

PART FOUR

Applied Research

CHAPTER TEN

Single-Case Designs and Small-*n* Research

CHAPTER OUTLINE

OVERVIEW

THE CASE STUDY METHOD

Characteristics

Advantages of the Case Study Method

Disadvantages of the Case Study Method

Thinking Critically About Testimonials Based on a Case Study

SINGLE-SUBJECT (SMALL-*n*) EXPERIMENTAL DESIGNS

Characteristics of Single-Subject Experiments

Specific Experimental Designs

Problems and Limitations Common to All

Single-Subject Designs

SUMMARY

OVERVIEW

So far in this book we have emphasized *group methodology*—research designed to examine the average performance of one or more groups of participants. This was particularly evident in Chapters 7, 8, and 9 when we were considering experimental methods. In this chapter we introduce two alternative methodologies that emphasize the study of a single individual. We call these methodologies *single-case research designs*.

Single-case designs have been used since scientific psychology began in the 19th century. Psychophysical methods had their origin in the work of Gustav Fechner and were described in his 1860 book, *Elemente der Psychophysik*. Fechner, and countless other psychophysicists since, relied on data obtained through experiments with one or two individuals. Hermann Ebbinghaus is another major figure in the early history of psychology who used a single-case design. In fact, the single case that Ebbinghaus studied was himself. He was both the participant and the experimenter for the research he published in his monograph on memory in 1885. Over a period of many months he learned and then attempted to relearn hundreds of series of nonsense syllables. His data provided psychologists with the first systematic evidence of forgetting over time.

Single-case studies appear regularly in psychology journals, dealing with issues ranging from cognitive therapy for Vietnam veterans (Kubany, 1997) to the study of brain processes in amnesic patients (Gabrieli, Fleischman, Keane, Reminger, & Morrell, 1995) and the treatment of motor and vocal tics associated with Tourette's syndrome (Gilman, Connor, & Haney, 2005). Cognitive psychologists who study expert performance, whether it be that of a ballet dancer, chess player, or musician, rely heavily on these methods (e.g., Ericsson & Charness, 1994). For example, several researchers recently reported on their observations of "Donny, a young autistic savant who is possibly the fastest and most accurate calendar prodigy ever described" (Thioux, Stark, Klaiman, & Schultz, 2006, p. 1155). In less than a second he can tell you the day of the week when you were born! Donny had been diagnosed with autism at age 6 years and had an IQ near the borderline of mental retardation. Yet, he was accurate 98% of the time when quizzed about days of the week between March 1, 1900, and February 28, 2100. He made systematic errors outside this range due to the fact that he seemed not to recognize that century years are leap years only if they are divisible by 400. Donny was evaluated over a 1-year period utilizing a variety of testing conditions. The researchers developed a cognitive model to explain Donny's performance and speculated on the development of savant skills in autistic individuals. Similarly, a single-case methodology was used to investigate a Japanese memorist who "set a world record for reciting the first 40,000 digits of pi" (Takahashi, Shimizu, Saito, & Tomoyori, 2006, p. 1195). His memory ability in this domain was apparently based not so much on superior natural ability but on his ability to utilize imagery mnemonics (memory aids).

In this chapter we discuss two specific single-case research methodologies, the case study method and single-subject experimental designs. The *case study method* is frequently associated with the field of clinical psychology, but investigators from fields such as anthropology, criminology, neurology, and sociology also

make use of this important method. For example, neurologist Oliver Sacks (1985, 1995) captivated millions with his vivid case studies of individuals with peculiar, and rather fascinating, brain disorders. One of Sacks' best-known books is *The Man Who Mistook His Wife for a Hat and Other Clinical Tales* (1985). In it, as the title indicates, we are introduced to a man who literally, when searching for his hat, seized his wife's head and tried to lift it off and put it on. In addition, the 1990 film *Awakenings* with Robert DeNiro and Robin Williams was based on one of Oliver Sacks' stories. These "clinical tales," as Sacks calls them, not only provide insights into the relationship between mind and brain, but also reveal how individuals adapt, cope, and succeed when faced with profound neurological deficits. We will review the advantages and the disadvantages of the case study method.

The emphasis in a *single-subject experimental design* typically is on manipulation of variables and interpretation for a single subject, even if a few subjects or a single "group" is observed. Single-subject experimental designs are also frequently called "*N* = 1 experimental designs" or "small-*n* research designs." These designs are characteristic of approaches called the *experimental analysis of behavior* and *applied behavior analysis*. As you will see, these approaches represent basic and applied applications, respectively, of a small-*n* approach. Single-subject designs are more systematic and controlled than are case studies. We will examine the rationale behind the use of these designs and provide specific illustrations of the more common single-subject experimental designs. These experimental designs represent a special case of the repeated measures design introduced in Chapter 8.

THE CASE STUDY METHOD

Characteristics

- Case studies, intensive descriptions and analyses of individuals, lack the degree of control found in small-*n* experimental designs.
- Case studies are a source of hypotheses and ideas about normal and abnormal behavior.

Key Concept

A **case study** is an intensive description and analysis of a single individual. Case studies frequently make use of qualitative data, but this is not always the case (e.g., Smith, Harré, & Van Langenhove, 1995). Researchers who use the case study method obtain their data from several sources, including naturalistic observation (Chapter 4), interviews and psychological tests (Chapter 5), and even archival records (Chapter 6). A clinical case study frequently describes the application and results of a particular treatment. For example, a clinical case study may describe an individual's symptoms, the methods used to understand and treat the symptoms, and evidence for the treatment's effectiveness. Thus, case studies provide a potentially rich source of information about individuals.

Treatment variables in clinical case studies are rarely controlled systematically. Instead, several treatments may be applied simultaneously, and the psychologist may have little control over extraneous variables (e.g., home and work environments that influence the client's symptoms). Thus, a *fundamental characteristic of case studies is that they often lack a high degree of control*. Without

control, it is difficult for researchers to make valid inferences about variables that influence the individual's behavior (including any treatment). Degree of control is one distinguishing feature between the case study method and single-subject experimental designs, with single-subject experimental designs having a higher degree of control (see, for example, Kazdin, 1998).

The form and content of case studies are extremely varied. Published case studies may be only a few printed pages long or may fill a book. Many aspects of the case study method make it a unique means of studying behavior. It differs from more experimental approaches in terms of its goals, the methods used, and the types of information obtained (Kazdin, 1998). For example, the case study method is often characterized as "exploratory" in nature and a source of hypotheses and ideas about behavior (Bolgar, 1965). Experimental approaches, on the other hand, are frequently viewed as opportunities to test specific hypotheses. The case study method has sometimes been viewed as antagonistic to more controlled methods of investigation. A more appropriate perspective is suggested by Kazdin (1998), who sees *the case study method as interrelated with and complementary to other research methods in psychology*.

The case study method offers both advantages and disadvantages to the research psychologist (see, for example, Bolgar, 1965; Hersen & Barlow, 1976; Kazdin, 1998). Before reviewing its advantages and disadvantages, however, we will illustrate the method with a summary of an actual case study reported by Kirsch (1978). It is important that you read this slightly abbreviated version of a case study carefully because we will review it when discussing the advantages and disadvantages of the case study method.

CAN CLIENTS BE THEIR OWN THERAPISTS? A CASE STUDY ILLUSTRATION

This article reports on the use of self-management training (SMT), a therapeutic strategy which capitalizes on the advantages of brief therapies, while at the same time reducing the danger of leaving too many tasks not fully accomplished. . . . The essence of this approach involves teaching the client how to be his or her own behavior therapist. The client is taught how to assess problems along behavioral dimensions and to develop specific tactics, based on existing treatment techniques, for overcoming problems. As this process occurs, the traditional client-therapist relationship is altered considerably. The client takes on the dual role of client and therapist, while the therapist takes on the role of supervisor.

The case of Susan

Susan, a 28-year-old married woman, entered therapy complaining that she suffered from a

deficient memory, low intelligence, and lack of self-confidence. The presumed deficiencies "caused" her to be inhibited in a number of social situations. She was unable to engage in discussions about films, plays, books, or magazine articles "because" she could not remember them well enough. She often felt that she could not understand what was being said in a conversation and that this was due to her low intelligence. She attempted to hide her lack of comprehension by adopting a passive role in these interactions and was fearful lest she be discovered by being asked for more of a response. She did not trust her own opinions and, indeed, sometimes doubted whether she had any. She felt dependent on others to provide opinions for her to adopt.

Administering a Wechsler Adult Intelligence Scale (WAIS), I found her to have a verbal IQ of about 120, hardly a subnormal score. Her digit

(continued)

span indicated that at least her short-term memory was not deficient. The test confirmed what I had already surmised from talking with her: that there was nothing wrong with her level of intelligence or her memory. After discussing this conclusion, I suggested that we investigate in greater detail what kinds of things she would be able to do if she felt that her memory, intelligence, and level of self-confidence were sufficiently high. In this way, we were able to agree upon a list of behavioral goals, which included such tasks as stating an opinion, asking for clarification, admitting ignorance of certain facts, etc. During therapy sessions, I guided Susan through overt and covert rehearsals of anxiety-arousing situations . . . structured homework assignments which constituted successive approximations of her behavioral goals, and had her keep records of her progress. In addition, we discussed negative statements which she was making to herself and which were not warranted by the available data (e.g., “I’m stupid”). I suggested that whenever she noticed herself making a statement of this sort, she counter it by intentionally saying more appropriate, positive statements to herself (e.g., “I’m not stupid—there is no logical reason to think that I am”).

