meaningfully applied to others. A third disadvantage is that the client may be changed by the very act of undergoing treatment; that is, the familial, social, and psychological status of the subject of the case study may change as a result of the therapeutic process. Last, the data underlying most psychological research can be reviewed by peers and the experiment replicated if it seems necessary. Peer review, however, cannot be carried out for case studies because of the confidentiality required in mental health treatment.

There are, however, several important implications of the case study of Anita as an example of research. First, it explores the present status of a woman who was emotionally and physically abused in childhood; thus, it sheds some light on the impact the abuse may have had on her adult personality. Abuse often creates difficulties for children because it prevents them from resolving the conflicts at each stage of development in a normal fashion. In addition, Anita's lifestyle suggests a possible diagnosis of borderline personality disorder (BPD); thus, her case may encourage additional research into the origins of this pathology or of personality disorders in general. The case of Anita may provide evidence for the way personality disorders develop or are exacerbated. Finally, Anita's response to treatment with REBT/CBT suggests that even disorders arising from childhood trauma may be effectively addressed by changes in the client's present-day beliefs, attitudes, and behaviors.

▣ PERSONALITY AND PSYCHOMETRIC TESTS

Objective Personality Measures

The term **objective test** does not necessarily mean that the test lacks bias in its construction, but rather that the scoring is objective and requires no judgment or subjective interpretation on the part of the examiner. Objective tests offer researchers the advantage of

68 PERSONALITY THEORIES

excluding the problem of having as many test evaluations as evaluators. In addition, objective tests can be scored by computer.

Minnesota Multiphasic Personality Inventory. Starke R. Hathaway, a psychologist, and J. Charnley McKinley, a psychiatrist, developed the Minnesota Multiphasic Personality Inventory (MMPI) over several years beginning in the late 1930s (Friedman, 2004), from a pool of more than 1,000 possible test items. Their goal was to develop a simple objective test that would measure many different types of psychopathology based on empirical rather than intuitively developed scales. Empirically derived tests rely on research data to determine the meaning of a pattern of responses. Hathaway and McKinley began with approximately 1,000 test questions culled from a wide array of sources, including other psychometric tests, case histories, and diagnostic reports.

The two researchers gave these questions to 724 adults who were mostly friends and relatives of patients in the University of Minnesota hospitals. The people in their sample had been assessed by clinicians to find if they met any major psychiatric diagnoses. They then tried to see which questions distinguished the various diagnostic groups from one another, and from those who did not warrant a diagnosis. Thus, if a majority of subjects answered a given question as “true,” that question did little to distinguish one group from another and was rejected. On the other hand, if one diagnostic group endorsed an item much more frequently than another, the question would help to distinguish between it and other diagnostic groups and was retained. Hathaway and McKinley hoped that the inventory they designed would allow them simply to test someone and determine their diagnosis. Their finished product, although flawed by being normed on a very homogenous sample, set the standard for what became the most widely used and widely researched tests in psychology (Butcher, 1990).

Millon Clinical Multiaxial Inventory. The Millon Clinical Multiaxial Inventory or MCMI was developed by Theodore Millon, an American expert on personality theory, to put into practice his own model of personality or **personology**. Unlike the MMPI, the MCMI was developed from the responses of people in clinical treatment, and its results are keyed to the American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV). Millon served on the Personality Disorders Work Group for DSM-IV and its predecessor, DSM-III-R. Since the introduction of the MCMI in 1977, it has become one of the most widely used and researched clinical assessment instruments in use, investigated in more than 500 articles and at least six books. The MCMI is used to evaluate both the personality issues and clinical problems that a person seeking mental health treatment might confront. The test has 175 true/false questions. The MCMI-III version of the test includes 14 personality pattern scales coordinated with DSM-IV Axis II disorders and 10 clinical syndrome scales related to DSM-IV Axis I disorders. In addition, three modifying indices and a validity index help the administrator to detect careless, confused, or random responding.

With regard to validity, the MCMI has been shown to be a reasonably valid measure of personality disorders and other pathology defined by DSM-IV (Craig, 1999, 2005), including its measures of depression (Davis & Hays, 1997), antisocial personality disorder (Hart, 1991; Messina, Wish, Hoffman, & Nemes, 2001), and antisocial behavior (Kelln, 1998). Although MCMI scales may be related to depressive disorders (Choca, Bresolin, Okonek & Ostrow, 1988) most studies have found poor concurrent validity of the Major Depression scale for both the MCMI-I and MCMI-II tests.



