Chapter 14. Experimental Designs: Single-Subject Designs and Time-series Designs

Introduction to Single-Subject Designs

Advantages and Limitations

Advantages of the single-subject approach

Limitations of the single-subject approach

Why Some Researchers Use the Single-Subject Method

Procedures for the Single-Subject Design

Establishing a baseline

Optimal baseline

Baselines to avoid

Analysis of treatment effects

AB and ABA designs

ABAB design

Intra-participant replication

Inter-participant replication

Reversible and irreversible behavior

Multiple baseline procedures

Time-series designs

Case Analysis

General Summary

Detailed Summary

Key Terms

Review Questions/Exercises

Introduction to Single-Subject Designs

A three-year old boy diagnosed with autism shows characteristic language deficits. His level of spontaneous speech is equivalent to what is expected of a boy less than two years old. Monica Bellon, Billy Ogletree, and William Harn (2000) conduct a study to increase the level of spontaneous speech in this young boy. They begin by recording the boy's normal level of spontaneous speech during four 45-minute sessions in which an adult reads storybooks to the child and periodically asks questions. During the next phase (treatment phase) that consists of eight 45-minute sessions, the adult again reads storybooks but also uses a technique called scaffolding. The scaffolding procedure includes pauses to allow the child to provide information, choices posed to the child, elaborations of the story by the adult, and questions asked of the child. The final phase consists of two 45-minute sessions that were identical to the baseline phase. Results show that spontaneous speech was relatively low and stable during the baseline phase, increased during the treatment phase, and remained elevated during the final phase. The authors concluded that repeated storybook reading with adult scaffolding effectively increased spontaneous speech in an autistic boy.

The above example illustrates the single-subject approach. It is a method designed to study the behavior of individual organisms. As the method continues to evolve and improve, it also has become more popular for both scientific and therapeutic purposes. Its track record in both areas is impressive. The single-subject approach should not be confused with the case-study or case-history approach where a single individual is also studied exhaustively. The case-study approach is often an uncontrolled inquiry into history (retrospective) and it may yield interesting information. However, the lack of control severely limits any conclusions that can be drawn. There are two serious problems with the case-study approach: (1) lack of experimental control, and (2) obtaining precise measures of behavior. Neither of these problems applies to the single-subject approach.

The method is relatively popular today but it hasn't always been. Research in psychology started out using small numbers of participants, and investigators relied heavily on their ability to control conditions so that the conditions were reasonably constant among participants. Rigorous methodology was only beginning to evolve. After the data were gathered, conclusions about effects of the independent variable were based on subjective visual inspection of the data. Groups were not formed randomly and objective statistical analyses for decision-making were not yet available. Investigators realized the shortcomings of their method and made attempts to minimize subjectivity in their analyses.

The introduction of random assignment and statistical analyses were tremendous advances for research. Random assignment enhanced the likelihood that groups were initially equal on all variables. Statistical procedures permitted researchers to decide objectively whether the observed effect was more likely a chance occurrence or an outcome of the treatment condition. Investigators readily accepted these

powerful research tools, and large sample statistical studies rapidly became popular. As interest in large sample methods increased, it became difficult to publish nonstatistical research or even studies based on a small number of participants. Some researchers strongly preferred the single-subject approach refined by B. F. Skinner and elaborated by others. They continued using and refining it. Controversies and arguments frequently erupted between researchers using the single-subject approach and those using a statistical one. It is ironic that, even though psychology was defined as the study of individual behavior, investigators studying individual behavior could not easily get their research published in the established journals. This was the case even though strong behavioral control by the treatment condition was shown repeatedly in individual participants. It was this difficulty in getting their research published that led to the formation of the Society for the Experimental Analysis of Behavior and the subsequent establishment of the journal entitled *Journal of the Experimental Analysis of Behavior.* The journal publishes basic research involving the study of individual participants. Subsequently, a second journal devoted to the study of individual participants was established focusing on applied research and entitled *Journal of Applied Behavior Analysis.*

With the passage of time, both the large sample and single-subject procedures have become better developed and their strengths and weaknesses more apparent. These methods continue to evolve, as do other research methods. Because of this, a greater variety of useful tools are becoming available to those interested in either basic or applied research.

Using the single-subject approach does not mean that you must investigate only a single participant, although you can. More often than not, several participants are studied very intensively, usually somewhere between three and five. However, in each case interest is always in the careful analysis of the individual participant separately and not in the average performance of the group. With the single-subject approach there is very little interest in averaging across participants and great emphasis is placed on careful and rigorous experimental control. Unwanted environmental variables are either excluded from the study or they are held constant so that their effects are the same across participants and conditions. As we shall see, important features of this procedure for determining the reliability of the findings are actual **replications** rather than inferential statistics. We shall describe two types of replication. These are intra-participant replication (replications within an individual participant) and inter-participant replication (replications between individual participants). As with other research methods, the single-subject approach has both advantages and limitations.

Advantages and Limitations

Advantages of the single-subject approach

Those who use the single-subject approach find it both a powerful and satisfying research method. One reason for this is that the method provides feedback quickly to the investigator about the effects of the treatment conditions. The experimenter knows relatively soon whether the treatment is working or not working. Day-to-day changes can be observed first hand, quickly and in individual participants. If changes are necessary on a day-to-day basis, they can be made. Seldom do scientists have available procedures that do this. In contrast to the single-subject approach, a large sample statistical approach may take weeks or months of testing participants, calculating means, then performing statistical analyses, etc., and unfortunately, often nothing may be known about the effects of the treatment conditions until the final statistical analysis is complete. Even then, as we have seen, the derived knowledge is limited to statements regarding group performance and not to the performance of specific individual participants.

The single-subject method also allows us to draw strong conclusions regarding the factors controlling the dependent variable, yet the method does not use random assignment. The method allows strong conclusions because investigators employing it use procedures that provide rigorous control over environmental-experimental conditions with great emphasis on obtaining stable behavior with each participant. To be an acceptable scientific work, the research must demonstrate for each participant that behavior is controlled by the treatment condition and he or she must also show both intra- and inter-participant replication. That is, control must be shown both within a single participant and also between the participants.

Limitations of the single-subject approach

One obvious limitation of the single-subject approach is that the method is unsuitable for answering actuarial types of questions. Questions such as, "How many of the one-hundred people exposed to a particular treatment will respond favorably and how many will respond unfavorably?" A similar question relates to studies comparing two or more different treatments on the same behavioral measure. For example, which of the various treatments is the most effective? Ineffective? Debilitating? The method cannot be used if you are interested in treating an entire group of participants, such as a classroom, in an identical way on a daily basis, i.e., when changes in procedures are made, they are made for everyone in the group at the same time and for the same period. A different method is also required if "after the fact" studies (ex post facto, correlational, passive observational) are of interest. Moreover, the single-subject approach makes heavy time demands. It may, on occasion, take several months to completely test a single participant under the various conditions of interest. Often researchers are unwilling or unable to devote the required time. In addition to these limitations, there are also some recurring problems. Establishing a

criterion and acquiring stable baselines for the response of interest are sometimes very difficult. Further, determining whether variability in behavior is intrinsic or extrinsic can be troublesome. Nonreversible (irreversible) behavior poses its own set of problems and it precludes the use of a design in which the researcher removes the treatment to observe a return to baseline levels of responding. Failure to obtain intra- and inter-participant replication for whatever reason creates problems for the single-subject approach. Sometimes decisions regarding the necessary number of both intra- and inter-participant replications are largely subjective. Nevertheless, in spite of the limitations and problems described here, the single-subject method does provide researchers with another powerful way to assess behavior.