During the fifth session of therapy, Susan reported the successful completion of a presumably difficult homework assignment. Not only had she found it easy to accomplish, but, she reported, it had not aroused any anxiety, even on the first trial. . . . It was at this point that the nature of the therapeutic relationship was altered. During future sessions, Susan rated her progress during the week, determined what the next step should be, and devised her own homework assignments. My role became that of a supervisor of a student therapist, reinforcing her successes and drawing attention to factors which she might be overlooking.

After the ninth therapy session, direct treatment was discontinued. During the following month, I contacted Susan twice by phone. She reported feeling confident in her ability to achieve her goals. In particular, she reported feeling a new sense of control over her life. My own impressions are that she had successfully adopted a behavioral problem-solving method of assessment and had become fairly adept at devising strategies for accomplishing her goals.

Follow-up

Five months after termination of treatment, I contacted Susan and requested information on her progress. She reported that she talked more than she used to in social situations, was feeling more comfortable doing things on her own (i.e., without her husband), and that, in general, she no longer felt that she was stupid. She summarized by saying: “I feel that I’m a whole step or level above where I was.”

I also asked her which, if any, of the techniques we had used in therapy she was continuing to use on her own. . . . Finally, she reported that on at least three separate occasions during the 5-month period following termination of treatment, she had told another person: “I don’t understand that—will you explain it to me?” This was a response which she had previously felt she was not capable of making, as it might expose her “stupidity” to the other person.

Three months after the follow-up interview, I received an unsolicited letter from Susan (I had moved out of state during that time), in which she reminded me that “one of [her] imaginary exercises was walking into a folk dancing class and feeling comfortable; well, it finally worked.”*

*Source: Kirsch, I. (1978). Teaching clients to be their own therapists: A case study illustration. *Psychotherapy: Theory, Research, and Practice*, 15, 302–305. (Reprinted by permission.)

Advantages of the Case Study Method

- Case studies provide new ideas and hypotheses, opportunities to develop new clinical techniques, and a chance to study rare phenomena.
- Scientific theories can be challenged when the behavior of a single case contradicts theoretical principles or claims, and theories can receive tentative support using evidence from case studies.

- Idiographic research (the study of individuals to identify what is unique) complements nomothetic research (the study of groups to identify what is typical).

Sources of Ideas About Behavior Case studies provide a rich source of information about individuals and insights into possible causes of people's behavior. These insights, when translated into research hypotheses, can then be tested using more controlled research methods. This aspect of the case study method was acknowledged by Kirsch (1978) when discussing the successful psychotherapy with the woman named Susan. He stated that the "conclusions [of this case study] . . . should be viewed as tentative. It is hoped that the utility of [this technique] will be established by more controlled research" (p. 305). The case study method is a natural starting point for a researcher who is entering an area of study about which relatively little is known.

Opportunity for Clinical Innovation The case study method provides an opportunity "to try out" new therapeutic techniques or to try unique applications of existing techniques. The use of self-management training (SMT) in psychotherapy represents a clinical innovation because Kirsch changed the typical client–therapist relationship. The SMT approach is based on teaching clients to be their own therapists—in other words, to identify problems and design behavioral techniques for dealing with them. The client is both client and therapist, while the therapist acts as supervisor. In a similar vein, Kubany (1997) reported the effect of a "marathon" 1-day cognitive therapy session with a Vietnam War veteran suffering from multiple sources of combat-related guilt. Therapy of this kind generally occurs over many sessions, but the fact that this intensive session appeared to be successful suggests a new way to conduct this type of clinical intervention.

Method to Study Rare Phenomena Case studies are also useful for studying rare events. Some events appear so infrequently in nature that we can describe them only through the intensive study of single cases. Many of the case studies described in books by Oliver Sacks, for example, describe individuals with rare brain disorders. The study of autistic savants and other individuals with exceptional memory abilities, which we mentioned at the beginning of this chapter, are also examples of how the case study is used to investigate rare events.

Challenge to Theoretical Assumptions A theory that all Martians have three heads would quickly collapse if a reliable observer spotted a Martian with only two heads. The case study method can often advance scientific thinking by providing a "counterinstance": a single case that violates a general proposition or universally accepted principle (Kazdin, 1998). Consider, for example, the theory that human language development depends on exposure to normal language during a *critical period*, from about 2 years of age to puberty (suggested by Lenneberg, 1967). A human child deprived of language exposure during

this period would, then, not be expected to acquire language. Of course, depriving a child of language experience for purposes of testing this hypothesis would be immoral. Yet, there are “nature’s experiments.” One such “experiment” involved a child, “Genie,” discovered in 1970 who had been cruelly isolated from most human contact and normal language from about the age of 2 to age 13 (Curtiss, 1977). The circumstances surrounding Genie’s imprisonment and abuse, as well as the attempts by psychologists and linguists to rehabilitate her, were chronicled by Rymer (1993). Genie’s experiences represented an opportunity for scientists to test a critical-period theory of language development:

At best, Genie could have provided a flawed endorsement of Lenneberg’s theory. But she was capable of a ringing refutation. If Genie could not learn language, her failure would be attributed ambiguously—either to the truth of the critical-period hypothesis or to her emotional problems. If Genie did learn language in spite of all that had happened to her, how much stronger the rebuttal! (Rymer, 1993, pp. 121–122)

According to the psychologist most closely associated with Genie’s language training, Genie showed some language development but it was never completely normal. This outcome challenged a “strong” version of Lenneberg’s theory that language could not be learned at all without exposure during the critical period. Genie’s case study does not refute, however, a “weak” version of Lenneberg’s theory that *normal* language development would not appear after puberty without exposure during the critical period (Curtiss, 1977, p. 209).

Tentative Support for a Psychological Theory Evidence from a case study can provide tentative support for a psychological theory. Although results of case studies are not used to provide *conclusive* evidence for a particular hypothesis, the outcome of a case study can sometimes provide important evidence in support of a psychological theory. The case of Genie is a good illustration since this case study provided important evidence in support of the weak version of Lenneberg’s critical-period theory.

Another illustration of evidence from case studies providing support for a theory comes from the memory literature. In 1968, Atkinson and Shiffrin proposed a model of human memory that was to have considerable influence on research in this field. The model, which was based on principles of information processing, described both a short-term memory (STM) system and a long-term memory (LTM) system. Although results of numerous experiments provided evidence for this dual nature of our memory, Atkinson and Shiffrin considered the results of several case studies as “perhaps the most convincing demonstrations of a dichotomy in the memory system” (p. 97). These case studies involved patients who had been treated for epilepsy via surgical removal of parts of the brain within the temporal lobes, including a subcortical structure known as the hippocampus. Of particular importance to Atkinson and Shiffrin’s theory was the case study of a patient known as H.M. (see Hiltz, 1995;

Scoville & Milner, 1957). Following the brain operation, H.M. was found to have a disturbing memory deficit. Although he could carry on a normal conversation and remember events for a short period of time, H.M. could not remember day-to-day events. He was able to read the same magazine over and over again without finding its contents familiar. It looked as though H.M. had an intact short-term memory system but could not get information into a long-term memory system. Subsequent testing of H.M. and patients with similar memory deficits revealed that the nature of this memory problem is more complex than originally suggested, but the case study of H.M. continues to be important whenever theories of human memory are discussed (for example, see Schacter, 1996).

Key Concept

Complement to the Nomothetic Study of Behavior Psychology (like science in general) seeks to establish broad generalizations, “universal laws” that will apply to a wide population of organisms. As a consequence, psychological research is often characterized by studies that use the nomothetic approach. The **nomothetic approach** involves large numbers of participants, and it seeks to determine the “average” or typical performance of a group. This average may or may not represent the performance of any one individual in the group. Rather, a researcher hopes to be able to predict, on the basis of this mean performance, what organisms will be like “in general.”

Key Concept

Some psychologists, notably Allport (1961), argue that a nomothetic approach is inadequate—that the individual is more than what can be represented by the collection of average values on various dimensions. Allport argues that the individual is both unique and lawful; the individual operates in accordance with internally consistent principles. Allport argues further that the study of the individual, called an **idiographic approach** to research, is an important goal for psychological research (see also Smith et al., 1995).

Allport illustrates the need for an idiographic approach by describing the task confronting the clinical psychologist. The clinician’s goal “is not to predict the aggregate, but to foretell ‘what any one man [sic] will do.’ In reaching this ideal, actuarial predictions may sometimes help, universal and group norms are useful, but they do not go the whole distance” (p. 21). Allport suggests that our approach to understanding human nature should be neither exclusively nomothetic nor exclusively idiographic but should represent an “equilibrium” between the two. At the very least the idiographic approach, as represented by the case study method, permits the kind of detailed observation that has the power to reveal various nuances and subtleties of behavior that a “group” approach may miss. And, as you have seen, case studies have the ability to teach us about typical or average behavior by carefully studying individuals who are atypical.

Disadvantages of the Case Study Method

- Researchers are unable to make valid causal inferences using the case study method because extraneous variables are not controlled and several “treatments” may be applied simultaneously in case studies.

STRETCHING EXERCISE

In this exercise you are to respond to the questions that follow this brief description.