SOURCE: Copyright © 1994, 2005 Dicandrian, Inc. All rights reserved. Published and distributed exclusively by NCS Pearson, Inc. P.O. Box 1416, Minneapolis, MN 55440. Reproduced with permission by NCS Pearson, Inc.

70 PERSONALITY THEORIES

California Personality Inventory. The California Personality Inventory (CPI), which is sometimes called the CPI-434 because the present version of the test contains 434 items, was designed by Harrison Gough, a professor at the University of California, Berkeley, in 1956. It is similar to the MMPI in the structure of its true/false questions; however, its goal is not to detect psychological maladjustment or severe mental disorders but rather to measure the more positive aspects of personality. The CPI is scored on 18 scales grouped into four categories: (1) measures of poise, self-assurance, and interpersonal adequacy; (2) measures of responsibility, intrapersonal values, and character; (3) measures of intellectual efficiency and the potential for achievement; and (4) measures of intellectual modes and interest modes. It is used most often by business corporations and other organizations to identify people with strong leadership potential and to form teams of employees who will work well together.

Projective Personality Measures

Projective personality measures invite the subject taking the **projective test** to respond to ambiguous stimuli and thus to reveal their inner thoughts, hidden feelings, and internal conflicts. More than half a century ago, the following explanation was given for the use of projective techniques in personality psychology:

The concept of projection as used in projective procedures is one formed on the pattern of projector and screen. In this sense, a projection has occurred when the psychological structure of the subject becomes palpable in his actions, choices, products, and creations. Therefore, when a procedure is so designed as to enable the subject to demonstrate his psychological structure unstilted by conventional modes, it is projective. (Rapaport, Gill, & Schafer, 1946, p. 7)

Rorschach Inkblot Test. The Rorschach inkblot test, which is one of the best known and most frequently administered projective personality measures, was devised in the early 1920s by Hermann Rorschach (1884–1922), a young Swiss psychiatrist who published a book called *Psychodiagnostics* (*Psychodiagnostik* in the original German) in 1921. In this book, Rorschach detailed his method of using inkblots to reveal a person's unconscious thinking. He obtained his data by studying the responses of 305 psychiatric patients (Alexander & Selesnick, 1966). Since Rorschach had been strongly influenced during his medical school years by Eugen Bleuler (1857–1940) the professor of psychiatry who had also taught Carl Jung (Ellenberger, 1970) he believed that much of the patient's mental processes could not be accessed directly. That is, Rorschach thought that simply talking to patients about their emotions and how they understood their problems would not produce an accurate understanding of their minds.

Jung envisioned the unconscious mind as possessing a great degree of complexity represented by numerous symbolic entities, so it followed that Rorschach, who used Jung's word-association test in his own work with schizophrenics, would have sought a way to uncover these entities through a test that used vague and complex symbolic stimuli. It is thought that Rorschach's interest in inkblots as a stimulus was influenced by his father's work as an art teacher and his own early interest in becoming an art teacher as well. He entered medical school, however, on the advice of Ernst Haeckel, an eminent German biologist. Rorschach died from a ruptured appendix only a year after the publication of *Psychodiagnostics*, at the relatively young age of 37.

Rorschach himself usually used 15 inkblots in his examination of patients, but the test as it is administered nowadays consists of only 10 inkblots, each presented on a separate card. Five blots are in color (black and red or pastel colors), and the remaining five in black and gray. Rorschach's original method of presentation left most details of interpretation to the examiner. The subject is shown each card and asked what they see, and their response is recorded. The interviewer notes the speed of the subject's response as well as his or her social behavior—that is, whether the subject seems anxious, challenged, intimidated, indifferent, or cooperative.

Each inkblot card is shown to the participant twice. The first time is a free-association phase in which the participant freely discusses any impressions or images the card evokes. The second presentation of the card is used for the inquiry phase. In this phase, the examiner asks the participant to identify the part of each blot that reminded him or her of the image. The examiner notes whether the participant bases his or her response on the whole inkblot, on a common detail, on an unusual detail, or on a confabulatory detail—that is, whether the participant responded to something that is not actually on the card. The examiner will also score the determinant of the responses, the determinant being the portion of the blot that evoked the subject's response. The form quality of the response is also scored. Here the examiner determines whether the response is well fitted to the blot. If the participant sees movement in the blot, the examiner notes whether the participant sees a human, animal, or an inanimate object moving. Finally, the subject's reference to the use of color and shading in the card is also scored.