Why Some Researchers Use the Single-Subject Method

Investigators who use the single-subject method do so for different reasons. One of the main reasons is that their interest is in the behavior of individual participants. The large sample group approach places emphasis on group averages rather than individual participants. Unfortunately, the behavior reflected by the group average may not represent the individual participant. The following example illustrates how distant the overall results for the group may be from the performance of any given individual participant. Say that we are interested in learning as a function of practice. The particular form or shape of the curve is what we are trying to determine. We choose twenty participants to participate in our study, choose a learning task that we want to evaluate, and then give practice trials to the participants until the task is learned. After all the data are gathered, we plot a learning curve to determine its form or shape (see Figure 14.1), which in turn will reveal to us how quickly and smoothly participants learned the task. The learning curve in Figure 14.1 is based on the performance of all twenty participants. Each data point on the graph represents an average (five trials) of an average (twenty participants).
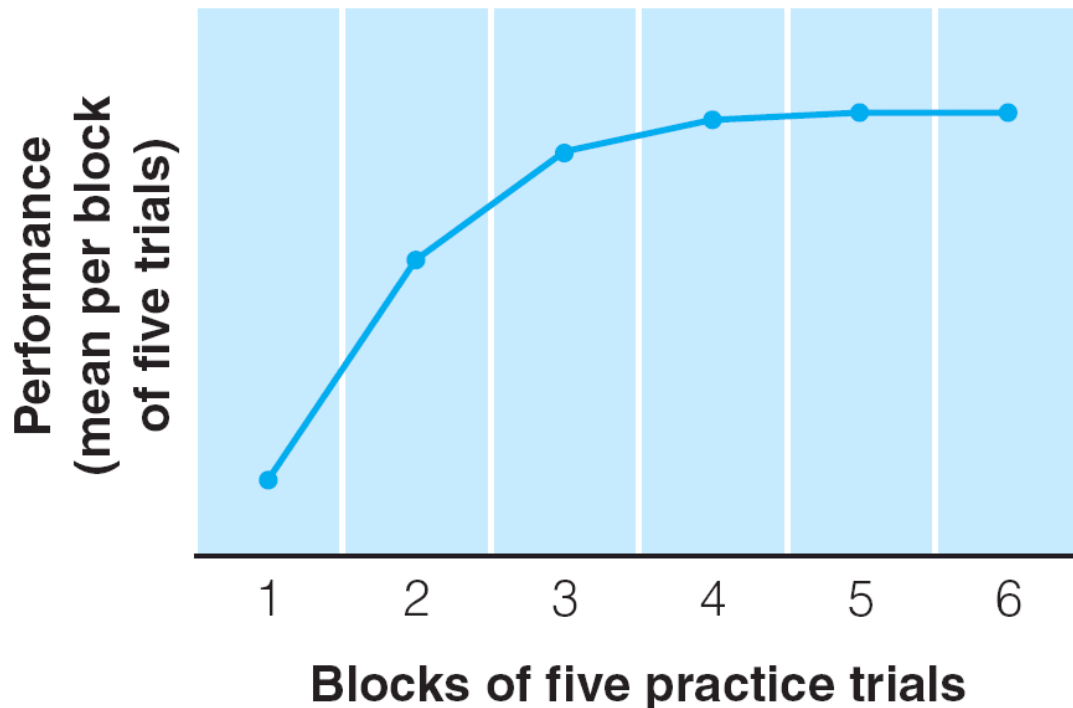
Fig. 14.1 Mean performance of twenty participants on each of six blocks of five practice trials

A description of how these averages were computed may be helpful. First the performance of each participant on each block of five trials was averaged. Then the average for each average block of five trials was obtained for all twenty participants. This average of averages produces a smooth, negatively accelerated learning curve. But does this group curve reflect the performance of a single individual? It is quite unlikely that any one individual in a group of twenty participants would perform like the group curve. In other words, plots of each individual participant may be different from the group curve. Usually a statistical approach that relies on the analysis of group means masks the performance of each participant, whatever the problem being studied. A related point follows.

A group performance curve may not only mask the performance of an individual but may also be misleading. Although the group average may indicate an increase in performance as a result of the treatment condition, not all participants may have increased; some individuals comprising the group may, in fact, perform at a lower than normal level. The point is that individual reactions to the experimental conditions are not taken into account. Failure to address individual reactions may be especially unfortunate in more applied research, particularly if assessing different therapeutic techniques. If the therapy is harmful (or helpful) to certain individuals, this fact may be lost in the group mean. Others have made a similar argument in terms of the statistical analysis.

The analysis may reveal statistically significant differences between group comparisons but the differences may be due to only a few participants. On the other hand, however, the analysis may not be statistically significant overall but some participants may change markedly as a result of the treatment conditions.

The dependence on statistical evaluation of the data with large sample methods is also a source of unhappiness for some researchers. Have the assumptions underlying the statistical test been satisfied? Is the sample size sufficiently large to give the needed power? Is the sample size too large so that trivial differences between group comparisons will be significant? What about Type I and Type II errors? Some researchers are concerned that investigators are placing greater concern on statistical issues per se and placing less concern on rigorous methodology. Statistical analyses cannot salvage a poor experiment. Complete confounding of variables cannot be corrected by statistical analysis.

Other researchers favor the single-subject method because, for some interests, large numbers of participants may not be available. Consequently, a large sample procedure cannot be used. In applied research dealing with specific behavioral problems, the researcher-therapist might have to wait months or years before obtaining a sufficiently large sample. Applied psychologists are often interested only in a small number of individuals. They need a method sufficiently flexible to allow treatment of individual cases, one that can be altered quickly to adjust to the responsiveness of the individual. Large sample statistical procedures do not have this flexibility.

Table 14.1 compares characteristics of both the single-subject approach and the large sample statistical approach.

Table 14.1	Comparison of Single-Subject and Large-Sample Approaches	
ISSUE	**SINGLE SUBJECT**	**LARGE SAMPLE**
Random assignment	Not applicable	Yes
Control groups	Not applicable	Yes
Manipulation	Yes	Yes
Determining reliability	Replications	Statistically
Determining generality	Replications	Sampling
Number of participants	Usually 1–5	Usually >10 per group
Flexibility of procedure	High	Low
Measurement of behavior	Continuous monitoring	Varies
Focus of interest	Individual	Group
Time to complete experiment	Relatively long term	Relatively short term
Knowledge of results	Moment to moment	After study is complete
Type I and Type II errors	Not applicable	Yes
Statistical assumptions	Not applicable	Yes
Statistical power	Not applicable	Yes
Laboratory experiments	Yes	Yes
Field experiments	Yes	Yes
Actuarial experiments	No	Yes
Comparison experiments	No	Yes

Procedures for the Single-Subject Method

As noted, when using the single-subject method the effects of the treatment must be shown in individual participants. To accomplish this the experimenter must have considerable control over the experimental situation at all stages of the research. Moreover, he or she must use the proper methodology. As with other research methods, the dependent variable must be clearly defined. Where possible, it should be defined in terms of operations that objectively identify the occurrence or nonoccurrence of the response. In single-subject research the dependent variable is often "rate of responding" and great emphasis is placed on steady state (stable) performance rather than behavior in transition, i.e., in the process of changing.

Establishing a Baseline

When assessing steady-state behavior in a given condition the behavior is assessed relative to some comparison point. With the single-subject approach, the comparison point is the **baseline** condition. To

establish a baseline, repeated observations of the natural frequency of the behavior of interest (dependent variable) are first made. In effect, you observe the frequency with which the behavior occurs before the treatment (independent variable) is introduced. This baseline serves as a sort of benchmark against which to ascertain whether the subsequent introduction of the treatment condition has an effect. The behavioral effect may be either an increase over baseline responding **(facilitation)** or a decrease under baseline responding **(suppression).**

Because the baseline serves as a point from which the treatment effects are judged, it is important that a stable baseline be established. There is no set number of days or experimental sessions that define baseline stability. Instead, a criterion of stability is established such as "four experimental sessions in which the frequency of the target behavior does not vary by more than 5 percent." In other instances a less demanding criterion of l0 percent may be used. Some participants may take only four days to meet the criterion, while others may take a week or more before the session-to-session variability is less than 5 or l0 percent. The choice between 5 percent or l0 percent is somewhat arbitrary but these values are often used. If baseline behavior is so variable that a 5 or l0 percent criterion of stability cannot be met, then the investigator should strive to acquire greater control over all variables related to the experimental situation. This can be a very difficult task. What is needed is a careful assessment of all aspects of the experiment for possible sources of unwanted variability. This would include assessing the instructions, procedure, apparatus, independent variable, dependent variable, and any other possibilities. It is wiser to assume that the reason for the variability is extrinsic (environmentally induced) and then seek ways to reduce it, rather than to assume that the variability is intrinsic (inherent) and cannot be reduced. If all efforts to reduce variability fail, then the percentage criterion under baseline conditions, e.g., 5 or 10 percent, may have to change upward. Some criterion is necessary to avoid arbitrary decision-making.