One of your friends is taking an introductory psychology class this semester, and she is describing to you over lunch her reactions to what happened in her class that morning. The topic for the day's class was adult development, and the professor described two research studies related to marriage and divorce. The professor emphasized that both studies represented excellent research that had been done by leading experts in the field. The first study involved a large sample of married couples that had been randomly selected from a well-defined population. The results of this study indicated that slightly more than half of marriages end in divorce and that factors such as persistent conflict between spouses and a family history of divorce were reliable predictors of divorce. The professor highlighted statistical analyses that confirmed the reliability of these predictors. The second study was a lengthy narrative description of a couple's experiences in therapy with a marriage and family counselor. The case study described how the couple entered therapy seriously considering divorce, but they decided after a year in therapy to stay married. The professor described several specific techniques the therapist

used while working with the couple to help them understand and deal with issues such as conflict in their marriage and a family history of divorce that put them at risk for divorce.

The class period ended before the professor had a chance to describe how the findings of these two studies were related and what conclusions about divorce could be drawn from them. How would you respond to the questions and concerns your friend had after this class?

- 1 One of your friend's questions is how she can decide which study's results to believe. The first study seems to say that marital conflict and a history of divorce lead to divorce, but the second study indicates that these factors need not lead to divorce. Your friend describes that she is inclined to believe the results of the second study. She finds the personal examples the professor described from the second study more compelling than the numbers he used to support the findings of the first study. What do you think?
- 2 Your friend also questions whether either of these studies will have implications for her own life experience. That is, can she tell based on the results of these studies whether she will experience a divorce if she someday chooses to get married? What do you think?

- Observer bias and biases in data collection can lead to incorrect interpretations of case study outcomes.
- Whether results from a case study may be generalized depends on the variability within the population from which the case was selected; some characteristics (e.g., personality) vary more across individuals than others (e.g., visual acuity).

Difficulty of Drawing Cause-Effect Conclusions You are well aware by now that one of the goals of science is to discover the causes of phenomena—to identify the specific factors that produce a particular event as unambiguously as possible. One disadvantage of the case study method is that cause-effect conclusions can rarely be drawn on the basis of results that are obtained from case studies. This disadvantage arises primarily because researchers are unable to control extraneous variables in case studies. Thus, the behavior changes that take place in case studies can be explained by several plausible alternative hypotheses.

Consider, for instance, the treatment of Susan through SMT reported by Kirsch (1978). Although Susan apparently benefited from the SMT therapy, can we be sure that SMT *caused* her improvement? Many illnesses and emotional disorders improve without treatment. Case study researchers must always consider the alternative hypothesis that individuals may have improved *without* treatment. In addition, several aspects of the situation may have been responsible for Susan's improvement. Her care was in the hands of a "clinical psychologist" who provided reassurance. Also, Susan may have changed her attitudes toward herself because of the insights of her therapist and the feedback she received from her test results, not because of SMT. The therapist also asked Susan, as part of her therapy, to rehearse anxiety-arousing situations covertly and overtly. This technique is similar to rehearsal desensitization, which may itself be an effective treatment (Rimm & Masters, 1979).

Because several treatments were used simultaneously, we cannot argue convincingly that SMT was the unambiguous "cause" of Susan's improvement. As we have seen, Kirsch himself was sensitive to the limitations of the case study method and suggested that the inferences he drew based on the results of his study should be considered tentative until they were investigated more rigorously.

The case studies associated with abused and deprived children such as Genie illustrate well the difficulty with isolating causal factors. We can only try to imagine the kinds of terror experienced by a child who is isolated and physically abused by an emotionally disturbed parent. How much of the failure to acquire normal language is the result of such emotional disturbance? What else don't we know about these children? Genie had been severely ill at age 14 months and an attending physician had commented that she showed signs of retardation (Curtiss, 1977). Was Genie destined to be developmentally delayed even without the horrible isolation she experienced? The data from case studies frequently leave us wondering about the causes of behavior.

Potential Sources of Bias The outcome of a case study often depends on conclusions drawn by a researcher who is both participant and observer (Bolgar, 1965). That is, a therapist observes the client's behavior *and* participates in the therapeutic process. It is reasonable to assume that the therapist may be motivated to believe that the treatment helps the client. As a result, the therapist, even if well intentioned, may not accurately observe the client's behavior. The potential for biased interpretation is not peculiar to the case study method. We have previously considered the problems of observer bias (Chapter 4) and experimenter bias (Chapter 7).

The outcome of a case may be based mainly on the "impressions" of the observer (Hersen & Barlow, 1976). For example, Kirsch (1978) described the patient Susan's "feelings" about her ability to achieve her goals and told how she reported a "sense of control" over her life. He stated that his "impressions are that she successfully adopted a behavioral problem-solving method of assessment and had become fairly adept at devising strategies for accomplishing her goals" (p. 304). A serious weakness of the case study method is that

interpretation of the outcome is often based solely on the subjective impressions of the observer.

Bias can also occur in case studies when information is obtained from sources such as personal documents, session notes, and psychological tests. Archival records, as we described in Chapter 6, are open to several sources of bias. Further, when individuals provide information about themselves (self-reports), they may distort or falsify the information in order to “look good.” This possibility existed in Susan’s treatment. We have no way of knowing whether she exaggerated her self-reports of improvement. Another potential source of bias occurs when reports are based on individuals’ memory. Cognitive psychologists have demonstrated repeatedly that memory can be inaccurate, particularly for events that happened long ago.

Problem of Generalizing from a Single Individual One of the most serious limitations of the case study method concerns the external validity of case study findings. To what extent can we generalize the findings for one individual to a larger population? Our initial response might be that the findings for one person cannot be generalized at all. Our ability to generalize from a single case, however, depends on the degree of variability in the population from which the case was selected. For example, psychologists who study visual perception are often able to generalize their findings based on the study of one individual. Vision researchers assume that visual systems in all humans are very similar. Therefore, only one or several cases may be used to understand how the visual system works. In contrast, other psychological processes are much more variable, such as learning, memory, emotions, personality, and mental health. When studying processes that vary greatly in the population, it is impossible to claim that what is observed in one individual will hold for all individuals.

Thus, even if we accept Kirsch’s (1978) conclusion regarding the effectiveness of the SMT technique of psychotherapy, we do not know whether this particular treatment would be as successful for other individuals who might differ from the patient Susan in any of numerous ways, including intelligence, age, family background, and gender. As with findings from group methodologies, the important next step is to *replicate* the findings across a variety of individuals.

Thinking Critically About Testimonials Based on a Case Study

- Being mindful of the limitations of the case study method can be helpful when evaluating individuals’ testimonials about the effectiveness of a particular treatment.

Case studies sometimes offer dramatic demonstrations of “new” findings or provide evidence for the “success” of a particular treatment. Consider advertisements for products you see in the media (e.g., infomercials). How many people who worry about their weight can resist the example of a formerly overweight individual who is shown to have lost considerable weight by using Product X? Evidence from case studies can be very persuasive. This is both an advantage and a disadvantage for the scientific community. Case studies demonstrating new or unusual findings may lead scientists to reconsider their

theories or may lead them to new and fruitful avenues of research. Case studies, then, can help advance science.

The disadvantage of case studies, however, is that their findings are often accepted uncritically. Individuals eager to lose weight or be cured of an illness may not consider the limitations of case study evidence. Instead, the evidence offers a ray of hope for a cure. For people who have (or think they have) few alternatives, this grasping at straws may not be totally unreasonable. Too often, however, people do not consider (perhaps they do not want to consider) the reasons a particular treatment would *not* work for them.

When we fail to consider the limitations of case study evidence, we risk unfortunate consequences. For example, during the early 1980s considerable controversy surrounded the supposed cancer-curing drug laetrile (Sun, 1981). Few respectable scientists or medical researchers considered this drug to be beneficial in the treatment of cancer. Advocates of laetrile, however, presented case studies reporting positive results. Largely because of public pressure, the government carried out systematic and expensive tests of the drug under controlled conditions. Researchers did not find beneficial effects of the drug in controlled experiments. As critics of laetrile have commented, many patients who used laetrile instead of traditional therapies may have postponed or interrupted valid courses of treatment and thus contributed to the spread of their cancer.

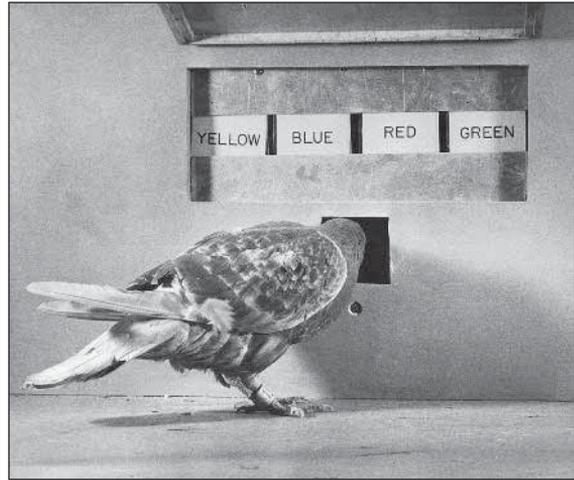
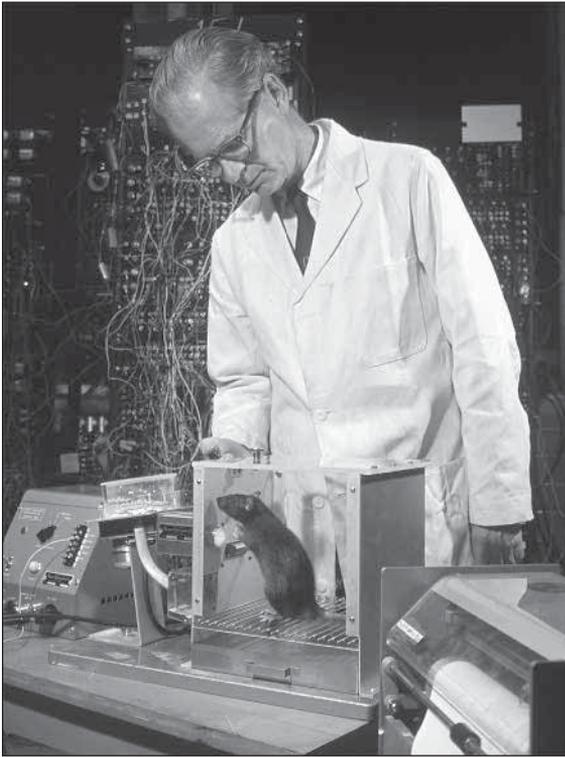
SINGLE-SUBJECT (SMALL-*n*) EXPERIMENTAL DESIGNS

- In applied behavioral analysis, the methods developed within the experimental analysis of behavior are applied to socially relevant problems.