In Table 3.3 are some examples (Gilbert, 1980) of clinical judgments based on Rorschach responses. They are derived from either the Klopfer or Phillips and Smith scoring methods.

With regard to the validity of the Rorschach test, the advent of the Exner (1993) scoring system, which can be used with children over the age of 5 as well as with adults, improved the validity of this widely used measure of personality and pathology. A rigid insistence on the use of the Rorschach is questionable, however, as many clinicians see the inkblot test as

Table 3.3 Example of Clinical Judgments Based on Rorschach Responses

<i>Clinical Interpretation</i>	<i>Subject's Response to Card</i>
Obsessive-compulsive personality	Overelaboration of form content in responses; high number of animal responses; or a high number of responses focusing on an unusual detail of the blot
Hypochondriasis	Perceiving anatomical features in the whole of card
Antisocial personality disorder	Seeing internal organs in the blot; or a mediocre or poor use of form
Maturity	Simultaneous use of Form and Shading in responses
Extraverted personality	Low number of human movement responses; or more colors used in blot interpretation than human movement

72 PERSONALITY THEORIES

an extension of the clinical interview rather than a separate evaluation. Most agree that the Rorschach should not be used as the only basis for diagnosing a mental disorder.

Is the Rorschach nonetheless a valid and scientific measure of personality or pathology? The overall evidence suggests a very weak yes. Atkinson (1986), in reviewing research articles examining the validity of the Rorschach from 1930 to 1980, concluded that there is evidence for its validity. He suggests that poor research was responsible for the impression among many psychologists that the test itself is not valid.

In recent meta-studies comparing the validity of both the MMPI and the Rorschach (Garb, 1998; Hiller, 1999), the latter is found to be a substantially less valid measure of personality and pathology than the MMPI. The Rorschach has been shown, however, to have some predictive validity in making a diagnosis of schizophrenia (Morgan-Gillard, 2003). This should not be surprising, as 188 of Hermann Rorschach's original sample of 305 subjects were schizophrenics.

Draw-A-Person Test. Some projective tests ask the subject to engage in a creative process, usually drawing a picture. In the Draw-A-Person (D-A-P) Test, developed by Karen Machover (1949, 1953), the subject, generally a child, is asked to draw a human being and then describe the person they just drew. Based on children's drawings of human figures, this test can be used with two different scoring systems for different purposes. One system measures nonverbal intelligence while the other system screens test subjects for emotional or behavioral disorders.

During the testing session, which can be completed in 15 minutes, the child is asked to draw three figures: an adult man; an adult woman; and him- or herself. To evaluate the child's intelligence, the test administrator uses the Draw-A-Person quantitative scoring system, or QSS. This system analyzes 14 different aspects of the drawings related to specific body parts and clothing, including the presence or absence of the body part or garment, the amount of detail, and the proportions of the figures in the drawing. In all, there are 64 scoring items for each drawing. A separate standard score is recorded for each drawing, and a total combined score is obtained for all three.

The administrator of the D-A-P has three pieces of plain white 8.5 × 11 paper ready for use. The first sheet is given to the child as the examiner says, "Here, I want you to draw a person as well you can." The examiner then asks such questions as: Who is this person? How old are they? What's their favorite thing to do? What's something they do not like? Has anyone tried to hurt them? Who looks out for them?

Then the child is given the second sheet of paper as the examiner says, "Now draw a [person of the other sex]." Again the examiner asks the child questions about the person's identity, age, likes and dislikes, and similar queries.

The child is then given the third sheet and told, "Now draw yourself." As with the previous two drawings, the examiner asks the child a series of questions about the person in the picture he or she just drew.

The use of a nonverbal and nonthreatening task to evaluate intelligence is intended to eliminate possible sources of bias by reducing such variables as primary language, verbal skills, communication disabilities, and sensitivity to working under pressure. The results of the D-A-P test, however, can be influenced by previous drawing experience—a factor that may account for the tendency of middle-class children to score higher on this test than lower-class children, who often have fewer opportunities to draw pictures.