Optimal Baseline. An optimal baseline requirement for any given response is that it be stable, i.e., there is little change in frequency from session to session under natural (baseline) conditions. In addition, if the treatment is expected to lead to increases in frequency of responding, then baseline responding should not be so high that further increases would be difficult to obtain (ceiling effect). On the other side of the coin, what if the treatment is expected to lead to decreases in frequency of responding? Now the opposite is true. Baseline responding should not be so low that further decreases would be difficult to achieve (floor effects). In some situations the experimenter may be interested in demonstrating both increases and decreases in responding but at different phases of the experiment. If this is the case, then a baseline level that permits both increases and decreases in responding would be necessary. Such a baseline level is shown in Figure 14.2.
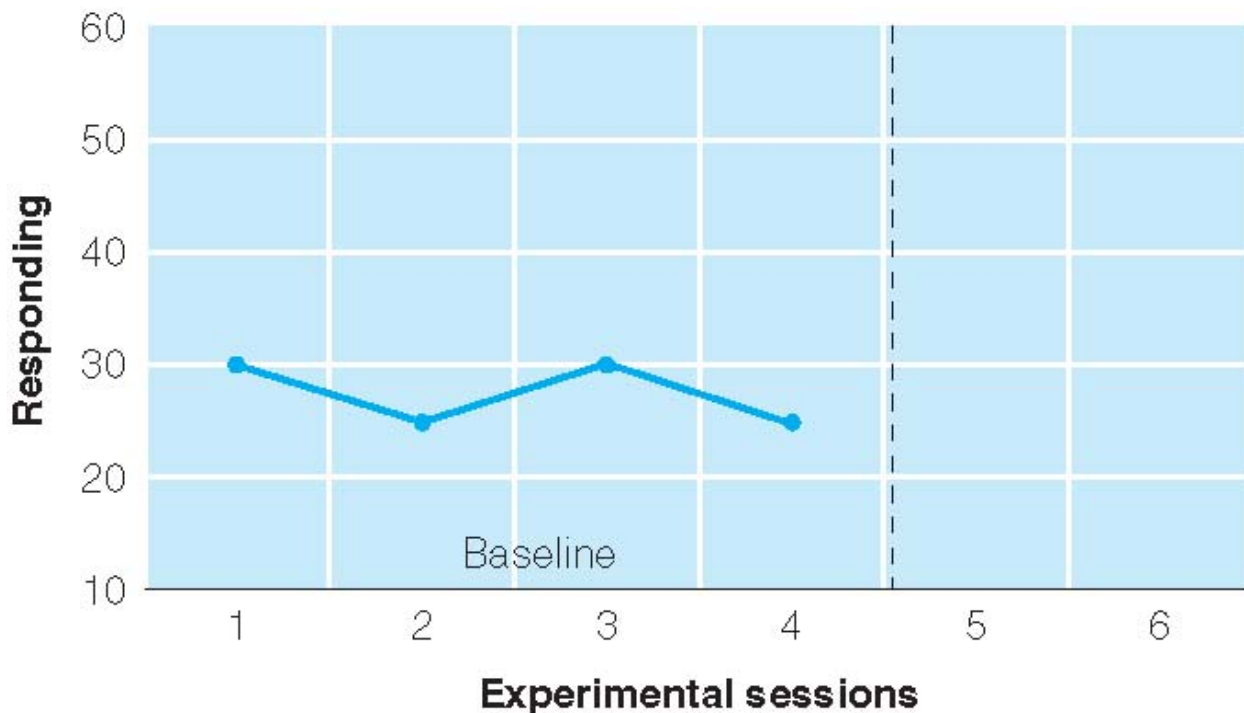
Fig. 14.2 Baseline when treatment is expected to lead to both increases and decreases in responding at different phases of the experiment.

After the treatment condition is introduced, departures from the baseline, either upward or downward, can be easily observed. If the frequency of responding neither increased nor decreased nor changed in terms of session-to-session variability, then our independent variable (treatment condition) obviously had no measurable effect.

Recall an earlier chapter in which we discussed the use of different degrees of an independent variable for purposes of identifying a function or trend. We saw that a minimum of three different points or values was needed. A similar requirement is necessary when establishing a baseline across sessions; never less than three sessions should be devoted to establishing a stable baseline since it is not possible to identify a stable pattern with less than three sessions. Reasons for this will become more apparent as we describe different possible baseline conditions.

Baselines to Avoid. There are several types of baselines that should be avoided simply because they evidence trends that make it difficult to interpret the effects of the treatment condition. For example, if you were evaluating the effects of praise on the amount of time spent studying, the baseline depicted in Figure 14.3 would not be appropriate. It would be difficult to assess whether obtaining an increase in study time on the fifth session when praise was introduced was a result of the treatment (praise) or a result of continued increases in study time under the baseline condition.
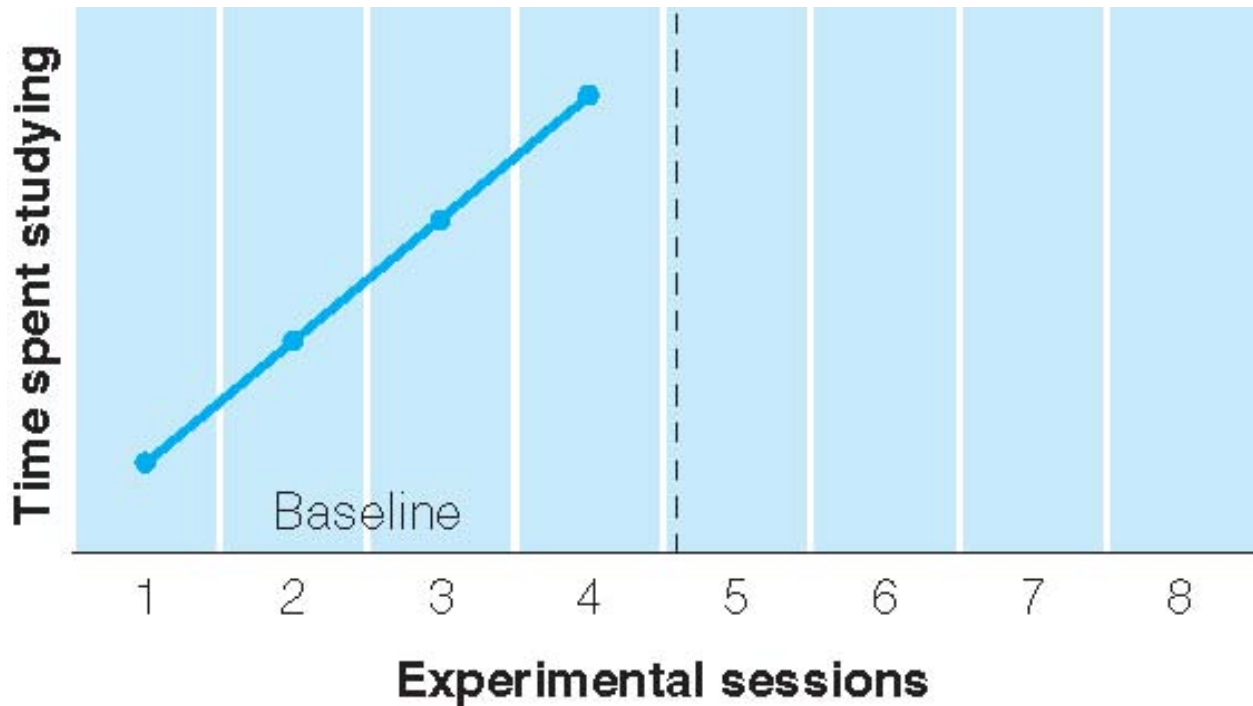
Fig. 14.3 An inappropriate baseline to use in a single subject design when evaluating a condition that is expected to lead to increases in the dependent variable.