In the remainder of this chapter we will describe single-subject experimental (small-*n*) designs. These experimental designs have their roots in an approach to the study of behavior that was developed by B. F. Skinner in the 1930s. The approach is called an *experimental analysis of behavior*. It presents a unique behavioral view of human nature that not only contains prescriptions for the way psychologists should do research but also has implications for the way society should be organized. Several of Skinner's books, including *Walden Two* and *Beyond Freedom and Dignity*, describe how the principles derived from an experimental analysis of behavior can be put to work to improve society.

In the experimental analysis of behavior (unlike the group methodologies discussed in previous chapters), it is often the case that the sample is a single subject or a small number of subjects (small-*n*). Experimental control is demonstrated by arranging experimental conditions such that the individual's behavior changes systematically with the manipulation of an independent variable. As Skinner (1966) commented,

Instead of studying a thousand rats for one hour each, or a hundred rats for ten hours each, the investigator is likely to study one rat for a thousand hours. The procedure is not only appropriate to an enterprise which recognizes individuality; it is at least equally efficient in its use of equipment and of the investigator's time and energy. The ultimate test of uniformity or reproducibility is not to be found in the methods used but in the degree of control achieved, a test which the experimental analysis of behavior usually passes easily. (p. 21)

FIGURE 10.1 Applied behavior analysis is an extension of B. F. Skinner's basic research on animal behavior.

Often there is a minimum of statistical analysis associated with single-subject experimental designs. Conclusions regarding the effects of an experimental variable (treatment) typically are made by visually inspecting the behavioral record in order to observe whether behavior changes systematically with the introduction and withdrawal of the experimental treatment. Therefore, there is considerable emphasis on appropriately *defining, observing, and recording* behavior. Has the behavior been defined clearly and objectively so that it can be reliably observed and recorded? Will a continuous (cumulative) record of behavior be kept or will observations be made at regular intervals? Although frequency of responding is a common measure of behavior, duration of behavior or other characteristics are sometimes measured. Moreover, as you will see later in this chapter, statistical issues sometimes do arise, such as excessive variability in the behavioral record, and must be dealt with. A discussion of other statistical issues associated with single-subject research designs would necessarily go beyond our brief introduction (see, for example, Kratochwill & Levin, 1992).

In *applied behavior analysis*, the methods that are developed within an experimental analysis of behavior are applied to socially relevant problems. These applications are frequently referred to as *behavior modification*, but when applied to clinical populations the term *behavior therapy* is preferred (Wilson, 1978). Behavior therapy is seen by many psychologists as a more effective approach to clinical treatment than that based on a psychodynamic model of therapy. Instead of seeking insight into the unconscious roots of problems, behavior therapy focuses on observable behavior. For example, self-stimulatory behaviors (e.g., prolonged body rocking, gazing at lights, or spinning) that often characterize autistic children may be conceptualized as behaviors under the control of reinforcement contingencies. In this way, clinicians and teachers may be able to control their frequency of occurrence by using behavior modification techniques (see Lovaas, Newsom, & Hickman, 1987). Numerous studies have been published showing how behavior modification and behavior therapy can be employed successfully to change the behavior of stutterers, normal and mentally impaired children and adults, psychiatric patients, and many others. Approaches based on applied behavior analysis have also been successfully used by school psychologists in educational settings (see Kratochwill & Martens, 1994). A primary source for these published studies is the *Journal of Applied Behavior Analysis*.

Characteristics of Single-Subject Experiments

- Researchers manipulate an independent variable in single-subject experiments; therefore, these designs allow more rigorous control than case studies.
- In single-subject experiments, baseline observations are first recorded to describe what an individual's behavior is like (and predicted to be like in the future) without treatment.
- Baseline behavior and behavior following the intervention (treatment) are compared using visual inspection of recorded observations.

Key Concept

The **single-subject experiment**, as its name suggests, typically focuses on an examination of behavior change in one individual or, at most, a few individuals. However, as we will see later in this chapter, the behavior of a single "group" of individuals also may be the focus. In a single-subject experiment the researcher contrasts treatment conditions for one individual whose behavior is being continuously monitored. That is, the independent variable of interest (usually a treatment) is manipulated systematically for one individual. Single-subject experimental designs are an important alternative to the relatively uncontrolled case study method (Kazdin, 1982). Single-subject experiments also have advantages over multiple-group experiments as described in Box 10.1.

Key Concept

The first stage of a single-subject experiment is usually an observation stage, or **baseline stage**. During this stage researchers record the subject's behavior prior to any treatment. Clinical researchers typically measure the frequency of the target behavior within a unit of time, such as a day or an hour. For example, a researcher might record the number of times during a 10-minute interview that an excessively shy child makes eye contact, the number of headaches reported

BOX 10.1

ADVANTAGES OF SINGLE-SUBJECT DESIGNS OVER GROUP DESIGNS:
LESS CAN BE MORE

Single-subject experimental designs may be more appropriate than multiple-group designs for certain kinds of applied research (see Hersen & Barlow, 1976). One such situation is when research is directed toward changing the behavior of a specific individual. For example, the outcome of a group experiment may lead to recommendations about what treatments are effective “in general” in modifying behavior. It is not possible to say, however, what the effect of that treatment would be on any particular individual based on a group average. Kazdin (1982) summarizes this characteristic of single-subject experiments well: “Perhaps the most obvious advantage [of single-case experimental designs] is that the methodology allows investigation of the individual client and experimental evaluation of treatment for the client” (p. 482).

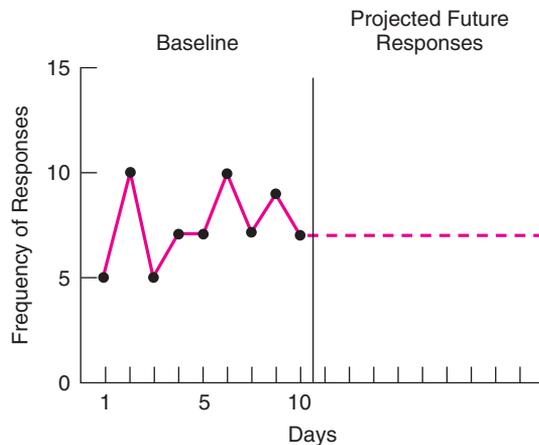
Another advantage of single-subject experiments over multiple-group experiments involves

the ethical problem of withholding treatment that can arise in clinical research. In a multiple-group design, a potentially beneficial treatment must be withheld from individuals in order to provide a control group that satisfies the requirements of internal validity. Because single-subject experimental designs contrast conditions of “no-treatment” and “treatment” within the same individual, the problem of withholding treatment can be avoided. Moreover, investigators doing clinical research often have difficulty gaining access to enough clients to do a multiple-group experiment. For instance, a clinician may be able to identify only a few clients experiencing claustrophobia (excessive fear of enclosed spaces). The single-subject experiment provides a practical solution to the problem of investigating cause-effect conclusions when only a few participants are available.

each week by a migraine sufferer, or the number of verbal pauses per minute made by a chronic stutterer. Using the baseline record, researchers are able to *describe* behavior before they provide treatment. Most importantly, the baseline allows researchers to *predict* what behavior will be like in the future without treatment (Kazdin, 1998). Of course, unless behavior is actually monitored, researchers don’t know for sure what future behavior will be like, but baseline measures allow them to predict what the future holds. Figure 10.2 illustrates this function of the baseline record. In this hypothetical example, all three measures of central tendency (mean, median, mode) converge in the prediction of future behavior. Actual observations of complex behaviors are likely not to demonstrate this uniformity, and researchers must consider the advantages and disadvantages of different measures of central tendency (see Chapter 12).

Once researchers observe that the individual’s behavior is relatively stable—that is, it exhibits little fluctuation between recording intervals—they introduce an intervention (treatment). The next step is to record the individual’s behavior with the same measures used during the baseline stage. By comparing the behavior observed immediately following an intervention with the baseline performance, researchers are able to determine the effect of the treatment. The effect of the treatment is seen most easily using a graph of the behavioral record. How did behavior change, in other words, following the experimental treatment? Keep in mind that we can predict what behavior would be like without treatment; this is represented by the dashed line in Figure 10.2. By

FIGURE 10.2 Hypothetical example of baseline observations for the frequency of responses. Data in baseline stage (solid line) are used to predict the likely rate of responses in the future if treatment is not implemented (dashed line). (Adapted from Kazdin, 1998, p. 209.)



visually inspecting the difference between behavior following treatment and what was predicted would occur without treatment, we can infer whether the treatment effectively changed the individual's behavior. Traditionally, the analysis of single-subject experiments has not involved the use of tests of statistical significance, but there has been some controversy about this (Kratochwill & Brody, 1978). Later in this chapter we will discuss some of the problems that can arise when visual inspection is used to determine whether a treatment was effective (see also Kazdin, 1998).