To assess the test subject for emotional problems, the administrator uses the Draw-A-Person: SPED (Screening Procedure for Emotional Disturbance) to score the drawings. This system is composed of two sets of criteria. For the first type, eight dimensions of each drawing are evaluated against norms for the child's age group. For the second type, 47 different items are considered for each drawing.

Thematic Apperception Test. The Thematic Apperception Test (TAT) is a projective personality test that was designed at Harvard University in the 1930s by Christiana D. Morgan (1897–1967) and Henry A. Murray (1893–1988). Along with the MMPI and the Rorschach, the TAT is one of the most widely used psychological tests. It is regarded as more structured and less ambiguous than the Rorschach. The TAT consists of 31 pictures that depict a variety of social and interpersonal situations. The subject is asked to tell the examiner a story about each picture. The story must include a description of the card (“This is a picture of a young boy looking at a violin inside a shop window”) and an explanation of what the characters are doing, what they have done in the past or what led to the situation, and what the characters might do in the future. The subject is also asked to describe the feelings and thoughts of the characters and the outcome of the story, if possible. The examiner records the subject's words verbatim and also notes the reaction time.

Of the 31 pictures, 10 are gender-specific while 21 others can be used with adults of either sex and with children. One card is blank. The pictures were intended to run the gamut of life situations; Murray thought that the first 10 cards would evoke descriptions of everyday situations, whereas the second set of 10 would encourage the subject to tell more unusual, dramatic, and bizarre stories about what he or she saw in the picture.

The original purpose of the TAT was to reveal the underlying dynamics of the subject's personality, such as internal conflicts, dominant drives and interests, or motives. The specific motives that the TAT assesses include the need for achievement, need for power, need for intimacy, and problem-solving abilities. After World War II, the TAT was used by psychoanalysts and clinicians from other schools of thought to evaluate emotionally disturbed patients. Another shift took place in the 1970s, when the influence of the **human potential movement** led many psychologists to emphasize the usefulness of the TAT in assessment services—that is, using the test to help clients understand themselves better and stimulate their personal growth.

The TAT is widely used to research certain topics in psychology, such as dreams and fantasies, mate selection, the factors that motivate people's choice of occupations, and similar subjects. It is sometimes used in psychiatric evaluations to assess disordered thinking and in criminal examinations to evaluate crime suspects, even though it is not a diagnostic test. As mentioned earlier, the TAT can be used to help people understand their own personality in greater depth and build on that knowledge in making important life decisions. Last, it is sometimes used as a screener in psychological evaluations of candidates for high-stress occupations (law enforcement, civil aviation, the military, religious ministry, etc.).

The TAT has been criticized for its lack of a standardized method of administration as well as the lack of standard norms for interpretation. Studies of the interactions between examiners and test subjects have found that the race, sex, and social class of both participants influence both the stories that are told and the way the stories are interpreted by the examiner. Attempts have been made to design sets of TAT cards for African American and for elderly test subjects, but the results have not been encouraging. In addition, the 31 standard pictures have been

74 PERSONALITY THEORIES

criticized for being too gloomy or depressing, therefore limiting the range of personality characteristics that the test can assess.

There is no standardized procedure or set of cards for administering the TAT, except that it is a one-on-one test. It cannot be administered to groups. In one common method of administration, the examiner shows the subject only 10 of the 31 cards at each of two sessions. The test sessions are not timed but average about an hour in length. There is no specific preparation necessary before taking the TAT, although most examiners prefer to schedule sessions (if there is more than one) over two days.

The chief risks involved in taking the TAT are a bad fit between the examiner and the test subject and misuse of the results. For example, the way the examiner asks the subject to tell the story may influence the results. Some researchers found that the examiner's use of loaded words produces distress in the subjects taking the TAT, thus skewing the results.

Because the TAT is used primarily for personality assessment rather than diagnosis of mental disorders, it does not yield a score in the usual numerical sense. The test has, however, more guides to evaluating the subject's responses than does the Rorschach. The examiner is expected to interpret the congruence between the subject's answers and the picture stimuli; the subject's compliance with the examiner's directions, the existence of conflict in the story, and finally the literal story content. Most psychologists who administer the TAT find Murray's scoring system to be overly elaborate, complex, and time-consuming.