Imposing a treatment on a steadily increasing baseline should be avoided where it is possible. Similarly, the effects of an independent variable may be difficult to interpret with a baseline that continues to decrease, and the effects of the treatment are also expected to lead to a decrease in performance. For example, if we were interested in assessing the effects of punishment on disruptive classroom behavior we would not want to use a baseline as shown in Figure 14.4. Further decreases at session 5 and beyond may be a result of the natural downward trend, a result of punishment, or both factors. In fact, with a baseline either increasing throughout or decreasing throughout, any change or no change in the pattern would be difficult to assess. The soundest procedure would be for the researcher to continue baseline measurement until it leveled off and reached a rigorous criterion of stability. If the measure fails to reach the stability criterion, then we should attempt to achieve greater control over the conditions or find a different measure. Additional options are available to experienced researchers (Sidman, 1960).
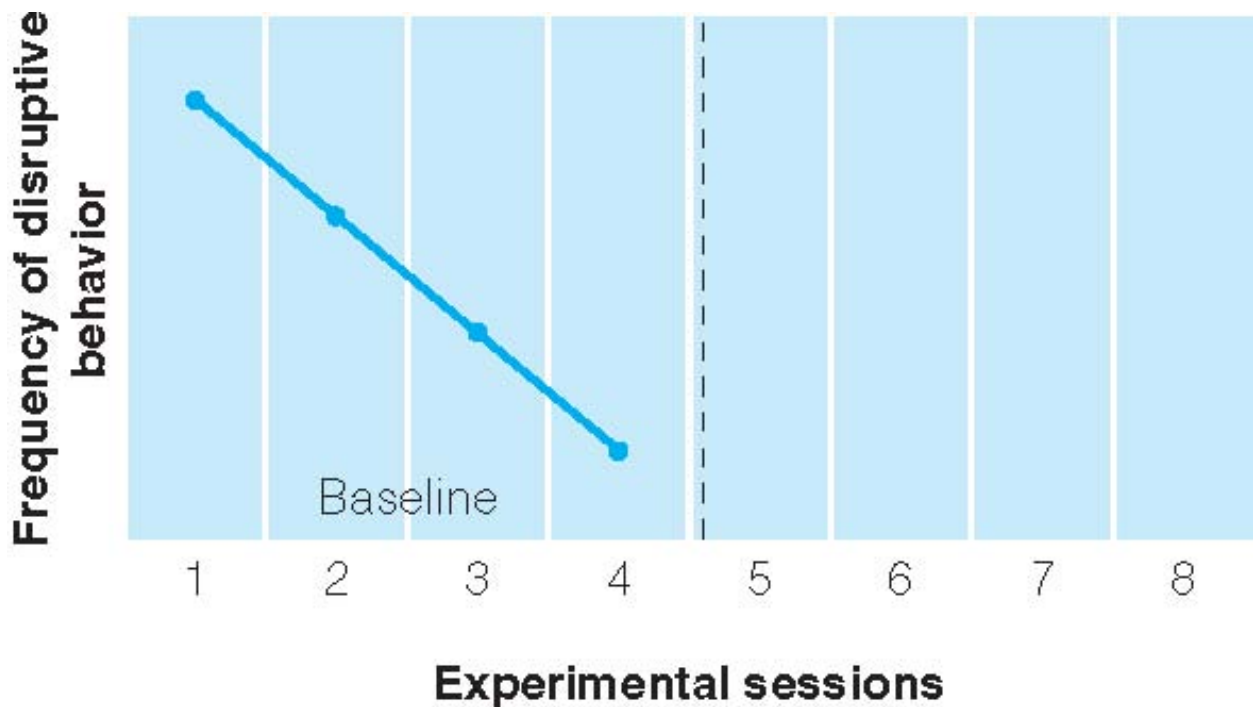
Fig. 14.4 An inappropriate baseline to use when evaluating a condition that is expected to lead to decreases in the dependent variable.

Finally, if marked variability in responding occurs from one experimental session to the next, it is difficult to interpret any effect that the treatment might have. Figure 14.5 depicts such a pattern. In basic laboratory research, a baseline pattern of this type is of little use. The investigator should make an effort to reduce the variability by eliminating sources of extrinsic (environmental) variability. If unsuccessful in doing so, a different response measure should be considered. At times, simply extending the period across more sessions results in a more stable baseline. However, most investigators would suggest that a careful, systematic assessment of the experimental situation be undertaken to identify sources of variability and then remove or alter them. Again, this means assessing the procedure, apparatus, task, instructions, experimenter, etc.
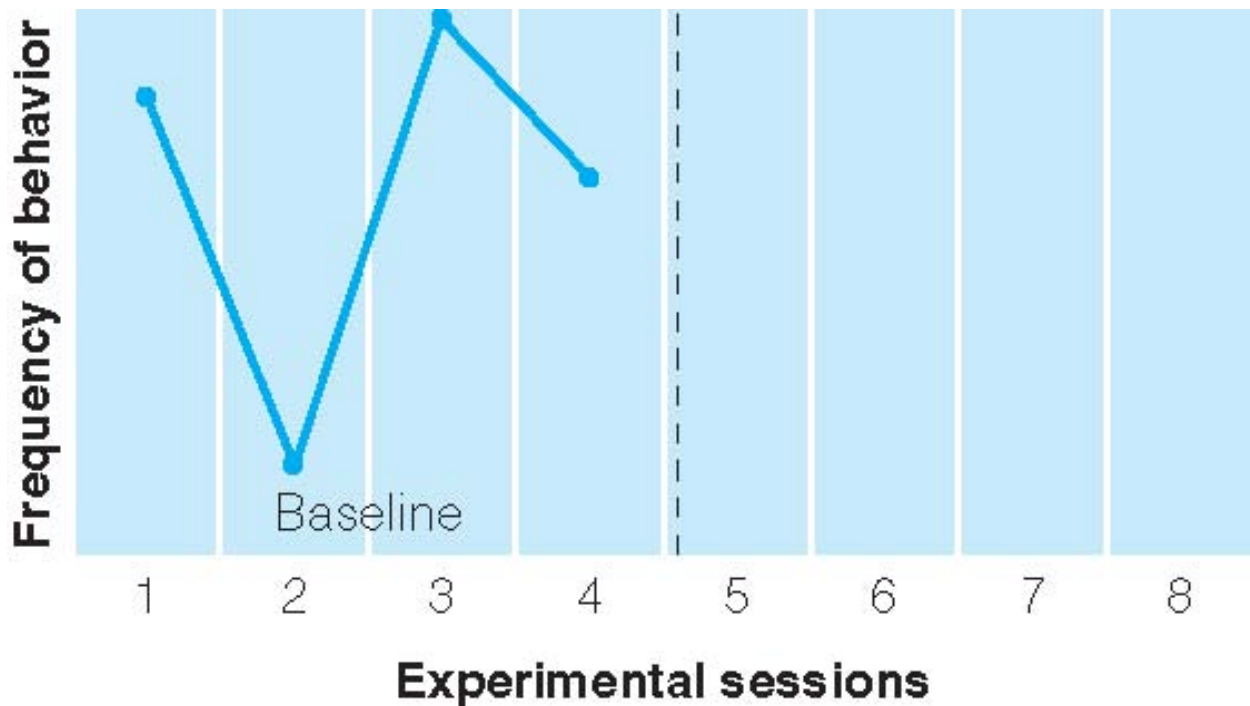
Fig. 14.5 An inappropriate baseline to use when evaluating conditions that are expected to lead to either increases or decreases in the dependent variable.

In applied areas, such as evaluation of therapeutic techniques, efforts to obtain a stable baseline may be less successful and the investigator, after an exhaustive search for solutions, may have to impose a treatment condition over an unstable baseline. If the effects of the treatment are strong, then they may be seen in terms of both greater stability and a higher (or lower) frequency of responding.

We have not exhausted the different kinds of difficult baselines that are encountered when doing research but we have described the more bothersome ones. The issue of what constitutes an acceptable baseline is a complex one that we have tried to simplify. We will now discuss the treatment phase of research.

Analysis of Treatment Effects

The analysis of treatment effects will be more understandable to you if we give an overview of the design strategy. It is customary to refer to the baseline phase of an experiment as the "A" condition and the treatment phase as the "B" condition. If there are different kinds of treatment conditions, then the others are referred to as "C," "D," etc.

AB and ABA designs. The weakest design in terms of drawing conclusions and ruling out alternative interpretations is the AB design. This design does not permit the systematic assessment of the treatment condition. There are problems with a single presentation of the baseline and treatment condition (i.e., AB).

For example, what would be the natural course of the behavior across the same time period if the treatment had not been presented? It is similar to conducting an experiment without using a nontreatment control group. Without a control group, we cannot be sure that the behavior was altered by the treatment condition, rather than by some extraneous condition. The same is true of the AB design. It is possible that changes in behavior during the treatment phase result from some unknown environmental event not related to the treatment. The AB design does not permit ruling out this alternative hypothesis. It is sometimes tempting to accept the results of an AB design and conclude that the treatment had an effect when low levels of baseline behavior (A) are followed by sudden dramatic increases with the introduction of the treatment (B). But to do so would be inappropriate, since proper control procedures were not present. Nevertheless, results of this kind would certainly be very encouraging and should be pursued further, but with a more powerful design. One such design is the ABA procedure. The AB design should be used only under circumstances that do not permit a more adequate method. These instances are more common in applied settings.

