

*ALTERNATING TREATMENTS DESIGN: ONE
STRATEGY FOR COMPARING THE EFFECTS OF TWO
TREATMENTS IN A SINGLE SUBJECT*

DAVID H. BARLOW AND STEVEN C. HAYES

STATE UNIVERSITY OF NEW YORK AT ALBANY AND
UNIVERSITY OF NORTH CAROLINA AT GREENSBORO

A little used and often confused design, capable of comparing two treatments within a single subject, has been termed, variously, a multielement baseline design, a multiple schedule design, and a randomization design. The background of these terms is reviewed, and a new, more descriptive term, Alternating Treatments Design, is proposed. Critical differences between this design and a Simultaneous Treatment Design are outlined, and experimental questions answerable by each design are noted. Potential problems with multiple treatment interference in this procedure are divided into sequential confounding, carryover effects, and alternation effects and the importance of these issues vis-a-vis other single-case experimental designs is considered. Methods of minimizing multiple treatment interference as well as methods of studying these effects are outlined. Finally, appropriate uses of Alternating Treatments Designs are described and discussed in the context of recent examples.

DESCRIPTORS: Single-subject design, methodology, comparison of two treatments

To compare the effects of two or more treatments in applied research, each treatment is usually administered to a different group of subjects and differences are noted. Because considerable intersubject variability exists in each group (some subjects change and some do not), inferential statistics are often necessary to determine if an effect exists. This leads to problems in generalizing results from the group average to the individual subject or patient who should benefit from the research (Hersen & Barlow, 1976; Sidman, 1960). To avoid intersubject variability, an ideal solution would be to divide one subject in two and apply two different treatments simultaneously to each identical individual. This would eliminate intersubject variability and allow effects, if any, to be directly observed. Statements about other individuals could then be made through the usual process of replication

and "logical generalization" (Edgington, 1966; Hersen & Barlow, 1976).

Such a procedure exists in the family of single-case experimental designs although it has been little used and often confused. It has been termed variously a *multiple schedule design* (Barlow & Hersen, 1973; Hersen & Barlow, 1976; Leitenberg, 1973), a *multielement baseline design* (Sidman, 1960; Ulman & Sulzer-Azaroff, Note 1), and a *randomization design* (Edgington, 1967). In addition, Kazdin and Hartmann (in press) use the term *Simultaneous Treatment Design* (see below). These terms were originated for somewhat different reasons, reflecting the multiple historical origins of single case research (Hersen & Barlow, 1976). Several proponents of the term *multiple schedule* (see Hersen & Barlow, 1976; Leitenberg, 1973) were associated in Vermont in the 1960s in an effort to apply operant procedures and methods to clinical problems (e.g., Agras, Leitenberg, Barlow, & Thomson, 1969). These procedures and terminology were derived directly from operant laboratories; but the term *multiple schedule* implies a distinct schedule associated with each stimulus

Reprint requests should be sent to David H. Barlow, Psychology Department, State University of New York, Albany, New York 12222. We would like to thank Harold Leitenberg, Mike Zeiler, and Warren Steinman for comments on an earlier draft of this manuscript.

component, and this may not always obtain in applied research, resulting in an unnecessary narrowness in the term. Ulman and Sulzer-Azaroff (1975) use Sidman's term *multielement baseline design* to describe a procedure in which different conditions or "treatments" are associated with different stimuli to establish experimental control. In fact, multielement baselines as conceived by Sidman (1960) have little parallel in applied research since the purpose is to "investigate relations between some single experimental operation and more than one behavioral baseline" (p. 326). Of more direct relevance is Sidman's term *multielement manipulation* in which the purpose is to study "the interaction between a single behavioral baseline and several qualitatively or quantitatively different experimental operations" (p. 323); in other words, a comparison of the effects of two or more treatments on one behavior. As in much basic research, however, Sidman's examples illustrate high-rate behavior brought to a point of stability before introduction of experimental operation or "treatments." Applied research, on the other hand, is more often concerned with low-rate behaviors which are unstable.