Although researchers have many design possibilities available (Hersen & Barlow, 1976; Kazdin, 1980), the most common single-subject designs are the ABAB design and multiple-baseline designs (Kazdin, 1998).

Specific Experimental Designs

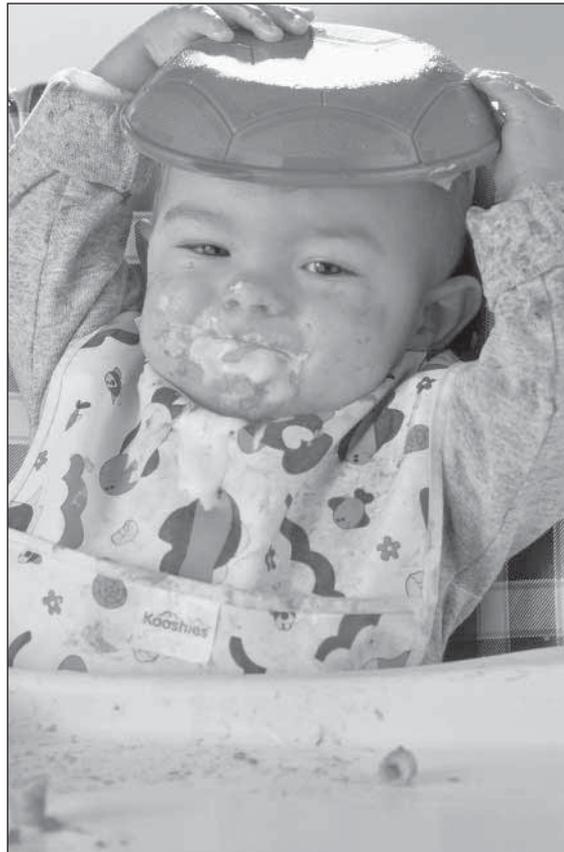
- In the ABAB design, baseline (A) and treatment (B) stages are alternated to determine the effect of treatment on behavior.
- Researchers conclude that treatment causes behavior change when behavior changes systematically with the introduction and withdrawal of treatment.
- Interpreting the causal effect of the treatment is difficult in the ABAB design if behavior does not reverse to baseline levels when treatment is withdrawn.
- Ethical considerations may prevent psychologists from using the ABAB design.
- In multiple-baseline designs, a treatment effect is shown when behaviors in more than one baseline change only following the introduction of a treatment.
- Multiple baselines may be observed across individuals, behaviors, or situations.
- Interpreting the causal effect of treatment is difficult in multiple-baseline designs when changes are seen in a baseline before an experimental intervention; this can occur when treatment effects generalize.

Key Concepts

The ABAB Design Researchers use the **ABAB design** to demonstrate that behavior changes systematically when they alternate “no-treatment” and “treatment” conditions. An initial baseline stage (A) is followed by a treatment stage (B), next by a return to baseline (A), and then by another treatment stage (B). Because treatment is removed during the second A stage, and any improvement in behavior is likely to be reversed at this point, this design is also called a **reversal design**. The researcher using the ABAB design observes whether behavior changes immediately upon introduction of a treatment variable (first B), whether behavior reverses when treatment is withdrawn (second A), and whether behavior improves again when treatment is reintroduced (second B). If behavior changes following the introduction and withdrawal of treatment, the researcher gains considerable evidence that the treatment caused the behavior change.

Horton (1987) used an ABAB design to assess the effects of facial screening on the maladaptive behavior of a severely mentally impaired 8-year-old girl. Facial screening is a mildly aversive technique involving the application of a

FIGURE 10.3 Applied behavior analysis is used to investigate methods of controlling maladaptive behavior of children and adults.



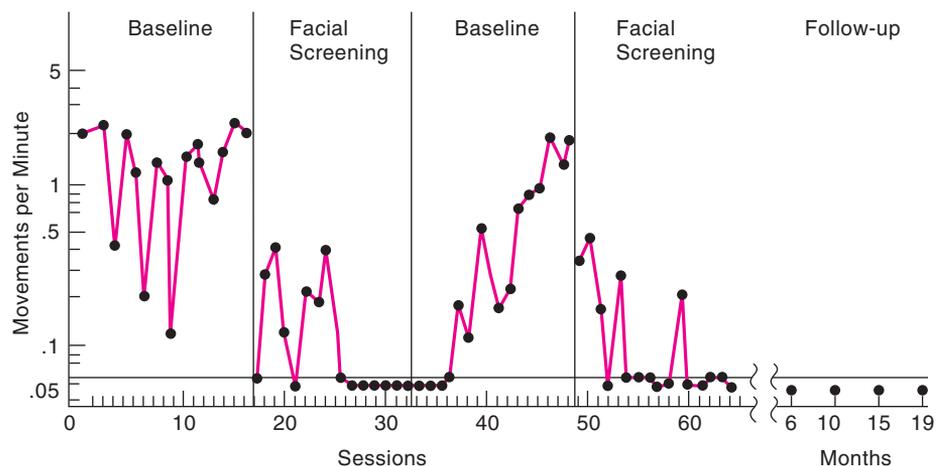
face cover (e.g., a soft cloth) when an undesirable behavior occurs. Previous research had shown this technique to be effective in reducing the frequency of self-injurious behaviors such as face slapping. Horton sought to determine whether it would reduce the frequency of spoon banging by the young child at mealtime. The spoon banging prevented the girl from dining with her classmates at the school for exceptional children that she attended. The banging was disruptive not only because of the noise but also because it often led her to fling food on the floor or resulted in her dropping the spoon on the floor.

A clear definition of spoon banging was made to distinguish it from normal scooping motions. Then, a paraprofessional was trained to make observations and to administer the treatment. A frequency count was used to assess the magnitude of spoon banging within each 15-minute eating session. During the initial, or baseline, period the paraprofessional recorded frequency and, with each occurrence of the response, said “no bang,” gently grasped the girl’s wrist, and returned her hand to her dish. The procedure was videotaped, and an independent observer viewed the films and recorded frequency as a reliability check. Interobserver reliability was approximately 96%. The baseline stage was conducted for 16 days.

The first treatment period began on Day 17 and lasted for 16 days. Each time spoon banging was observed, the paraprofessional continued to give the corrective feedback of “no bang” and returned the girl’s hand to her dish. However, the paraprofessional now also pulled a terry-cloth bib over the girl’s entire face for 5 seconds. Release from facial screening was contingent on the participant’s not banging for 5 seconds. The first treatment phase was followed by a second baseline period and another treatment phase. Posttreatment observations were also made at 6, 10, 15, and 19 months.

Figure 10.4 shows changes in the frequency of the girl’s spoon-banging behavior as a function of alternating baseline and treatment phases. Facial screening was not only effective in reducing this behavior during treatment phases;

FIGURE 10.4 Frequency of spoon-banging responses across baseline, treatment, and follow-up phases of study. (Adapted from Horton, 1987.)



follow-up observations revealed that the spoon banging was still absent months later. Following the final treatment phase, the girl no longer required direct supervision during mealtime at either school or home and was permitted to eat with her peers. There was clear evidence that the application of the facial screening was responsible for eliminating the spoon banging. The facial screening was the only treatment that was administered, and visual inspection of Figure 10.4 shows that behavior changed systematically with the introduction and withdrawal of treatment. The facial-screening technique was a successful procedure for controlling the maladaptive behavior of the young child when other, less intrusive procedures had failed.

Methodological Issues Associated with ABAB Designs A major methodological problem that sometimes arises in the context of an ABAB procedure can be illustrated by looking again at the results of the Horton (1987) study shown in Figure 10.4. In the second baseline stage, when application of the facial screening was withdrawn, spoon banging increased. That is, the improvement observed under the preceding treatment stage was reversed. What if the spoon-banging behavior had remained low even when the treatment was withdrawn? What can the researcher conclude about the effectiveness of the treatment when behavior in a second baseline stage does not revert to what it was during the initial baseline period? In Box 10.2 we describe reasons why behavior might not revert to the baseline level when the treatment is withdrawn.

BOX 10.2

WHY REVERSAL MAY NOT OCCUR IN THE REVERSAL DESIGN

One reason the behavior may not revert to the baseline level is that the behavior may not be expected logically to change once the treatment led to improvement. This occurs in situations in which the treatment involves teaching individuals new skills. For example, a researcher's treatment might be teaching a developmentally disabled individual how to commute to work. Once the skill is learned, it is unlikely to be "unlearned" (revert to baseline) when the treatment is withdrawn. The solution to this problem is straightforward. Researchers should not use the ABAB design when they can logically expect that the target behavior would not revert to baseline when treatment is withdrawn.

What other reasons are there for behavior not to return to baseline in the second stage? One possibility is that a variable *other than* the treatment variable caused behavior to change in the first shift from baseline to treatment stages. For example, the individual may receive increased attention

from staff or friends during treatment. This increased attention—rather than the treatment—may cause behavior to improve. If the attention persists even though the specific treatment is withdrawn, the behavior change is likely to persist as well. This explanation suggests a confounding between the treatment variable and some other, uncontrolled factor (such as attention).