The ABA design is a far more powerful design than the AB design simply because the treatment condition is *introduced* for a period of time and then *withdrawn.* There are two opportunities to assess whether the treatment condition is effective—introducing it and withdrawing it. If behavior shows a systematic change, then your confidence is increased that the treatment, rather than some unknown environmental event, is the reason for the behavioral change. It is quite unlikely that natural conditions would increase and then decrease behavior as it did when the treatment was presented and then withdrawn. Showing the same or similar relationships in other participants would further strengthen your confidence that the treatment was responsible.

The ABA design is generally criticized on two counts. One is that replication of the effect within a participant is not shown. The importance of this type of replication will be described in more detail below. The second problem relates to the applied setting where behavior modification is considered desirable. If the treatment (e.g., therapy) is effective in modifying behavior, then it is desirable to end the investigation on a treatment phase rather than a baseline phase.

ABAB design. The most powerful design strategy (best method for assessing treatment effects) that we will discuss is the ABAB design. The ABAB design is a shorthand way for stating that we first determine a baseline (A), then we introduce the treatment for the first time (B). After the criterion of ability is achieved we then withdraw the treatment and reintroduce the baseline condition (A). Finally, after baseline stability is reestablished, we present the treatment condition (B) for the second time. This ABAB design, when used, is a very powerful design that allows the researcher to make strong conclusions regarding the treatment effects. With this design the researcher demonstrates the degree of control over

behavior in two ways—first by *introducing* the treatment condition, then by *removing* it. Again, we will repeat the procedure. After the baseline is established (A), the treatment condition (B) is *introduced* and the extent to which the treatment influences behavior (the extent to which behavior departs from baseline) is assessed. Then, following stable performance, the treatment condition is *removed* (baseline condition (A) again presented). Performance should then return to the original baseline. The final phase requires that we again present the treatment condition (B) and end the experiment with it. We will now give an example of an ABAB design strategy.

Over the years, researchers have been interested in whether participants prefer predictable over unpredictable painful events. Many used the single-subject method with a sample of 3 or 4 participants. It is interesting to note that the studies used very similar procedures even though different species were involved, e.g., fish, birds, rats, humans. The initial studies in this area used rats as participants, a brief electric shock as the mildly painful stimulus, and a tone to signal if shock was to occur. Researchers first exposed the animals to predictable shock (a five-second tone signaled when a .5 second shock occurred) and to unpredictable shock (unsignaled shock) to acquaint them with the conditions and to make sure that they had equal experience with both. (The number of shocks was the same whether predictable or unpredictable. The only difference was that a signal preceded one condition but not the other.)

During this initial exposure to the two conditions, participants could not alter (change) the condition from one to the other. However, their responses on a response lever were recorded, even though responses on this lever had no effect at all. This period served as a baseline period (A) to measure how frequently they pressed the lever when there were no consequences. Responses on the lever occurred but were low in frequency during the baseline phase. After four days of being exposed to both signaled and unsignaled shock and with baseline responding stable, animals were given a choice between the signaled and unsignaled conditions. During this choice phase (treatment phase), the response lever was functional and responses now changed the conditions from one to the other. Animals at this time were placed in the unsignaled condition but if the lever was pressed the condition changed. A response on the lever changed the condition to the signaled one for a period of one minute. At the end of this one-minute period, the condition automatically changed back to the unsignaled condition and remained there unless another lever response was made. If the predictable (signaled) condition was reinforcing (preferred), response rate should increase over baseline; if it was punishing (not preferred), response rate should decrease. After choice behavior stabilized and preference was determined, the baseline condition was reinstated. This was followed by another treatment condition (preference testing). The results of the experiment were similar to those shown in Figure 14.6.
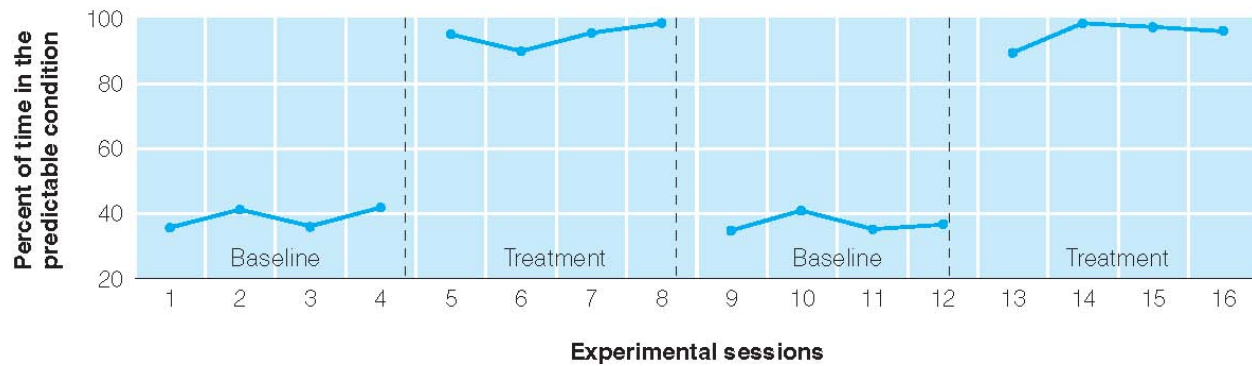
Fig. 14.6 Single-subject ABAB design in which the participant could choose between predictable (signaled) or unpredictable (unsignaled) shock. The results would be similar whether percent of time or number of lever presses were used as the dependent variable.

During the baseline conditions (A) participants lever-pressed at a rate sufficient to remain in the predictable shock schedule (had the levers been effective) only about 20 percent of the time. When the treatment condition (B) was introduced, participants changed from the unpredictable schedule at a rate sufficient to spend 90 percent of the time in the predictable condition. When the treatment condition was withdrawn and the baseline condition reinstated (session 9), responding again returned to a low level. This showed that withdrawing the treatment *reversed* performance from high to low responding. Finally, when the treatment condition was introduced for the second time (session 13), responding on the levers increased to a high level. Data such as this demonstrate convincingly, without the need for a statistical analysis, that the treatment condition is systematically controlling behavior.

Let's apply the ABAB design to our question regarding the effect of TV violence on aggressive behavior in children. It should not be too difficult for you to imagine how such a single-subject design could be implemented. First, a child is selected for the study. Typically, the participant is someone who is readily available to the researcher and has the characteristics of interest (e.g., particular age). Then a baseline level of aggressive behavior is established during a week in which the child does not watch TV programs that contain violence. All of the issues regarding observation and measurement that have been discussed in previous chapters must be considered to develop a quality protocol for recording the dependent variable (level of aggressive behavior). After the one-week baseline, the treatment is imposed on the second week. During this second week, the participant is exposed to TV programs with violence and aggressive behavior continues to be recorded in the same manner as the previous week. The third and fourth weeks are replications of the first and second weeks. That is, the third week involves TV programs without violence and the fourth week involves TV programs with violence. Remember that measurement of the participant's aggression (dependent variable) remains consistent throughout the experiment.

Some products advertise their effectiveness by pointing to "single-subject research" that consists of testimonials from individuals. Let's examine the information in Box 14.1 regarding claims that slippers can help you lose weight.

Box 14.1 Thinking Critically About Everyday Information – Diet slippers that help you to lose weight

A Japanese company sells diet slippers that are designed to help a person lose weight. In support of their product, the company provides testimonials from individuals who have tried the slippers. Some of the testimonials include the following:

"Testimonial No. 1
I always wear diet slippers. It's been more than two years since I tried these slippers for the first time. I have lost weight. Also minor health problems that I had are gone. I feel great every day. I want to share the benefits with lot of people. Therefore, I encourage them to try Diet Slippers. They are very happy with the results.

Testimonial No. 2
Hello, I'm a great fan of diet slippers. It's been almost one and half years since I started wearing these. I lost about 5 lb. Before I wore the Diet Slippers, I could not afford to take three meals a day because the fear of gaining weight. Now I don't have to worry about it. It took a little while for me to get used to the Diet Slippers. First I felt a little tired after the first use. But now, I feel totally comfortable in them and can't go without them even one day. I thank you for your wonderful creation.

Testimonial No. 3
Thanks to Diet Slippers, I have lost 9 lb.

Testimonial No. 4
My mother-in-law is quite impressed with Diet Slippers, because without causing any negative effect to her health, she was able to lose 5 lb. I thank you on behalf of my mother-in-law."