Edgington (1967, 1972), from a position outside of operant psychology, originated the term *randomization design* to describe his variation of a time series approach amenable to statistical analysis. The design differs slightly from traditional operant application in that treatments are deliberately randomized across times of application: for example, ABBABAA rather than ABAB. Treatments are repeated often enough to allow statistical comparison of A and B phases (continuing a tradition begun by R. A. Fisher (0000) who explored the abilities of a lady to discriminate tea prepared in two different ways). This frequent repetition of treatments usually requires fast alternation to obtain the necessary number of random observations, and the almost unavoidable discriminability of each condition in applied research with human (e.g., McCullough, Cornell, McDaniel, & Mueller, 1974)

makes this approach procedurally similar to the two approaches described above.

The basic feature of this design, under its various names, is the fast alternation of two different treatments or conditions, each associated with a distinct and discriminative stimulus. As Leitenberg (1973) points out, this design "is based on discrimination learning principles; that is, if the same behavior is treated differently in the presence of different physical or social stimuli, it will exhibit different characteristics in the presence of these stimuli" (p. 93). Thus in the typical design, after a baseline period, two treatments (A and B) are administered, alternating with each other, and the effects on one behavior are observed. For example, A may be administered in the morning and B in the afternoon, preceded by instructions such as, "This is treatment A" and "This is treatment B." Conditions which might affect data other than treatments are counterbalanced as the experiment continues, such as time of day, therapist administering the treatment, or location of the treatment. For example, B might be given in the morning one day and the afternoon the next. The data are plotted separately for each intervention to provide a ready visual representation of the effects of each treatment. Because confounding factors such as time of administration have been neutralized (presumably) by counterbalancing, and because the two treatments are readily discriminable by subjects through instructions or other discriminative stimuli, differences in the individual plots of behavior change corresponding with each treatment should be attributable to the treatment itself, allowing a direct comparison between two (or more) treatments. [Also see Kazdin and Hartmann (in press) for a discussion of the logic of this design.]

For example, McCullough et al. (1974) described treatment of disruptive behavior in a 6-year-old boy. Following a 5-day baseline period in which cooperative behavior was measured, two treatments were introduced for a total of 4 days: (a) social reinforcement for coopera-