With regard to psychometrics, many experts think that the TAT is unsound. For almost every positive empirical finding, there is a negative counterpart. On the other hand, exploration of such specific variables as the need for achievement produces respectably reliable results. The median test-retest correlation is only about .30; moreover, split-half reliabilities for the TAT have been consistently poor. Criterion-related evidence for the test's validity is difficult to document. A **meta-analysis** by Spangler (1992) indicated that average correlations between the TAT and other criteria run between .19 and .22.

Rotter Incomplete Sentences Blank. The Rotter Incomplete Sentences Blank (Rotter & Willerman, 1947), or RISB, is designed to screen for overall psychological adjustment. First published by Julian Rotter (1916–) in 1950, the RISB is a semi-objective projective measure that provides direct information about the subject's personality conflicts. The subject is given a set of 40 sentence "stems" and asked to complete the sentences. The stems, which are usually only two or three words long (for example, "I regret. . ."), are worded to elicit responses that will reveal the subject's feelings. The RISB takes about 20 to 25 minutes to administer. The subject's answers are rated on a 7-point scale measuring omissions, incomplete responses, conflict responses, positive responses, or neutral responses. The minimum possible score is 0 and the highest possible score, 240. Most subjects score somewhere between 80 and 205. Higher scores indicate greater maladjustment. There are three forms of the RISB, for high school students, college students, and adults respectively.

The RISB has certain advantages compared to the Rorschach and the TAT. It elicits concise rather than diffuse responses; it purports to measure only one construct—adjustment—rather than three or four; answers are scored on a 7-point numeric scale; and the manual that accompanies the RISB provides detailed guidelines for scoring, similar to those for the Wexler Adult Intelligence Scale, the most widely used test to measure intelligence.

The psychometric properties of the RISB include good interscorer reliability (about .90) and an internal consistency reliability of about .80. The test-retest reliability for the RISB is

close to .80 for one to two weeks after administration to the test subject; however, it drops to about .50 after several months or years. Validity studies tend to support the contention that the RISB measures adjustment; however, the evidence is weak since it is based almost entirely on a relatively small sample of college students (Lah, 1989).

▣ ETHICS IN PERSONALITY RESEARCH

The Milgram Study

The most famous—even parodied—study of human behavior was performed by Stanley Milgram (1963), a psychologist who was concerned with the problem of ordinary people’s obedience to authority in relation to such atrocities as the Holocaust. Milgram’s study, which in its original form would certainly not be permitted in the early 2000s, dramatically showed that ordinary individuals may comply with orders to inflict pain on others, leading to a collective outcry for firmer rules governing psychological research on human beings.

Milgram placed a classified advertisement in the local New Haven newspaper seeking participants for a psychological experiment at Yale University. The stated purpose of the experiment was to investigate the improvement of learning and memory through the use of punishment. People who responded encountered a formal-looking individual wearing a laboratory coat, who explained why the study might have important consequences. The subjects were paired with another participant who was actually a confederate of the researcher. All participants were told that they would be assigned either to the role of the teacher or the learner by chance. In actuality, one of Milgram’s confederates, a middle-aged Caucasian male who had been trained for the part, was always assigned the learner’s role.

Imagine that you are the subject. Drawing random slips of paper seemingly determines who will be the “learner” or “teacher.” There is a 50-50 chance that you will become the teacher, or so you believe. You are, therefore, not surprised to be told that you will be the teacher. The three of you then proceed to an adjacent room, where the learner is strapped into a chair. The experimenter explains that this apparatus is to prevent excessive movement during the experiment, but it is obvious to you that the learner could not escape from the chair if he wished.

An electrode is then attached to the learner’s arm, and conductive gel is applied to the electrode. The experimenter explains that this gel is to prevent burning and blisters. Both you and the learner are told that the electrode is attached to a generator in the other room and that electric shocks will serve as punishment for incorrect responses to the teacher’s questions. The learner asks the experimenter if “the shocks will hurt,” to which the experimenter replies, “Although the shocks will be painful, they cause no permanent tissue damage.” You, the would-be “teacher,” are given a sample 45-volt shock—a slightly painful experience—so that you will have some idea of what the higher voltages mean for the learner.

You leave the learner in his room and return to the other room, where the experimenter shows you the shock generator. The generator has 30 switches, each labeled with a voltage ranging from 15 volts up to 450 volts, in 15-volt increments. Each switch also has a rating, ranging from “slight shock” to “Danger: severe shock.” The final two switches are simply labeled “XXX.” You are told that your role is to teach the learner a simple word-pairing task, but that you must also punish him for incorrect responses. You are told that for every incorrect response, you must increase the voltage by 15 volts (that is, move up to the next switch).