Consider the following questions:
1. How are these testimonials similar to single-subject designs?
2. Do the testimonials provide evidence of stable baselines?
3. Do the testimonials provide evidence of intra-participant replication of the effect?
4. What might be an alternative explanation for the reported weight loss?

Retrieved June 10, 2003 online at:
http://www.myshaldan.com/testimo.htm

Intra-Participant Replication

The preceding study regarding TV violence exposed each participant twice to the baseline and treatment condition. When the conditions are repeated with the *same* participant, we are using **intra-participant replication.** This is an important part of the single subject method. As we have noted, the primary interest among psychologists is focused on the behavior of individual organisms. Intra-participant replication focuses on the individual participant and identifies the factors affecting the participant. Systematic behavioral changes can be observed in individual participants by introducing and withdrawing the treatment condition. Intra-participant replication, then, demonstrates that our method is reliable, that the treatment effect is real, and that we have control over behavior.

The decision on the number of intra-participant replications that are necessary is sometimes difficult and may vary from experiment to experiment. Often a single replication is enough, i.e., ABAB. The number of intra-participant replications decided upon may vary according to the size of the treatment effects, the stability of the behavior, whether inter-participant replication is obtained, whether inter-species replications exist, whether similar related findings exist, and how well the present findings fit in with established findings.

A word should be said about the size of the effect. Small but consistent treatment effects combined with stable individual baselines can be important. Such effects, even though small, indicate experimental control over behavior. Perhaps with more effort and exploration, the conditions leading to a larger effect will be discovered.

Inter-Participant Replication

We have seen that it is possible to demonstrate repeatedly consistent behavioral changes as a function of the treatment in an individual participant. It is also possible to demonstrate the same effect consistently in other participants. This **inter-participant replication** establishes the generality of the findings, i.e., showing the effect occurs in more than one research participant. There are no hard and fast rules on the number of participants for which inter-participant replication must be shown. Much of the published research involves three to five participants per experiment. However, it is not unusual to find either fewer or more participants in different experiments. In addition to demonstrating that your findings can be generalized to other participants, inter-participant replication also demonstrates that the researcher has identified the controlling factors sufficiently to permit replication to other participants. On occasion, however, inter-participant replication is unsuccessful. When this occurs, additional detective work is usually necessary. It may be that greater control over the experimental situation is necessary. It is also possible that, because of individual differences, some participants react less to a given treatment. If the treatment were increased slightly (intensity, duration, frequency, etc.), inter-participant replication might be successful. This ability to treat individual participants in a flexible manner is one of the great strengths of the single subject approach.

Reversible and Irreversible Behavior

When intra-participant replication is achieved, we have demonstrated that baseline responding under the nontreatment condition can be recovered again once the treatment condition is withdrawn. We refer to the behavior as being **reversible behavior.** Without this feature, intra-participant replication is not possible. In contrast, **irreversible behavior** refers to those occasions when the original baseline cannot be recovered after the treatment has been withdrawn. The baseline level of responding remains at the same level as it was under the treatment condition. Many critics of the single-subject approach argue that this is

one of its weaknesses and that the method cannot be used when baseline responding is not recoverable. However, supporters argue that irreversible behavior may not be due to uncontrollable intrinsic factors (factors within the participant) but, instead, may be due to extrinsic controlling factors (experimental-environmental factors). They argue that a careful assessment of the situation and thoughtful changes based upon this assessment, more often than not, will produce reversible behavior. Often it is achieved only after a number of attempts. However, they do not argue that all behavior is reversible.

There may be situations in which the behavior is not reversible. For example, drug research may encounter carry-over effects in which the effects of the drug last longer than anticipated and continue to affect behavior after the treatment (drug) is thought to be removed. The solution to this problem is to allow enough time for the drug to dissipate. A more difficult problem may arise when a single-subject design involves experimentally-induced brain lesions in animal research. If the tissue damage is permanent, the treatment cannot be withdrawn, thus precluding the ABAB design. To overcome this difficulty, researchers have used pharmacological agents that only temporarily affect brain functions instead of producing permanent impairment. There are yet other instances where behavior appears irreversible because of learning factors. The example of spontaneous speech in an autistic boy at the beginning of this chapter is one example. Consider another example. If a number of problems have the same solution and the solution is learned through a reinforcement procedure, you can withdraw the reinforcement but the solution to the problem will remain. Conceivably, then, researchers on occasion will encounter behavior that does not reverse (return to baseline levels) when the treatment condition is withdrawn. Given this state, the powerful ABAB design strategy cannot be used. For this reason, and others, there is an alternative procedure, referred to as the multiple baseline procedure. The latter procedure is also useful in therapeutic situations where withdrawing a therapeutic treatment for purposes of identifying the controlling factor may be undesirable or unethical.

Multiple Baseline Procedures

As we noted, the ABAB design is not appropriate when we are unable to recover our baseline level of performance or when the withdrawal of treatment poses an ethical dilemma. Under these circumstances, a different design strategy such as the **multiple baseline procedure** is necessary. In effect, the multiple base-line procedure allows the researcher to perform intra-participant replication with *different responses* rather than the same response.

The multiple baseline procedure requires that baselines be provided for several different responses and that these responses be independent of each other. To say that the responses are independent implies that increases or decreases in frequency of one response do not affect (lead to increases or decreases) in the frequency of other responses. (See Figure 14.7.) Usually, baselines for three or more different responses

are needed when the multiple baseline procedure is used. Baseline responding is established in the same manner as described previously except that baselines for several responses are plotted at the same time. Then the different responses are treated one at a time.
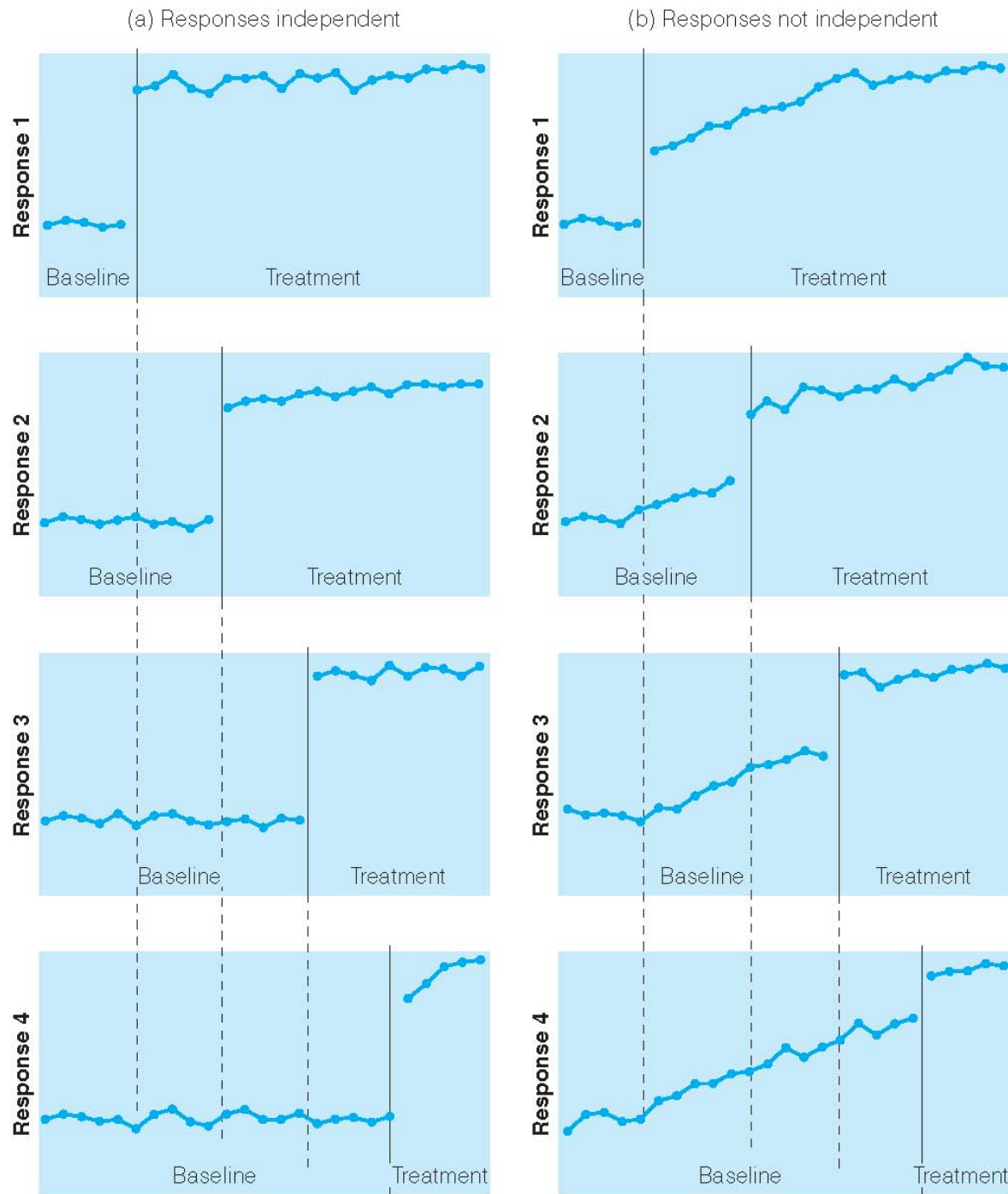

Fig. 14.7 Illustration of multiple baseline procedures when responses are independent (a) and not independent (b). Note that when experimental treatments are introduced in (a),

the baselines of the remaining responses are unaffected. In (b), the introduction of each

treatment produces a change in the baseline levels of the remaining responses (hypothetical data).