A CASE STUDY

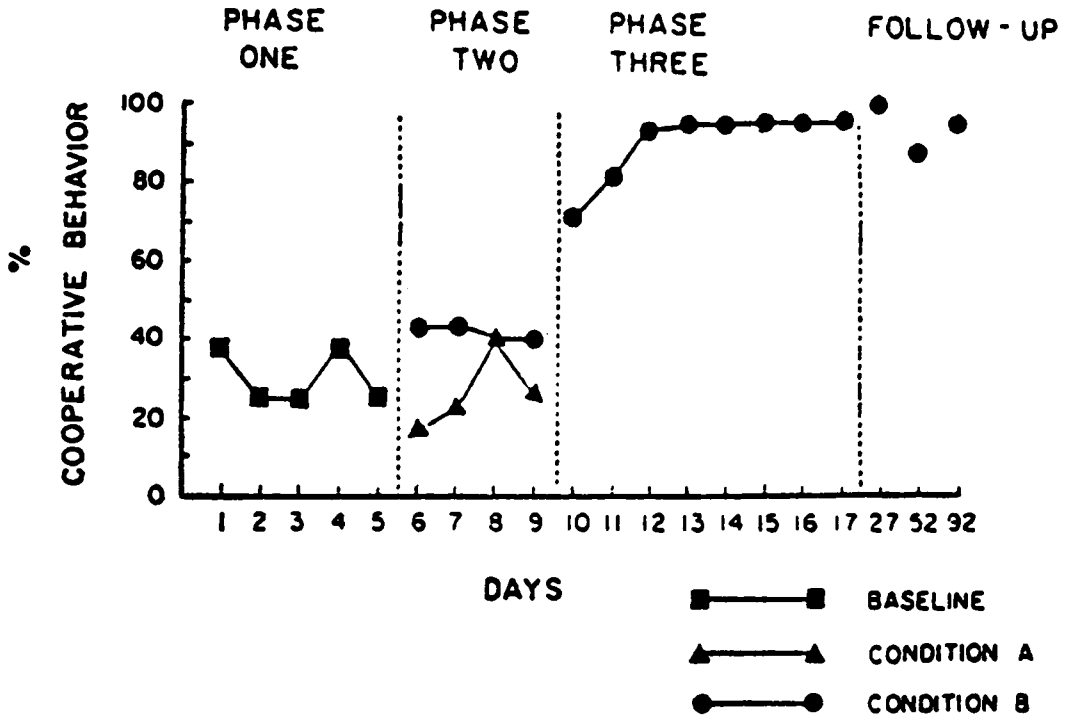


Fig. 1. Percentage of observation periods in which Cedric emitted cooperative behavior. (Reprinted from McCullough, J. P., Cornell, J. E., McDaniel, H. H., & Mueller, R. K. Utilization of the simultaneous treatment design to improve student behavior in a first-grade classroom. *Journal of Consulting and Clinical Psychology*, 1974, 42, 288-292. Copyright 1974 by the American Psychological Association. Reprinted by permission.)

tive behavior and ignoring uncooperative behavior (labeled Treatment A), and (b) social reinforcement for cooperative behavior plus time out for uncooperative behavior, in this case removal from the classroom for 2 minutes (labeled Treatment B). A teacher (T-1) and a teacher's aide (T-2) administered the treatments with the teacher administering Treatment A the first 2 days and Treatment B the last 2 days. For purposes of this discussion, it is most important to note that a treatment was administered during both a morning session (9:00 to 11:00 a.m.) and an afternoon session (12:00 noon to 2:00 p.m.) and that treatments were *alternated* during the day so that one treatment was offered in the morning and the other in the afternoon. Across all 4 days the time of administration of a partic-

ular treatment (a.m. or p.m.) was counterbalanced. The effect of the two treatments are presented in Figure One. Treatment B increased cooperative behavior more than Treatment B in Phase 1 and therefore was continued in Phase 3.

Because none of the names mentioned above, specifically multiple schedule, multielement baseline (or more accurately multielement manipulation), and randomization design is either wholly accurate or totally suited to describe the various conditions that obtain in applied research, a new name for the design is proposed. *Alternating Treatments Design*¹ has the advan-

¹Ulman and Sulzer-Azaroff (1975) suggested a similar name, "alternating conditions," as a possibly more descriptive substitute for multielement baseline. "Treatments" seems preferable, however, because of

tages of avoiding the inaccuracies associated with the above mentioned terms and, at the same time, describing the essential feature of this design, the fast alternation of two or more treatments in a single subject.

ALTERNATING TREATMENTS AND SIMULTANEOUS TREATMENT DESIGNS

A source of some confusion has been the similarities and differences between the Alternating Treatments Design (ATD) and its various names described above and the Simultaneous Treatment Design (STD). The STD has also been termed the concurrent schedule design in Hersen and Barlow (1976). Concurrent schedule design had the same historical origin as the term multiple schedule design described above (Barlow & Hersen, 1973; Leitenberg, 1973), but the implication that a distinct schedule of reinforcement is attached to each treatment produces the same unnecessary narrowness as in the term multiple schedule design. Browning's (1967) term, simultaneous treatment design, seems more descriptive and suitable. Nevertheless, both terms adequately describe the fundamental characteristics of this design, the concurrent or simultaneous application of two or more treatments in a single case. This contrasts with the fast alternation of two or more treatments in the ATD. Hersen and Barlow (1976) noted that only one example of the use of an STD exists in applied research, the original Browning (1967) experiment, also described in Browning and Stover (1971). This is still true.

In this experiment, as in a true concurrent schedule, the subject is able to choose the preferred schedule or treatment since those treat-

ments were simultaneously present. But it is unlikely that the subject will be equally exposed to each treatment. In fact, the very structure of concurrent schedule ensures that the subject will not be equally exposed to all treatments because a choice is forced (except in the unlikely event that both treatments are equally preferred). The data from Browning's subject indicate a "preference" for the treatment "verbal admonishment" as indicated by frequency and duration of bragging and a decided lack of preference for ignoring. There may be instances, however, when preference for a treatment may have little relation to its effectiveness. This point will be discussed later.