76 PERSONALITY THEORIES

Now, the experiment begins in earnest. The learner (pretends to) finds the task difficult and makes numerous errors. Each error results in a shock more intense than the previous one. The shocks are weak at first, but soon, they become more intense. At 75 volts, you can hear the “learner’s” grunting noises through the wall. The same thing happens at 90 and 105 volts. At 120 volts, the learner says the shocks are getting painful. You know they are because you can hear him through the wall. At 150 volts, he cries, “Get me out of here! I refuse to go on!” Thus, despite the fact that there was no actual physical pain involved, the participants were exposed to what appeared to be great suffering.

The learner’s protests continued as the voltage is set higher and higher. If at this point, or any other point, of the experiment, you question whether you should go on, the experimenter tells you that you must continue, using such reasons as “You can’t stop now,” “He is getting paid to participate in this experiment,” or simply that “The experiment depends on your continuing compliance.” The researcher may even say, “You have no choice.” As the shocks increase, the learner screams, “I can’t stand the pain!” At 300 volts, he begins pounding on the wall and demands to be released. After 330 volts, there is no longer any noise from the learner. At this point, the experimenter tells you that the learner’s failure to respond should be interpreted as an incorrect response and that you must continue increasing the shock level. The learning task will be completed only when the highest shock level is reached.

Would you proceed to the end? That is, would you continue to give electrical shocks of dangerous intensity to a fellow subject, who may be dying or already dead, simply because you are told to do so? Of course not, you are probably thinking—only a deranged sadist would answer yes. When Milgram asked a group of 40 psychiatrists to estimate the percentage of subjects who would complete the task, the estimates he got ranged from 0 to 3 percent. The psychiatrists also predicted that most people would drop out at the tenth level (that is, administering a shock of 150 volts). They were dumbfounded to discover that 26 of Milgram’s 40 subjects completed the task—65 percent!

The results of Milgram’s experiment were shocking indeed. Here was dramatic evidence that it is relatively easy to get ordinary individuals to torture others simply because an authority tells them to do so. Two thirds of Milgram’s subjects showed total obedience to the researcher’s dictates, to the extent that they might well have been administering shocks to a corpse—the body of a fellow human being they had killed—all because an authority figure in a laboratory coat told them that they could not withdraw. And Milgram’s confederate was indeed a “fellow man.” The early 1960s were still a time of lingering racial and ethnic discrimination, so the learner was a white male of middle-class appearance who could well have been the subject’s friend or neighbor.

What, however, was the effect of Milgram’s experiment on the subjects? Did they gamely continue to “fry” the learner without any apparent qualms? No, as Milgram reported, the teacher-subjects showed intense emotional pain. Some trembled, sweated, stuttered, bit their lips, or dug their nails into their own flesh. Over a third exhibited a bizarre form of nervous laughter. Three went into seizures.

For this reason—the effect on the teacher-subjects—Milgram’s experiment became as famous for generating ethical questions about the treatment of human subjects in psychological research as for the dark secret it revealed about humanity. It led to the enunciation of a set of principles on the ethical treatment of participants in research—the very term *subject* was now often seen as dehumanizing.

The Schachter-Singer Experiment

A basic ethical issue in psychological research using human subjects is transparency regarding the purpose of the experiment. Stanley Milgram's subjects, for example, were told that the experiment concerned the effects of punishment on learning and memory, which was not its true purpose. Rather, Milgram was interested in the psychology of human obedience to authority. Another famous experiment conducted around the same time as Milgram's raised the same question about the importance of full disclosure to human subjects. That was an experiment designed by Stanley Schachter and Jerome Singer (1962) to test their theory about the cognitive processes underlying emotional experiences.

Schachter and Singer hypothesized that identical states of physical arousal would produce different emotions if they were consciously attributed to different causes. The research subjects were told that they were participating in an experiment investigating the effects of a vitamin called Superoxin, which allegedly improved vision. They were divided into four groups, none of which received a vitamin injection. Subjects in the first three groups got a shot of **epinephrine** (adrenaline), a hormone naturally produced in the body by the adrenal glands, which is released when a person is under stress. Epinephrine speeds up the heartbeat, raises blood pressure, and raises blood-sugar levels. Subjects in the fourth (control) group received an injection of saline solution, a placebo.