It is also possible that, although the treatment caused behavior to improve, other variables took over to control the new behavior. Again, we can consider the effect attention has on behavior. When family and friends witness a change in behavior, they may pay attention to the individual. Think of the praise people get when they have lost weight or quit smoking. Positive reinforcement in the form of attention may maintain the behavior change that was initiated by the treatment and so we would not expect behavior to return to baseline levels when the treatment was withdrawn.

If for whatever reason behavior does not revert to baseline levels when treatment is withdrawn, researchers cannot safely conclude that the treatment caused the initial behavior change (Kazdin, 1980, 1998). The researcher must examine the situation carefully with the hope of identifying variables that might be confounding the treatment variable or replicate the procedure with different subjects (Hersen & Barlow, 1976).

Researchers can also face an ethical problem when using the ABAB design. Suppose the treatment seems to improve the individual's behavior relative to the baseline. Is it ethical to remove what appears to be a beneficial treatment to determine if the treatment actually caused the improvement? As you might imagine, withdrawing a beneficial treatment may not be justified in all cases. Some behaviors might be life-threatening or exceptionally debilitating, and it would not be ethical to remove treatment once a positive effect is observed. For example, some autistic children exhibit self-injurious behaviors such as head banging. If a clinical researcher succeeds in reducing the frequency of this behavior, it would be unethical to withdraw treatment to meet the requirements of the ABAB design. Fortunately, there is a single-case experimental design that does not involve withdrawal of treatment and that may be appropriate in such situations—the multiple-baseline design.

The Multiple-Baseline Design The multiple-baseline design also makes use of baseline and treatment stages, but not by withdrawing a treatment as in the ABAB design. As the name suggests, researchers establish several baselines when using a multiple-baseline design. The multiple-baseline design demonstrates the effect of a treatment by showing that behaviors in more than one baseline change following the introduction of a treatment.

One example of the multiple-baseline design is to treat one person's behavior in different situations. In this case, the first step in the multiple-baseline design is to record behavior (such as the aggressiveness of a child) as it normally occurs in several situations (such as at home, in the classroom, and at an after-school daycare facility). The researcher establishes the baseline frequency of the behavior in each situation (i.e., multiple baselines). Next the treatment is introduced in one of the situations (e.g., at home), *but not* in the other situations. The researcher continues to monitor behavior in all of the situations. A critical feature of the multiple-baseline design is that treatment is applied to only one baseline at a time. The behavior in the treated situation should improve; the behavior in the baseline situations should not improve. The next step is to apply the treatment in a second situation (treatment may continue in the first situation as well) but leave the third situation as a continuing baseline. Behavior should change only in the treated situation, not in the baseline situation. The final step is to administer the treatment in the third situation; again, the behavior should change when the treatment is administered in the third situation. The key evidence for the effectiveness of a treatment in the multiple-baseline design is the demonstration that behavior changes only when the treatment is introduced.

There are several variations on the multiple-baseline design, depending on whether multiple baselines are established for different individuals, for different behaviors in the same individual, or for the same individual in different

Key Concept

situations. Although they sound complex, multiple-baseline designs are frequently used and easily understood. We will describe each type of multiple-baseline design using an applied research example.

In the **multiple-baseline design across individuals**, baselines are first established for different individuals. When the behavior of each individual has stabilized, an intervention is introduced for one individual, then for another individual, later for another, and so on. As in all multiple-baseline designs, the treatment is introduced at a different time for each baseline (in this case, for each individual). If the treatment is effective, then a change in behavior will occur immediately following the application of the treatment in each individual.

An example of the use of a multiple-baseline design across individuals comes from the field of sports psychology. Allison and Ayllon (1980) were interested in evaluating the effectiveness of a coaching method that involved several behavioral techniques on the acquisition of specific football, tennis, and gymnastic skills. Although they found that the method was effective for each sport, we will describe their test of the effectiveness of behavioral coaching for the acquisition of a football skill. The participants in this experiment were second-string members of a citywide football program chosen because they “completely lacked fundamental football skills” (p. 299).

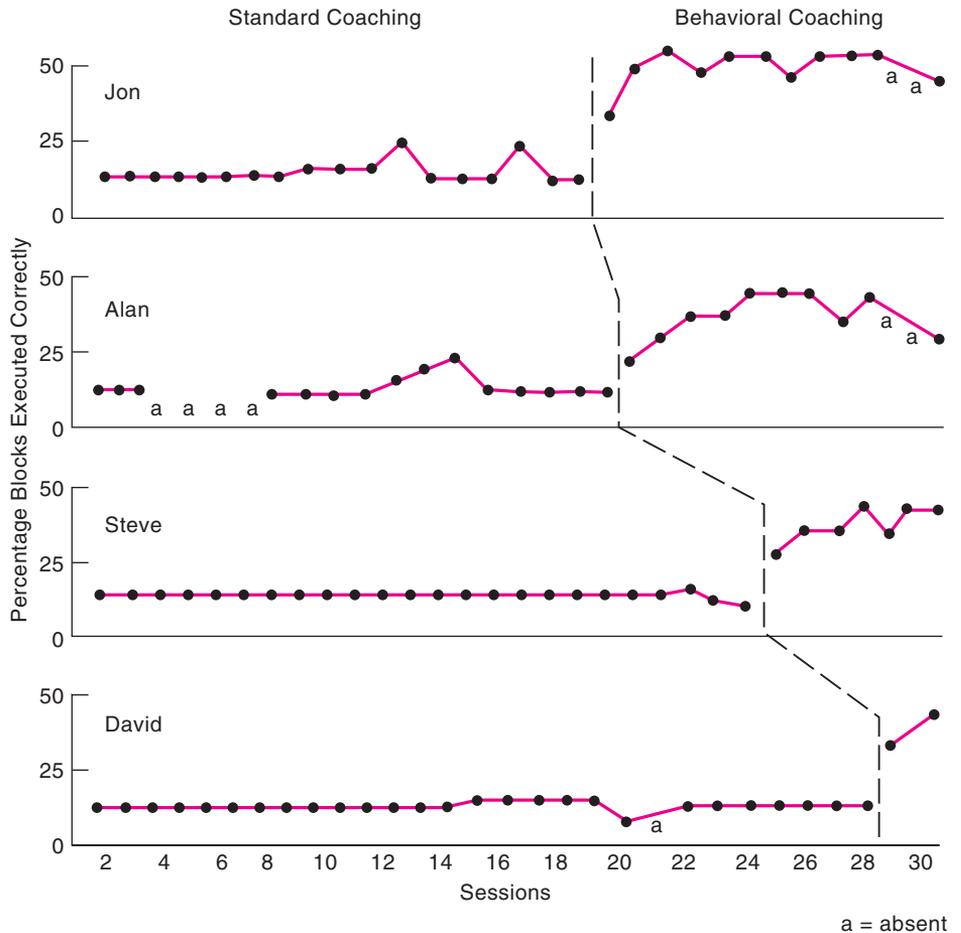
The skill to be acquired in the Allison and Ayllon (1980) study was blocking. Blocking skill was defined operationally in terms of eight elements, ranging from the body’s first being behind the line of scrimmage to maintaining body contact until the whistle was blown. Behavioral coaching involved specific procedures implemented by the team coach, including systematic verbal feedback, positive and negative reinforcement, and several other behavioral techniques. The experimenter first established baselines for several different members of the football team under “standard coaching” conditions. In the standard procedure, the coach used verbal instructions, provided occasional modeling or verbal approval, and, when execution was incorrect, “loudly informed the player and, at times, commented on the player’s stupidity, lack of courage, awareness, or even worse” (p. 300). In short, it was an all-too-typical example of negative coaching behavior.

The experimenter and a second observer recorded the frequency of correct blocks made in sets of 10 trials. Behavioral coaching was begun, in accordance with the multiple-baseline design, at different times for each of four football players. Results of this intervention are shown in Figure 10.5. Across four individuals, behavioral coaching was shown to be effective in increasing the frequency of correctly executed blocks. The agreement between the two observers on blocking performance ranged from 84% to 94%, indicating that the observation of behavior was reliable. The skill execution changed for each player at the point at which the behavioral coaching was introduced. Thus, there is evidence in this multiple-baseline design that the coaching method caused the change in each player’s performance.

Key Concept

A second type of multiple-baseline design involves establishing two or more baselines by observing different behaviors in the same individual, a **multiple-baseline design across behaviors**. A treatment is directed first at one behavior, then at another, and so on. Evidence for a causal relationship between treatment

FIGURE 10.5 Multiple baselines showing percentage of football blocks executed correctly by four players as a function of standard coaching and behavioral coaching. (From Allison & Ayllon, 1980.)



and behavior is obtained if performance changes for each behavior immediately after the treatment is introduced. For example, Gena, Krantz, McClannahan, and Poulson (1996) attempted to teach several different socially appropriate affective behaviors to youths with autism. As the researchers noted, children with autism often show inappropriate affective behaviors, which limit their opportunities to communicate effectively with others and to develop interpersonal relationships. Treatment included verbal praise and tokens (exchangeable for rewards) that were delivered contingent on appropriate affective responses in three or four different behavior categories. Target behaviors were selected from among the following: showing appreciation, talking about favorite things, laughing about absurdities, showing sympathy, and indicating dislike. Visual inspection of the behavioral records showed evidence for the effectiveness of the treatment. As required in the multiple-baseline design, the different affective behaviors changed immediately after introduction of the intervention for that behavior.