Contrast this with the McCullough et al. (1974) experiment which was termed an STD by the authors and also by Kazdin and Hartmann (in press). These two experiments have much in common. Both successfully compared the effects of two or more treatments in a single case. Both used therapists as discriminative stimuli for the treatments, and, therefore, both had to counterbalance therapists to control for the effects of an individual therapist. Because it is essential that discriminations be formed, it is remarkable that each teacher was associated with a given treatment for only 2 days in McCullough et al. (1974). As each "session" lasted 2 hours, the behavior and subsequent treatment application were evidently occurring at a high rate, allowing this discrimination to occur. Both experiments also employed a Latin square statistical analysis suitable for use with a single subject (Benjamin, 1965).

But one procedure was a simultaneous treatment design with choice or preference for treatments as the method of comparing results and the other was an alternating treatment design in which the subject experienced each of the rapidly alternating treatments for an equal amount of time with the effects on behavior noted. The ATD, of course, requires a further counterbalancing of times of administration and this was evident in the McCullough et al. (1974) experiment.

its parallel in the well-accepted term "simultaneous treatment design" and the tradition of utilizing treatment to refer to a distinct, independent variable in both applied and basic research. Warren Steinman has suggested substituting "intervention" for "treatment" to avoid an overly medical or clinical connotation, but given the parallels and traditions noted above, "treatment" probably communicates more information.

This is a major procedural difference, with implications for the types of experimental questions answerable with each design as well as the nature of the data used in comparing two treatments. Thus, it would seem important to make this distinction in the case of some recent excellent examples of ATDs which have been termed STDs (e.g., Kazdin, 1977; Kazdin & Geesey, 1977). For example, it is difficult to conceive how the McCullough et al. (1974) experiment could have been administered using an STD. For this to occur, both the teacher and teacher's aide would have to be present in the classroom administering different treatments simultaneously. Furthermore, the subject would have to approach one or the other in a free operant fashion for the treatment to be administered—an unwieldy procedure at best. After the discrimination was made, the subject might continue the disruptive behavior by approaching only the therapist administering Treatment A. One would have to infer, then, that Treatment B was more effective although very few trials with Treatment B might occur because preference is being measured rather than effects on a given behavior. Of course, one can conceive of many instances where preference among several treatments would be a significant applied question.

MULTIPLE TREATMENT INTERFERENCE

Multiple treatment interference (Campbell & Stanley, 1963) or condition change interactions (Ulman & Sulzer-Azaroff, 1975) pose the question: Will the results of Treatment A in an ATD where it is juxtaposed with Treatment B be the same as when Treatment A is applied in isolation? In other words, will the results of Treatment A be generalizable from the contrived experimental situation to the natural situation. This is no small issue, since the external validity or generalizability of the result is a major portion of any experimental inquiry. It is understandable that this issue should arise in relation to an experimental design that features

fast alternation of treatments or conditions as this is more unlike the real situation than the first (treatment) phase of a withdrawal design or treatment in the experimental group in a between-group comparison design. This issue must be put in perspective.

Few would question the internal validity of the ATD or the ability of the design to rule out rival hypotheses. In fact, the testing of two treatments in the same subject within the same time period produces one of the most elegant controls for most threats to internal validity. But critics who become overly concerned about external validity have, in Campbell and Stanley's (1963) view, evidenced "a recurrent reluctance to accept Hume's truism that induction or generalization is never fully justified logically" (p. 17). Because few applied behavioral researchers derive random samples, inference of results from a group to a population of individuals is not possible (Hersen & Barlow, 1976). Technically an experiment, although internally valid, is generalizable only to subjects with exactly the same set of characteristics, during the same time of day, under the same weather conditions and star constellations. Because this would get us nowhere, we often guess which factors will affect generalizability and which will not in a given experiment and proceed accordingly (Campbell & Stanley, 1963; Edgington, 1969). Perhaps we decide, based on some previous experience, that IQ will be a factor in generalizability, but star configuration will not. We would then replicate the experiment on subjects with different IQs. Campbell and Stanley (1963) note that although we should strive for as much representativeness of the natural environment as is possible while maintaining strong internal validity "we should keep in mind that the 'successful' sciences such as physics and chemistry made their strides without any attention to representativeness (but with great concern for repeatability by independent researchers)" (p. 18).