The subjects in the first three groups were given three different types of information about the effects of the "vitamin" injection. One group was given accurate information about the physiological effects they would experience from the injection (feeling their heart racing, mild shakiness, etc.) A second group was told nothing. The third group was misinformed; the subjects were told that their extremities would feel numb as a result of the injection and that they would develop a slight headache.

Then, the subjects mingled with one of Schachter's assistants, who had supposedly been given the same injection. The assistant behaved in one of two ways: either euphoric or even silly, joking, making paper airplanes, and generally playing around; or angry and aggressive, freely insulting the subjects. Subjects in the second or third groups (uninformed or misinformed about the effects of the injection) proved to be quite susceptible to the assistants' behavior, feeling giddy and silly if the assistant had acted that way or angry and hostile if the assistant had displayed angry behavior. The subjects who had been given accurate information about the effects of epinephrine (even though they still thought they had been given a vitamin), however, were influenced very little by the assistants' behavior.

These findings confirmed Schachter and Singer's hypothesis that people who do not have an explanation for a state of physiological arousal become dependent on situational cues (in this case, the assistants' behavior) to understand what they are feeling, whereas people who have a satisfactory explanation for the physiological changes they are experiencing do not need their present social situation to help them make sense of what they are feeling.

Other psychologists have pointed out that the results of Schachter and Singer's 1962 experiment were inconclusive—there are many other factors that influence people's interpretation of feeling states. What is important in the present context, however, is the ethical legitimacy of deceiving subjects about the nature of a substance or device used in an experiment. In this case, three groups of subjects were given an injection that they were told was a vitamin when it was really a stress-related hormone.

78 PERSONALITY THEORIES

Walster's Experiment With Self-Esteem

How does self-esteem affect receptivity to romantic overtures? Must we like ourselves before we can like others, or are we more receptive to affection from others precisely when our own self-esteem is low? The reader should remember that a true experiment requires subjects to be divided randomly into at least two groups. Elaine Walster (1965), however, decided to explore the question of romantic receptivity by actively manipulating the subjects' self-esteem rather than by dividing them into high self-esteem and low self-esteem groups based on responses to a self-esteem inventory.

All the subjects in this experiment were female. They were told that the study was intended to investigate "personality and the therapy process." The subjects were then given the MMPI or the CPI and were asked to return a few weeks later for a subsequent session. At that session, a handsome male confederate of the research team pretended to have a romantic interest in the subject and asked her for a date.

Each subject then received either a flattering or disparaging "report" from the personality test she had taken weeks earlier. She received a description of herself as either "immature" and "incapable of success," or having "great maturity" and "one of the most favorable personality structures." Subjects were then asked under the guise of further research to evaluate their feelings for the male confederate. Walster found that the subjects who had received the negative evaluations were more likely to report being attracted to the researcher's male confederate than were the subjects who had received positive evaluations. The researcher concluded that attacks on someone's self-esteem from an outside source can lead the person to like a potential date more rather than less.

The reader should note all the levels of deception involved in this experiment. First, the subjects were given false feedback from a bona fide personality test. Thus, they saw themselves, at least temporarily (until after the debriefing), as either wonderful people or failures. Second, the subjects were misled into thinking that a specific person—the handsome research assistant—found them attractive. On the strength of that misperception, many accepted the offer of a date with the assistant. Is the risk of potential psychological damage to the subjects acceptable for whatever—perhaps only slight—increase in knowledge about human psychology the experiment may provide? At least two publications (Diener & Crandall, 1978; Warwick, 1982) criticized Walster's work on the grounds that the subjects may have been hurt or angry about being deceived or about having a commonplace anxiety in the college population (about dating) exploited for purposes of research.

Using deception in research is considered ethically unacceptable by many researchers. Some psychologists will allow no exceptions to this principle, arguing that the use of deception delegitimizes science, as well as hurting participants and the reputations of researchers. A related danger is that if psychologists come to be commonly regarded as deceivers, they will not be able to function effectively as health providers in therapeutic settings. Among those who argue that deception violates basic human rights are Geller (1982), Kelman (1967), and Baumrind (1985).