To illustrate, after stable baseline levels are established, the treatment condition is applied to only one of the responses (target response). The other responses are not treated and remain under the baseline condition. Changes in the target response are recorded to assess the treatment effects but the other responses, those not receiving any treatment, continue to be monitored in the baseline condition. After the target response stabilizes to the treatment condition, the experimenter then applies the treatment to the second response until it stabilizes, then the third response is treated. Since a withdrawal phase is not used with this procedure, the treatment of each response following the establishment of the baseline is essentially an AB design. As with any of the single-subject designs, the effectiveness of the treatment condition is assessed by a change in behavior relative to the baseline level. In the case of the multiple baseline procedure, the treatment effects are assessed by comparing the response receiving the treatment with its no-treatment baseline and also with the baseline of the untreated responses. The latter would be meaningless if the responses were not independent. Although the multiple baseline procedure is not as effective as ABA or the ABAB design, it does demonstrate replication of the treatment condition *across responses.*

Let's consider an actual study that is not recent but is nonetheless interesting (cited by Leitenberg, 1973). Two investigators used punishment to deal with a case of transvestism. A male patient reported becoming sexually excited by dressing as a woman and was apparently unhappy about his feelings. Although, at the age of twenty-one, the patient had never had girlfriends, he wanted to have a normal heterosexual relationship. Aversion therapy was used twice daily. It consisted of shocks to his arm or leg while he was either dressing as a woman (cross-dressing) or while thinking (fantasizing) about dressing as one. In the latter instance, the participant signaled the investigator when fantasies occurred. Shocks were withheld if the participant discarded the female garments. One response measure was sexual arousal, as determined by the circumference of the penis, to female pajamas, panties, slips, skirts, or a slide of a nude woman. Latency of the response was also recorded. After baseline responses to these garments were determined, the treatment of shock was introduced. The investigators found that, after the response of putting on panties was shocked a number of times, penile erection to this stimulus was suppressed. However, erections to the other garments continued to occur. Then responses of dressing up in the other garments were shocked until erections ceased to each. It was shown by the investigators that while erection no longer occurred to the female clothing, it continued to occur to the nude slide.

The multiple baseline procedure is also illustrated in an interesting study in which investigators used punishment to successfully modify undesirable mealtime behavior among sixteen severely retarded males (Barton, Guess, Garcia, & Baer, 1970). The undesirable behavior included *stealing* food from others, *eating with fingers* rather than utensils, *pigging* (eating spilled food from the floor or lapping food directly from the tray), and *making a mess,* such as spilling or dropping food. The study ran nearly four months. Punishment consisted of removal (time-out) from the dining area whenever the undesired behavior occurred. In some instances, removal was for the entire meal; in other instances, for a shorter period. The investigators started by recording baseline frequencies for each behavior (stealing, pigging, etc.). The first response had only a six-day baseline before the punishment (treatment time-out) was introduced, e.g., stealing. While punishment was being given for stealing behavior, baseline recording continued for the other responses. When stealing stabilized, then punishment was introduced for the second response while baseline recording continued for the remaining two responses. This procedure continued until all responses were under the punishment procedure. For the most part, but not entirely, independence among the responses was observed, i.e., punishing one response affected primarily only that response and not the others. The efforts of the researchers were successful: The undesirable mealtime behavior greatly improved as did the morale of the workers. An idealized depiction of the study appears in Figure 14.8. The figure also clearly outlines the procedure.
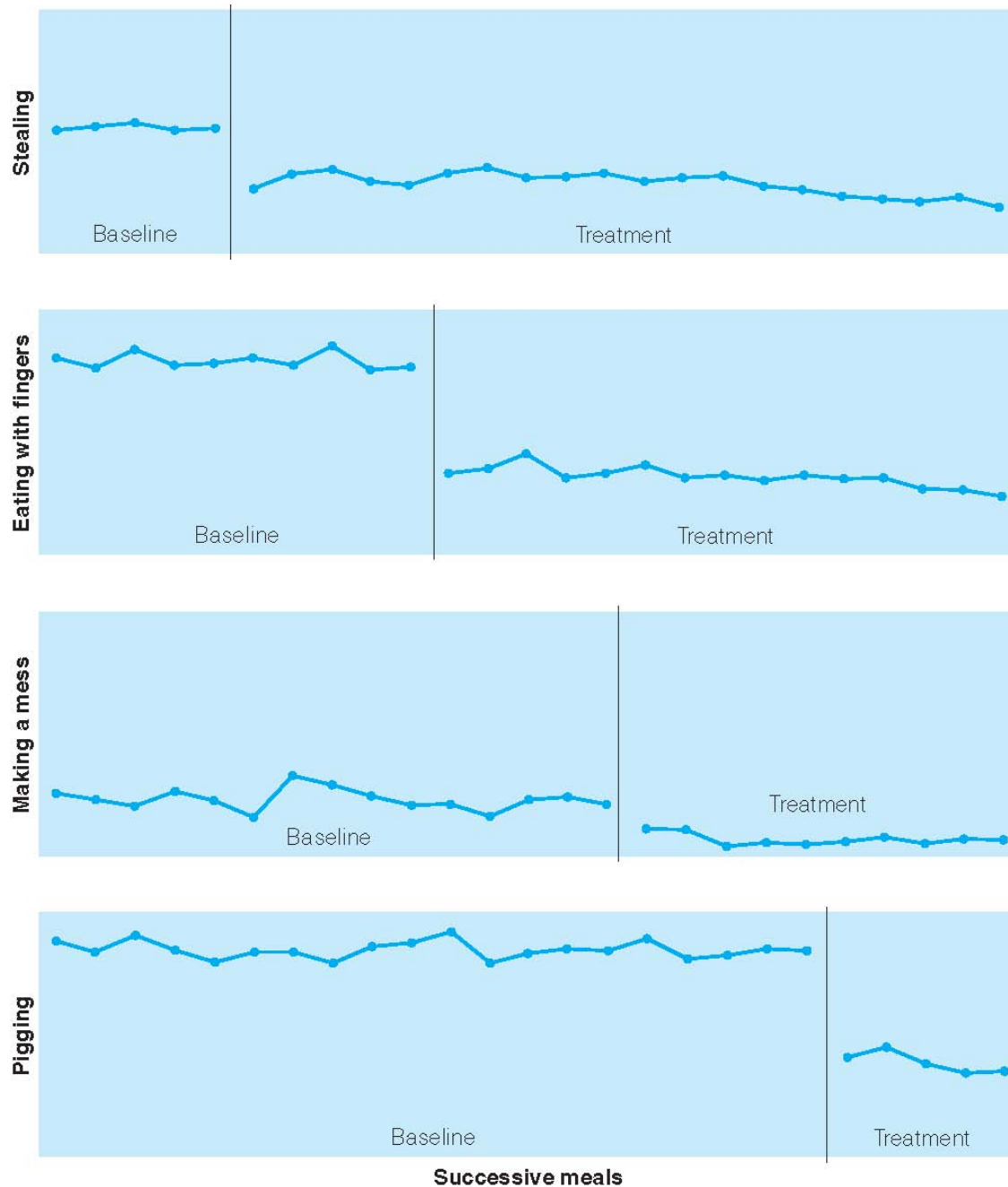
Fig. 14.8 An idealized graphic representation of a successful effort to suppress undesired behavior by the use of punishment using the multiple baseline procedure. Note that the treatment effects appear to be relatively independent and that all four undesirable responses were successfully suppressed.