Thus, in any science, external validity takes a (temporary) back seat to internal validity, but this is particularly true with the ATD, because

internal validity is so powerful and replication on additional individuals rather than statistical inference from groups to populations is the primary means of establishing external validity in our science of applied behavior analysis. Furthermore, a close look suggests that multiple treatment interference may be no more of a problem for this design than for some other designs, and may, in some instances, be less.

The "messy" area of applied research is fraught with multiple treatment interference. Unlike the splendid isolation of animal laboratories where rats are returned to their cages for 23 hours to await the next session, the children and adults who are the subjects of applied research are experiencing a variety of events before and between treatments. One subject may have recently lost a family member, another flunked an exam, a third had sexual intercourse, and a fourth was mugged on the way to a session. It is possible that these subjects responded differently to treatment than otherwise would have been the case, and these historical factors account for some of the enormous intersubject variability in between-group designs comparing two treatments. ATDs, on the other hand, attempt to control for this experience by dividing each subject in two and administering two or more treatments within the same period of time. As with all applied research, the results may be affected by interaction with events occurring in the environment which form a background (baseline) for the experiment, and the fact that single case experiments are replicable at all in view of this "multiple treatment interference" may be surprising. But, *within a single case*, ATDs handle outside interference more effectively than, say, withdrawal designs but, at the same time, introduce the issue of one *experimental* treatment interfering with another.

Ulman and Sulzer-Azaroff (1975), in an excellent discussion, divide the problem of a multiple treatment interference into sequential confounding and carryover effects. To this we would add alternation effects. Sequential confounding, of course, is the major reason why it is not pos-

sible to compare two treatments in a standard A-B-A design, or one of its variations. Because the B treatment follows the A treatment, its effects are confounded by the prior administration of A. For example, if a teacher institutes praise in a classroom and then add rules, one could say something about the effect of praise but nothing about rules except as it follows praise. To compare rules and praise in a straight A-B-A design, one would have to counterbalance the order of administration in a second subject (with several additional direct replications) using an interaction design strategy (Hersen & Barlow, 1976) to look at the separate and combined effects of each variable or treatment.² Later phases in the A-B-A design and its variations exist solely for purposes of internal validity. External validity must come from the first treatment phase, due to problems with sequential confounding. In an ATD, however, the effects of sequence are controlled by counterbalancing (e.g., ABBAAB). This is made possible by rapid alternation, which allows more administrations of A and B in a shorter period of time than is possible with the standard A-B-A design where phases may last days or weeks. This counterbalancing also allows statistical analysis of ATDs for those who so desire. Edgington's (1967) randomization procedure or Benjamin's (1965) special Latin square analysis (e.g., McCullough et al., 1974) have been used.

Carryover effects (or contextual effects), on the other hand, refer to the influence of one treatment on an adjacent treatment, irrespective of overall sequencing. These effects may be di-

²The corollary to this approach in applied group comparison research is the within subject design (e.g., Cochran & Cox, 1957), also termed a crossover or randomized block design (Edwards, 1968) among other names. In this design, the order of administration of two (or more) treatments is counterbalanced in additional groups of subjects. Overall response to treatments across the two (or more) groups is then determined and compared statistically since the counterbalancing and subsequent statistical analysis is seen as handling sequential confounding. However, the approach ignores carryover effects, which are then averaged into the group differences.

vided into contrast and induction (Reynolds, 1968; Ulman & Sulzer-Azaroff, 1975). Contrast refers to changes in behavior in a direction opposite to that expected due to a contrast with another treatment. For example, Azrin and Holz (1966) point out that comparing different magnitudes of punishment could make the lesser magnitude actually reinforcing. This is illustrated in a sequentially confounded experiment comparing 30-min versus 15-min versus 1-min time-out periods on children's disruptive behavior (White, Nielsen, & Johnson, 1972). If a 1-min time-out period was implemented first in the sequence, disruptive behavior in children was reduced; but if it followed a longer time out period in the experimental sequence, disruptive behavior actually rose above baseline, presumably due to the contrast with the much longer periods of time out which retained their suppressive effects.