Key Concept

The third major variation on the multiple-baseline design involves establishing two or more baselines for an individual's behavior across different situations, a **multiple-baseline design across situations**. For example, as we described when we introduced the multiple-baseline design, a researcher might establish baselines showing the frequency of a child's aggressive behavior at home, in the classroom, and at an after-school daycare facility. As with other variations of this design, the treatment is applied at different times and the behavioral records are examined to determine whether behavior changes systematically with the introduction of treatment.

Hartl and Frost (1999) successfully treated a 53-year-old woman for compulsive hoarding. The clutter in her house took up approximately 70% of the living space such that rooms could not be used for their intended purpose. In the TV room, for example, newspapers, paid and unpaid bills, letters and other items were piled 3 feet high on the couch and spilled onto the floor, burying a coffee table. In other rooms and hallways there were numerous gifts that the client had bought with no particular recipient in mind, the piles at time reaching the ceiling. Treatment consisted of "training in decision-making and categorization, exposure and habituation to discarding, and cognitive restructuring . . . each woven into the context of weekly excavation sessions" (p. 454). A multiple-baseline design across situations was used with the situations being different rooms in her house. To measure progress the researchers computed "clutter ratios" (CRs) based on the proportion of square feet in the room covered by clutter. Treatment began in the kitchen while four additional rooms provided baseline measures; after several treatment sessions in the kitchen, the researchers moved to another room to begin treatment, and so on. Each session lasted a couple of hours with the total number of sessions continuing for more than a year. A graph showing treatment and baseline conditions across situations (rooms) provided clear evidence that CRs "decreased substantially in each of the target rooms once treatment was applied" to that room (p. 456).

Methodological Issues Associated with Multiple-Baseline Designs

- How many baselines are needed?

As with many other aspects of single-case research, there are no hard-and-fast rules for answering the question "How many baselines do I need?" The bare minimum is clearly two baselines, but this is generally considered inadequate. Three or four baselines in a multiple-baseline design are recommended (Hersen & Barlow, 1976).

- What if behavior changes before the intervention?

Problems can arise in any of the types of multiple-baseline designs when changes in behavior are seen in a baseline before the treatment has been administered. The reasons for these premature changes in a baseline are not always clear. The logic of the multiple-baseline designs depends critically on the changes in behavior occurring directly after the introduction of the treatment. Thus, when changes in baseline performance occur prior to treatment, this makes it hard to conclude that the treatment was effective. If the pretreatment changes occur in only one of several baselines (especially if there is a plausible

explanation for the change based on procedural or situational factors), the multiple-baseline design can still be interpreted with some confidence. For instance, Kazdin and Erickson (1975) used a multiple-baseline design across individuals to help severely mentally impaired individuals respond to instructions. Participants who followed instructions were reinforced with candy-coated cereal and praise, and this intervention was introduced in each of four small groups at different points in time. Performance changed directly with the application of the positive reinforcement procedure in three groups, but not in the fourth. In this group, which had the longest baseline, behavior gradually improved prior to the intervention. The researchers reasonably suggested that this occurred because individuals in this group saw other participants comply with instructions and then imitated the treated participants' behavior.

- What if the treatment generalizes to other behaviors or situations?

A problem sometimes seen in multiple-baseline designs occurs when changes in one behavior *generalize* to other behaviors or situations. When Hartl and Frost (1999) successfully treated a woman for excessive hoarding, one might speculate that treatment in one room of her house would lead her to decrease clutter in other rooms. No such decrease was observed, however, and clutter even increased slightly in the bedroom that served as a control room with no intervention.

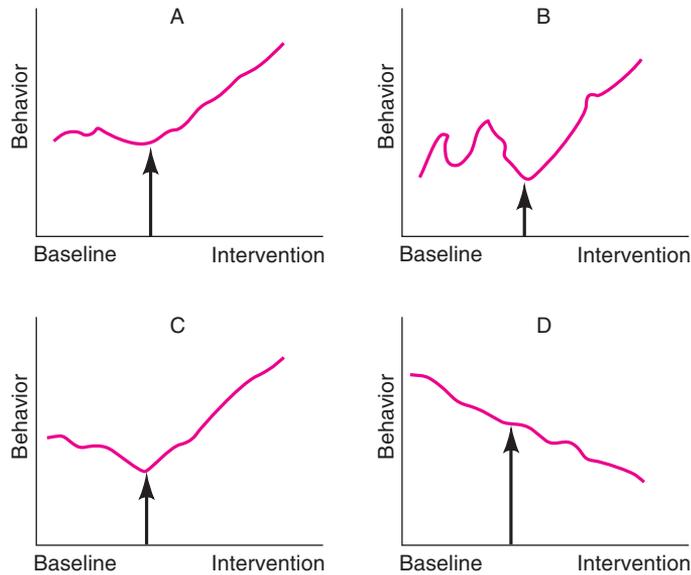
In dealing with possible problems of generalization, researchers need to keep in mind the maxim "An ounce of prevention is worth a pound of cure." If altering the behavior of one individual is likely to affect the behaviors of others, or if behavior in one situation is likely to influence behavior in another situation, or if changing one type of behavior is likely to affect other behaviors, then multiple-baseline designs may need to be modified or perhaps abandoned (Kazdin, 1998). Unfortunately, anticipating when changes will occur simultaneously in more than one baseline is not always easy, but these problems appear to be relatively infrequent exceptions to the effects usually seen in a multiple-baseline design (Kazdin, 1998). What is clear, however, is that concluding a treatment is effective using a multiple-baseline design requires that behavior changes directly follow the introduction of the treatment in each baseline.

Problems and Limitations Common to All Single-Subject Designs

- Interpreting the effect of a treatment can be difficult if the baseline stage shows excessive variability or increasing or decreasing trends in behavior.
- The problem of low external validity with single-subject experiments can be reduced by testing small groups of individuals.

Problems with Baseline Records An ideal baseline record and response to an intervention are shown in panel A of Figure 10.6. Behavior during the baseline stage is very stable, and behavior changes immediately following the introduction of treatment. If this were the outcome of the first stages of either an ABAB or a multiple-baseline design, we would be headed in the direction of showing that our treatment is effective in modifying behavior. However, consider the baseline and treatment stages shown in panel B of Figure 10.6. Although the

FIGURE 10.6 Examples of behavioral records showing possible relationships between baseline and intervention phases of a behavior modification program. The arrow indicates the start of an intervention.



desired behavior appears to increase in frequency following an intervention, the baseline shows a great deal of variability. It is difficult to know whether the treatment produced the change or behavior just happened to be on the upswing. In general, it is hard to decide whether an intervention was effective when there is excessive variability in the baseline.

There are several ways to deal with the problem of excessive baseline variability. One way is to look for factors in the situation that might be producing the variability and that could be removed. The presence of a particular staff member, for instance, might be causing changes in the behavior of a psychiatric patient. Another approach is to “wait it out”—to continue taking baseline measures until behavior stabilizes. It is, of course, not possible to predict when and if this might occur. Introducing the intervention before behavior has stabilized, however, would jeopardize a clear interpretation of the outcome. A final way to deal with excessive variability is to average data points. By charting a behavioral record using averages of several points, researchers can sometimes reduce the “appearance” of variability (Kazdin, 1978).

Panel C of Figure 10.6 illustrates another potential problem that can arise when baselines show an increasing or decreasing trend. If the goal of the intervention was to increase frequency of behavior, the decreasing trend shown in panel C poses no problem of interpretation. An intervention that reversed the decreasing trend can be taken as evidence that the treatment was effective. If the goal of the intervention was to reduce the frequency of a behavior, however, the problem would be more serious. This situation is illustrated in panel D. Here we see a decreasing trend in the baseline stage and continued reduction of frequency in the treatment stage. It would be difficult to know whether the treatment had an

effect because the decrease following the intervention could be due to the intervention or to a continuation of the baseline trend. When an intervention is expected to have an effect in the same direction as a baseline trend, the change following the intervention must be much more marked than that shown in panel D to support a conclusion that the treatment had been effective (Kazdin, 1978). This problem becomes even more troubling because a treatment effect in a single-subject design is usually judged by visually inspecting the behavioral record. It is often difficult to say what constitutes a “marked” change in the behavioral record (see, for example, Parsonson & Baer, 1992). It is an especially good idea in these circumstances to complement the observations of the target behavior with other means of evaluation such as making comparisons with “normal” individuals or asking for subjective evaluations from others familiar with the individual.

Questions of External Validity A frequent criticism of single-subject research designs is that the findings have limited external validity. In other words, the single-subject experiment appears to have the same limitation as the case study method. Because each person is unique, it can be argued that there is no way of knowing whether the effect of a particular intervention will generalize to other individuals. There are several reasons, however, why the external validity of findings from single-subject experiments may not be as limited as it seems.

First, the types of intervention used in single-subject experiments are often potent ones and frequently produce dramatic and sizable changes in behavior (Kazdin, 1978). Consequently, these types of treatments are often found to generalize to other individuals. Other evidence for the generality of effects based on single-subject experiments comes from the use of multiple-baseline designs. A multiple-baseline design across individuals, for example, is often able to show that a particular intervention was successful in modifying the behavior of several individuals. Similarly, multiple baselines across situations and behaviors can attest to the external validity of a treatment effect.