We want to repeat, before leaving this topic, that the multiple baseline procedure is used either because the baseline level of responding cannot be recovered when the treatment condition is withdrawn, or withdrawal of the treatment may have an adverse effect on the participant, especially in a therapeutic

setting. The multiple baseline procedure is not as powerful as the ABAB design where both intra- and inter-participant replication can be shown by both the introduction and withdrawal of the treatment. However, as we have shown, replication across responses within the same participant can be shown with this procedure.

Time-series Designs

In many ways the single-subject approach is similar to a time series analysis in that the stability and changes in behavior are studied across time or experimental sessions. Time series analysis is characterized by repeated measurements of the dependent variable over time with an introduction of the independent variable at a particular point in time. Trends or patterns of behavior are observed both before and after introduction of the independent variable. Consistent with the theme of this chapter, the time series analysis can be conducted with more than one participant but data analysis is typically focused on individual participants.

Because time series analysis is characterized by relatively long-term measurements of some dependent variable, you must be careful to consider extraneous variables often associated with repeated measures designs. These include history, maturation, attrition, instrumentation, and carryover effects. In some cases, a change in the level of behavior may be the result of one of these extraneous variables rather than the introduction of the independent variable.

Time series analysis is also a technique that is often used to track changes in behavior that occur on a large scale. For example, does a full moon make people more likely to commit crimes? One could track crime statistics on a daily basis over a long period of time and relate those statistics to the fullness of the moon. Note that this is not an experimental design because there is no independent variable manipulated by the researcher. Thus, cause/effect conclusions would not be warranted. A time series analysis in which there is a bit more control would involve the tracking of crime statistics both before and after a new law is passed that increases the punishment for a particular crime. One primary purpose of such a law is to cause a reduction in the incidence of the crime. Even in this latter example it is very difficult to verify the effect of this new law because there are so many other factors that influence crime rate that are likely to vary over time (e.g., the economy). However, this is not to suggest that time series analysis of such questions should not take place. In fact, they should. What we do suggest is that we all need to evaluate such information with a very critical eye.

Case Analysis

Self-injurious behavior is exhibited by some individuals with developmental disabilities and is one of the most disturbing behaviors for people to observe. Attention can sometimes be an effective treatment for self-injurious behavior (e.g., Vollmer, Iwata, & Zarcone, 1993). Let's consider the hypothetical case of Mark T. Mark is a 53 year-old male with profound mental retardation. Each morning, Mark engages in head-banging behavior. As a result, he has to wear a protective helmet. You decide to explore the possibility that attention from caregivers could be an effective treatment. The study will take place over a two-week period, with baseline recordings made during the first week and treatment imposed during the second week. During the first week, the number of head-banging incidents is recorded from 7:00-9:00 each morning. During the second week, a procedure is implemented during the same 7:00-9:00 time periods. The procedure is such that a caregiver will approach and talk to Mark for two minutes every time a 10-minute interval passes during which there was no head-banging. Thus, Mark is reinforced for behaviors other than head-banging. The data that represent the number of head-banging incidents during each 2-hour period are as follows:

Week 1 (Baseline)	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	6	10	8	13	12	15	16
Week 2 (Treatment)	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14
	15	12	10	11	8	8	5

Critical Thinking Questions

1. Which single-subject design was used?
2. Sketch a line-graph that shows the number of head-banging incidents across the two-week period.
3. Are you justified in drawing the conclusion that attention was an effective treatment in reducing head-banging behavior? Why or Why not?
4. Was an optimal baseline achieved? How would you improve this aspect of the study?
5. Describe how intra- and inter-participant replication could be achieved.
6. What other control procedures should be considered?

General Summary

Single-subject designs are experimental designs in which there is manipulation of an independent variable and careful analysis of the behavior of individual participants. Participant's scores are not grouped together and inferential statistics are not used to arrive at conclusions. Rather, a single participant's

behavior is compared across different treatment conditions and conclusions are strengthened by intra-participant replication and inter-participant replication. Intra-participant replication requires that the experimental conditions be repeated for the same individual. A consistent pattern of behavior differences between the experimental conditions suggests that the independent variable is causing the difference. Intra-participant replication is difficult when the behavior under investigation is not reversible. In such situations, a multiple-baseline procedure may be useful. Inter-participant replication requires that the experimental conditions be repeated with one or more other individuals. Again, a consistent pattern of behavior differences between experimental conditions that is observed in more than one participant strengthens the argument that the independent variable is causing the difference.

This chapter concludes a series of chapters that described a variety of experimental designs in which cause-effect conclusions could be made. Although these experimental designs are powerful tools for answering many research questions, not all questions can, or should, be answered with such techniques. The next chapter will review a variety of nonexperimental methods that can be used to learn more about behavior.

Detailed Summary

1. The single-subject approach is a method designed to study the behavior of individual organisms. As the method continues to evolve and improve, it also has become more popular for both scientific and therapeutic purposes.

2. The single-subject approach is different than the case-study or case-history approach. The latter approach is characterized by lack of experimental control and imprecise measures of behavior.

3. Historically, the introduction of random assignment and statistical analyses to the research world led to a bias to conduct research on relatively large groups of participants. However, there are limitations to what can be learned by studying groups. Single-subject designs, when used properly, can provide valuable information about individual behavior and can elucidate causal relationships between independent and dependent variables.

4. For single-subject designs, the reliability of findings involves actual replications rather than inferential statistics. The two types of replications are intra-participant replication (replications within an individual participant) and inter-participant replication (replications between individual participants).

5. Advantages of single-subject designs include immediate feedback to the researcher regarding changes in behavior, a focus on individual behavior, and strong conclusions regarding the effect of one variable on another.

6. Limitations of single-subject designs include their unsuitableness for answering actuarial types of questions, for comparing two or more different treatments on the same behavioral measure, for treating an entire group of participants in an identical way, and for "after the fact" studies (ex post facto, correlational, passive observational). Moreover, the single-subject approach makes heavy time demands and often relies on the establishment of stable baselines and reversibility of behavior.

7. Researchers favor the single-subject method for different reasons. Some reasons include interest in individual behavior, the fact that group data can mask individual behavior, unavailability of large groups of participants, and concerns with methods of statistical evaluation.

8. To determine an effect of an independent variable (treatment), the dependent measure should be operationally defined and should exhibit a stable baseline level of occurrence with repeated measurements. An optimal baseline is one that is stable and not subject to floor or ceiling effects.

9. An optimal baseline is achieved by identifying and controlling extraneous sources of variability in the environment.

10. The ABAB design is one of the most powerful design strategies. A baseline is first established (A), then we introduce the treatment for the first time (B). After the criterion of ability is achieved we then withdraw the treatment and reintroduce the baseline condition (A). Finally, after baseline stability is reestablished, we present the treatment condition (B) for the second time. With this design the researcher demonstrates the degree of control over behavior in two ways—first by *introducing* the treatment condition, then by *removing* it.

11. Intra-participant replication demonstrates that the method is reliable, that the treatment effect is real, and that there is control over behavior. Inter-participant replication establishes the generality of the findings.

12. Intra-participant replication in the ABAB design relies on the reversibility of the behavior to baseline levels after the treatment is withdrawn.

13. The multiple baseline procedure allows the researcher to perform intra-participant replication with different independent responses rather than the same response.

14. Time series analysis is characterized by repeated measurements of the dependent variable over time with an introduction of the independent variable at a particular point in time. Trends or patterns of behavior are observed both before and after introduction of the independent variable.

Key Terms

Baseline

Facilitation

Inter-participant replication

Intra-participant replication

Irreversable behavior

Multiple baseline procedure

Optimal baseline

Replication

Reversible behavior

Suppression

Time-series design

Review Questions/Exercises

1. Construct a hypothetical graph of the results of a single-subject design to evaluate the effects of TV violence on a child. The axes on the graph should be clearly labeled. The graph should illustrate an optimal baseline, reversible behavior, intra-participant replication, and an effect such that TV violence results in more aggressive behavior.
2. Summarize inter-participant replication and its purpose.
3. In 2002, after much debate, the high school in Conway, AR implemented a random drug testing program of students involved in any extracurricular activities. The expressed purpose of the program is to reduce drug use among teenagers. Describe how a time-series design might be used to evaluate the effectiveness of the drug testing program.