Induction refers to a positive transfer between treatments with the behavior during one treatment more closely approximating the behavior during a second treatment than would occur if the treatments were applied individually. For example, if the 1-min time-out period noted above produced greater suppression following a 15-min time-out period than it did coming first in sequence, this would be induction. This phenomena emerge from basic research on components in a multiple schedule and reviews (Dunham, 1968; Freeman, 1971) suggest that the effects, although reliable, are small (Ulman & Sulzer-Azaroff, 1975).

Carryover effects in humans within the context of a multiple schedule are most often transient as treatments are relatively widely spaced, but Waite and Osborne (1972) demonstrated sustained contrast in children in a mult VI 20-sec EXT with 2-min schedule components. Nevertheless, Ulman and Sulzer-Azaroff (1975) and Sidman (1960) suggest several methods for minimizing or eliminating contrast or induction.

First, counterbalancing the order of treatments, as is necessary to control for sequential confounding so that treatments will follow one

another in an unpredictable fashion, should minimize carryover effects. For example, in the White et al. (1972) experiment mentioned above, each of three time-out periods was administered to three different groups of children. In the group receiving the 1-min treatment initially, disruptive behavior was suppressed. However, 1-min time-out periods were not suppressive and perhaps even facilitative if they followed the 30-min time-out period. But each time-out period took 2 weeks with 2-week baselines interspersed. Shorter, more numerous, and unpredictable periods of treatment still separated by a reasonable period of time such as several hours (e.g., McCullough et al., 1974) might produce less contrast, particularly since O'Brien (1968) demonstrated that relatively brief periods of treatment minimize contrast effects.

Second, Powell and Hake (1971) minimized carryover effects in a study comparing two reinforcement conditions by presenting only one condition per session, a situation that usually obtains in applied research (e.g., McCullough et al., 1974; Agras et al., 1969). It is interesting to note that similar procedures have been suggested to minimize carryover effects in the traditional within-subjects group approach (Greenwald, 1976).

Finally, a third issue in studying carryover effects is the speed of alternation of treatment. For example, multiple-schedule work in basic research when carryover effects have been studied usually alternates schedules by the minute rather than once or twice a day as is now typical in applied research. This seems to heighten carryover effects, particularly contrast, as noted above (Powell & Hake, 1971; Waite & Osborne, 1972). But alternation must be frequent enough to allow a discrimination to be formed. The appropriate speed of alternation which allows discrimination learning but minimizes carryover effects in an experimental question will probably depend on the particular question asked.

With these steps and in view of the nature of applied research, including the ability of humans to discriminate quickly and efficiently, it

would seem that carryover effects should not be a stumbling block to the external validity of an experiment. But, as Ulman and Sulzer-Azaroff (1975) note, "In the absence of a systematic investigation, however, such interaction remains unspecified, and any generalization based on this design should be qualified accordingly" (p. 389).

Fortunately, it is possible to assess directly the extent to which such effects are present. Sidman (1960) suggests two methods. One is termed *independent verification*, which essentially entails conducting a control experiment in which one or the other of the component treatments in the ATD are administered independently. For example, two treatments might be compared through an ATD in a direct replication across three subjects. Three more subjects might then receive baseline, followed by Treatment A, in an A-B fashion. The second treatment could be administered to a third trio of subjects in the same manner. Any differences that occur between the treatment administered in an ATD or independently could be due to carryover effects. Alternatively, these subjects could receive Treatment A alone, followed by an ATD alternating Treatments A and B, returning to Treatment A alone. An additional three subjects could receive Treatment B in the same manner. Trends and levels of behavior during either treatment alone versus the same treatment in the ATD could be compared.

A more elegant method is termed *functional manipulation* by Sidman (1960). In this procedure, the strength or intensity of one of the components is changed. For example, if comparing flooding and structured approach in the treatment of fear, the amount of time in flooding could be doubled at one point. Changes in fear behavior occurring during the second unchanged treatment (structured approach) could be attributed to carryover effects.