Most psychologists, however, regard the use of deception as unavoidable at times if their research is to have external validity. So some point out that researchers will often be in the position of weighing the importance of discovery against the need for honesty. (Aronson, Ellsworth, Carlsmith, & Gonzales, 1990).

The American Psychological Association (APA, 2002) guidelines set three criteria that must be met for deception to be acceptable in research.

1. The benefits of the research must outweigh risks. Moreover, all alternatives to deception must be examined, and their rejection must be justified.
2. The deception must not be such that knowing about it would affect the participants' willingness to take part in the research. An example of unacceptable deception would be the concealment of physical risks.
3. Where deception is used, **debriefing** is required as soon as possible.

Debriefing involves the disclosure of any false information that may have been given to the subjects as part of the experiment. Full information about the experiment is also thought to prevent any lasting negative feelings. Researchers have a responsibility to prevent not only physical damage but also psychological damage to their participants. The empirical evidence shows that debriefing is generally effective, although some observers report mixed results. Walster, Berscheid, Abrahams, & Aronson, (1967) reported that debriefing is generally ineffective, but Holmes (1976a, 1976b) maintains that there is a preponderance of evidence supporting it.

Some researchers maintain that role playing can be as effective as deception. Evidence for the usefulness of role playing includes Greenberg (1967), Willis and Willis (1970), and Horowitz and Rothschild (1970), while other researchers (Simons & Piliavin, 1972; Yinon, Shoham, & Lewis, 1974) maintain that it is not effective. The reader should remember that when Milgram asked people how many subjects would deliver the highest level of electric shocks, estimates ranged from 0 to 3%.

And there is evidence that deception in psychological research does not hurt participants. Responding to Baumrind's (1964) criticism of the famous experiment regarding obedience to authority, Milgram (1964) wrote that fewer than 2% of his participants regretted having participated, while 84% said that they found their participation worthwhile. Gerdes (1979) found that participants did not object to deception across the 15 experiments that he surveyed. And Smith (1981) found that two years after participating in research that had involved deception and in a two-year follow-up study, none of the participants felt they had been harmed.

The debate regarding psychological research still goes on. Unlike other fields, psychology is particularly prone to a variation of the Heisenberg uncertainty principle of physics. In short, the act of measuring often alters that which is being measured. In physics, this is exemplified by the problem that detecting the location of a subatomic particle will change the state of that particle, thereby preventing complete knowledge of its condition. In psychology, letting people know that they are the subjects of an experiment will have a deep impact on the way they react to subsequent events. How is this resolved? An early answer was to either deceive the participants about the nature of the experiment or to hide the fact they were part of a research study. Deception is now rarely approved by the boards in universities that approve studies involving people. It is only done when there is no other way that the research can be performed, or the importance of the research far outweighs the risks to the subjects. Despite these requirements, human research still goes on. It does so by requiring a bit more cleverness in the design of experiments and by losing some insights that could be acquired only by deceit. Perhaps, this is a small price to pay when the alternative is the potential of doing harm.

80 PERSONALITY THEORIES

 CHAPTER SUMMARY

Personality research is particularly important as it is one of the few fields where the preponderance of theorizing has been largely predicated on the intuition of key figures. In addition, a great portion of what is taught about personality has only minimal validation by the scientific method. Psychoanalytic, person-centered, and Gestalt theories are examples of principles set forth by iconic psychologists; all have strong face validity but weak experimental support. Virtually, every other science—especially those with clinical applications—employs the scientific method to test its hypotheses. This method, which can trace its origins back nearly half a millennium, holds that a hypothesis, no matter how intuitively appealing, must be tested using experimental techniques that vitiate the common tendency for people to see what they want to see.

This is especially important in psychology, in which people are studying the attributes of other people; it is a field in which bias, culture, or emotions can profoundly affect observations and conclusions. Personality research, therefore, typically strives to be both valid and reliable using true experiments when possible or quasi experiments when not.

Personality tests, which are usually the product of years of research and statistical analysis, are also an important means to study personality. Such tests include objective measures such as the Minnesota Multiphasic Personality Inventory and the Millon Clinical Multiaxial Inventory or projective measures such as the Rorschach and Thematic Apperception Test. The objective tests, being more statistically based, are generally regarded as more efficient means of studying personality. Finally, no matter what aspect of personality is being studied, ethical treatment of study participants is always paramount.