Perhaps the best way to establish the external validity of a treatment effect in a single-subject experiment is to test a “single group” of subjects. The procedures associated with single-subject designs are sometimes used with small groups of individuals (i.e., small-*n*). For example, Kazdin and Erickson (1975) found that positive reinforcement improved responsiveness to instructions in small groups of mentally impaired individuals. The researchers were able to demonstrate that a treatment was, on average, effective for a small group of participants as well as for individuals in the group. In a sense, the treatment effect was replicated several times across members of a group. Single-subject experiments like these offer impressive evidence for both internal and external validity.

SUMMARY

Two important single-case research designs are the case study and the single-subject experiment, or small-*n* design. The case study method can be an important source of hypotheses about behavior, can provide an opportunity for clinical innovation (e.g., trying out new approaches to therapy), can

permit the intensive study of rare phenomena, can challenge theoretical assumptions, and can provide tentative support for a psychological theory. The intensive study of individuals that is the hallmark of the case study method is called idiographic research, and it can be viewed as complementary to the nomothetic approach (seeking general laws or principles) that is also characteristic of psychology. Problems arise when the case study method is used to draw cause-effect conclusions, or when biases in the collection of, or interpretation of, data are not identified. The case study method also involves potential problems of generalizing findings based on the study of a single individual. Moreover, the “dramatic” results obtained from some case studies, though they may give scientific investigators important insights, are frequently accepted as valid by people who are not aware of the limitations of this method.

B. F. Skinner developed the experimental analysis of behavior. Applied behavior analysis seeks to apply principles derived from an experimental analysis of behavior to socially relevant problems. The major methodology of these approaches is the single-subject experiment, or small-*n* research. Although there are many kinds of single-subject designs, the most common are the ABAB design and the multiple-baseline design.

An ABAB design, or reversal design, allows a researcher to confirm a treatment effect by showing that behavior changes systematically with conditions of no treatment (baseline) and treatment. Methodological problems arise in this design when behavior that changed during the first treatment (B) stage does not reverse when treatment is withdrawn during the second baseline (A) stage. When this occurs, it is difficult to establish that the treatment, rather than some other factor, was responsible for the initial change. One may encounter ethical problems when using the ABAB design if a treatment that has been shown to be beneficial is withdrawn during the second baseline stage.

A multiple-baseline design demonstrates the effectiveness of a treatment by showing that behaviors across more than one baseline change as a consequence of the introduction of a treatment. Baselines are first established across different individuals, or across behaviors or across situations in the same individual. Methodological problems arise when behavior does not change immediately with the introduction of a treatment or when a treatment effect generalizes to other individuals, other behaviors, or other situations.

Problems of excessive baseline variability as well as of increasing or decreasing baselines sometimes make it difficult to interpret the outcome of single-subject designs. The problem of excessive baseline variability can be approached by seeking out and removing sources of variability, by extending the time during which baseline observations are made, or by averaging data points to remove the “appearance” of variability. Increasing or decreasing baselines may require the researcher to obtain other kinds of evidence for the effectiveness of a treatment. Finally, the single-subject design is often criticized for its lack of external validity. However, because treatments typically produce substantial changes in behavior, these changes can often be easily replicated in different individuals. The use of single “groups” of subjects (small-*n* research) can also provide immediate evidence of generality across subjects.

KEY CONCEPTS

case study	308	multiple-baseline design across individuals	326
nomothetic approach	313	multiple-baseline design across behaviors	326
idiographic approach	313	multiple-baseline design across situations	328
single-subject experiment	319		
baseline stage	319		
ABAB design (reversal design)	322		

REVIEW QUESTIONS

- 1 Identify and give an example of each of the advantages of the case study method.
- 2 Distinguish between a nomothetic and an idiographic approach to research.
- 3 Identify and give an example of each of the disadvantages of the case study method.
- 4 What is the major limitation of the case study method in drawing cause-effect conclusions?
- 5 Under what conditions might a single-subject design be more appropriate than a multiple-group design?
- 6 Distinguish between baseline and intervention stages of a single-subject experimental design.
- 7 Why is an ABAB design also called a reversal design?
- 8 What methodological problems are specifically associated with an ABAB design?
- 9 Outline the general procedures and logic that are common to all the major forms of multiple-baseline designs.
- 10 What methodological problems are specifically associated with multiple-baseline designs?
- 11 What methodological problems must be addressed in all single-subject designs?
- 12 What evidence supports the external validity of single-subject designs?

CHALLENGE QUESTIONS

- 1 A case study showing how “mud therapy” was successful in treating an individual exhibiting excessive anxiety was reported in a popular magazine. The patient’s symptoms included trouble sleeping, loss of appetite, extreme nervousness when in groups of people, and general feelings of arousal that led the individual always to feel “on edge” and fearful. The California therapist who administered the mud therapy was known for this treatment, having appeared on several TV talk shows. He first taught the patient a deep relaxation technique and a “secret word” to repeat over and over in order to block out all disturbing thoughts. Then the patient was asked to lie submerged for 2 hours each day in a special wooden “calm tub” filled with mud. During this time the patient was to practice the relaxation exercises and to concentrate on repeating the secret word whenever the least bit of anxiety was experienced. The therapy was very costly, but after 6 weeks the patient reported to the therapist that he no longer had the same feelings of anxiety that he reported previously. The therapist pronounced him cured and attributed the success of the treatment to immersion in the calming mud. The conclusion drawn by the author of the magazine article describing this therapy was that “it is a treatment that many people could benefit from.” On the basis of your knowledge of the limitations of the case study method, answer the following questions:
 - A What possible sources of bias were there in the study?
 - B What alternative explanations can you suggest for the successful treatment?
 - C What potential problem arises from studying only one individual?
 - 2 A 5-year-old child frequently gets skin rashes, and the mother has been told by her family doctor that
- (continued)*

the problem is due to “something” the child eats. The doctor suggests that she “watch carefully” what the child eats. The mother decides to approach this problem by recording each day whether the child has a rash and what the child ate the day before. She hopes to find some relationship between eating a particular food and the presence or absence of the rash. Although this approach might help discover a relationship between eating certain foods and the appearance of the rash, a better approach might be one based on the logic and procedures associated with single-subject designs. Explain how the mother might use such an alternative approach. Be specific and point to possible problems that may arise in this application of behavioral methodology.

- 3 During the summer months, you find employment in a camp for mildly mentally impaired children. As a counselor you are asked to supervise a small group of children, as well as to look for ways to improve their attention to various camp activities that take place indoors (e.g., craft-making and sewing). You decide to explore the possibility of using a system of rewards (M&M candies) for “time on task.” You realize that the camp director will want evidence of the effectiveness of your intervention strategy as well as some assurance that it will work with other children in the camp. Therefore you are to
 - A Plan an intervention strategy based on reinforcement principles that has as its goal an increase in the time children spend on a camp activity.
 - B Explain what behavioral records you will need to keep and how you will determine whether your intervention has produced a change in the children’s behavior. You will need, for example, to specify exactly when and how you will measure behavior, as well as to justify your use of a particular design to carry out your “experiment.”
 - C Describe the argument you will use to convince the director that your intervention strategy (assuming that it works) will work with other, similar children.
- 4 A teacher asks your help in planning a behavioral intervention that will help manage the behavior of a problem child in his classroom. The child does not stay at her desk when asked to do so, does not remain quiet during “quiet times,” and exhibits other behaviors that disrupt the teaching environment. Explain specifically how a positive reinforcer, such as candy or small toys, might be used as part of a multiple-baseline across behaviors design to improve the child’s behavior.

Answer to Stretching Exercise

- 1 You may be inclined to agree with your friend. Personal examples are often more compelling than quantitative evidence. In evaluating these two studies, however, it is important to recognize that they represent two different approaches to doing research. The first study represents the nomothetic approach, which relies on the study of large groups and tends to use quantitative measures to describe the groups. The second study represents the idiographic approach, which involves the intensive study of individual cases and qualitative description. After recognizing these differences in the two approaches, careful examination of the findings indicates there is no need to choose between the two studies. The first study does indicate that slightly more than half of marriages end in divorce, but this means that slightly less than half of all marriages do not end in divorce. The second study indicates that even marriages that are at risk for divorce because of such factors as conflict and a family history of divorce do not necessarily end in divorce. The second study suggests that it may take additional effort to overcome these risk factors. For example, the couple considering divorce when they entered therapy was willing to spend a year in therapy to work on their marriage. The findings of these two studies illustrate the general idea that nomothetic and idiographic research can complement rather than compete with each other.
- 2 Your friend’s second question is an example of a general question that students of psychology often ask (and should ask): What does all this research evidence have to do with me? The findings of these two studies provide potentially useful information for your friend as she considers her future. The first study tells us that divorce does occur frequently and that certain factors have been identified as indicators of when divorce is more likely to occur. The second study tells us that marriages can succeed even when these risk factors are present. This information can be useful because it provides evidence from systematic and controlled study that complements

what we can learn from our own experience. Your friend will not be able to determine based on these findings whether she will, in fact, divorce should she choose to marry. More generally, the findings of psychological research cannot yet tell us the answer to Gordon Allport's question of what any one person will do.

Answer to Challenge Question 1

- A One source of bias in this case study was that the same individual served as therapist and as researcher with the commensurate problems of observer bias. A second source of bias is that the therapist based his conclusion solely on the self-reports of the patient.
- B The successful treatment may have resulted from the relaxation technique alone; the use of the "secret word" in the face of anxiety; attention the patient received from the therapist; or even the high cost of the treatment.
- C The major problem that arises from studying one individual is a potential lack of external validity.