As Sidman (1960) observed, the study of treatment interaction can be interesting in its own right. In addition to the important step of determining the presence of carryover effects in

an ATD, it is possible that some treatments juxtaposed in fast alternation could prove more effective than either component alone. That is, alternation effects, mentioned above, could prove therapeutic. For example, in some recent unpublished work, a sadistic rapist was treated by daily alternation of orgasmic reconditioning using first a sadistic fantasy and second an appropriate heterosexual fantasy. Sexual arousal to the appropriate fantasy seemed to increase more quickly during the fast alternation than during orgasmic reconditioning to the appropriate fantasy alone (Abel, Blanchard, Barlow, & Flanagan, Note 2). This may represent a contrast effect or possibly an intensification of the therapeutic effect due to a sharpening of stimulus control. Ulman and Sulzer-Azaroff (Note 1) also cite several studies reporting a possible intensification effect after multiple reversals in an A-B-A-B design. The appropriate method for studying these alternation effects would be to juxtapose a period of fast alternation with a period of slower alternation.

In summary, the unrepresentativeness of the ATD to natural situations as a threat to external validity is less of a drawback than it might be, due to the prevalent replication strategies in applied behavior analysis (as opposed to statistical inferential strategies of generalization) and the superior interval validity present in this design. Nevertheless, there are methods to minimize carryover effects as well as methods to study carryover effects which should be pursued to improve the external validity of the ATD and for possible applied value intrinsic to the fast alternation of two treatments.

USES OF ALTERNATING TREATMENT DESIGNS

ATDs have been used in two ways: (a) to compare the effect of treatment and no treatment (baseline) and (b) to compare two distinct treatments. Each of these approaches require separate comment.

Comparing Treatment with No Treatment

Several investigators have compared treatment and no treatment in an ATD. For example, O'Brien, Azrin, and Henson (1969) compared the effect of following and not following suggestions made by chronic mental patients in a group setting on the number of suggestions made by these patients. Doke and Risley (1972) alternated daily the presence of three teachers versus the usual one teacher and noted the effect on planned activities in a classroom (contingencies on individual versus groups were also compared in an ATD later in the experiment). Redd and Birnbrauer (1969) alternated reinforcement and no reinforcement, using two adult therapists as discriminative stimuli. The two adults switched treatments half way through the experiment to control for effects of person. Zimmerman, Overpeck, Eisenberg, and Garlick (1969) and Ulman and Sulzer-Azaroff (1975) also compared reinforcement and no reinforcement, although Ulman and Sulzer-Azaroff included two types of reinforcement, group versus individual. Finally, Agras, et al. (1969) studied the effects of social reinforcement in a severely claustrophobic patient by alternating social reinforcement with no reinforcement.

While data from these experiments are convincing, questions asked in the experiments mentioned above could all be answered by use of the more standard A-B-A-B withdrawal design. In the area of fear reduction, for example, the question of reinforcement versus no reinforcement and even the question of the role of relaxation in systematic desensitization have both been addressed using withdrawal designs (Agras, Leitenberg, & Barlow, 1968; Agras, Leitenberg, Barlow, Curtis, Edwards, & Wright, 1971).

The advantages of the ATD over the more usual withdrawal design have been enumerated by Ulman and Sulzer-Azaroff (1975) and Kazdin & Hartmann (in press). Clearly the major advantage of the ATD is that it does not require a withdrawal of treatment which may result in

a reversal of any therapeutic gains. This allows one to proceed without concern for the ethical issue of reversing clinically relevant behavioral gains, an issue which sometimes arises in clinical research. Occasional staff resistance to withdrawal of treatment is also avoided.

A second advantage is that the comparison can be made more quickly than in a withdrawal design. McCullough et al. (1974) for example, effectively compared two treatments in 4 days. Withdrawal designs, on the other hand, require relatively stable baselines followed by at least three, and usually more, data points in each of at least three phases (A-B-A). As Ulman and Sulzer-Azaroff (1975) note, this efficiency also allows sudden termination of the experiment with the likelihood of having obtained usable data. A withdrawal design, however, must be carried through to completion.

A final advantage is the possibility of proceeding without a formal baseline phase. Ulman and Sulzer-Azaroff (1975), in considering this point, suggest that behaviors yielding chronically unstable baselines can be studied with this design. In applied research, the most common observation is behavioral improvement during baseline, which does not allow for introduction of treatment in the usual withdrawal design (Hersen & Barlow, 1976). But this does not present a problem for the ATD.

Despite these advantages, there are distinct disadvantages. Foremost among them is the as yet unknown magnitude of multiple treatment interference existing in the ATD. Until these issues are thoroughly explored experimentally, the external (or ecological) validity is uncertain since there are very few straight applied situations where a treatment is alternated with no treatment. The first time a treatment is introduced in a withdrawal design, however, does very closely resemble the applied situation because the treatment is administered in a straightforward manner. As noted above, later withdrawal and reinstatement phases occur solely for the sake of internal validity. Thus, one can as-

sume more readily that the first treatment phase of a withdrawal design is externally valid.

One way to avoid this problem would be to administer the treatment alone in the first phase, followed by the ATD. This would still be more economical than a withdrawal design since one would not need a baseline nor the final reinstatement phase, yet one could make some estimates on the generalizability of the treatment from the first phase without the fast alternation.

A second disadvantage is that the ATD could be more cumbersome to arrange than the withdrawal design. Not only must treatments be quickly alternated, but discriminative stimuli, times, and locations (if different) must all be counterbalanced. However, increased experience with the ATD in our setting suggests that this is not a major problem.

Nevertheless, comparison of treatment with no treatment can be made with either design, and choice between the ATD and the withdrawal design will depend on the experimental questions asked and the practicalities of the experimental situation.

COMPARING TWO TREATMENTS

When the experimental question is the comparison of two treatments, there are few alternatives.³ Indeed, the majority of ATDs published have attempted to answer this question (e.g.,

³Other than the within subjects comparison (e.g., Edwards, 1968), one could address this question only by a series of withdrawal designs which administered two treatments with baseline interspersed and counterbalanced for sequence effects. For example, three subjects could receive A-B-A-C where B and C were two distinct treatments, and three could receive A-C-A-B. This design would approximate the counterbalanced within-subject group comparison with the exception of the individual analysis of the data and repeated measurement. But it would likely maximize contrast effects (White, Nielsen, & Johnson, 1972). Another possibility would be the interaction design strategy (Hersen & Barlow, 1976), but this would require that the two treatments or variables be combined at some point to examine their separate and combined effects, a strategy that would not be possible if the treatments could not be combined. An example would be flooding and desensitization.

Corte, Wolf, & Locke, 1971; Doke & Risley, 1972; Steinman, 1970). Here the advantages of the ATD mentioned above are all relevant, and because there are few alternatives, the issue of the ATD being cumbersome is not relevant. The only remaining disadvantage is the threat to external validity posed by multiple treatment interference pending clarification of these issues. In a manner similar to those discussed above, one way to avoid this problem would be to hypothesize which of the two (or more) treatments is more effective and administer that treatment (Treatment A) alone in the first phase, followed by the ATD. As noted earlier, one could then estimate external validity from the first phase. As a check on internal validity, one should administer Treatment B followed by the ATD to a second subject to control for sequential confounding of the ATD. In any case, this threat may not be as great as it appears, for reasons discussed in the section on multiple treatment interference.

Considerations in the Use of the ATD

The most elegant examples demonstrating the correct application of the ATD comparing two treatments have appeared only recently (Kazdin, 1977; Kazdin & Geesey, 1977; McCullough, et al., 1974; Ulman & Sulzer-Azaroff, 1975). These examples are elegant because each illustrates the proper use of the ATD based on our current knowledge. Among other considerations, each design controls for sequential confounding by randomizing the order of treatment, a procedure that was not carried out, for example, in Agras et al. (1969). This is illustrated in the data advanced by McCullough et al. (1974) and presented here in Figure 1. In each of the experiments mentioned above, time of administration and location of administration were also counterbalanced if these factors were relevant to the experiment. Finally, each experiment illustrates proper use of the discriminative stimuli.

When the discriminative stimuli themselves may influence the data, as in the case of different therapists, these must also be counterbalanced.

But when the S^Ds are closer to those used in basic research (colored lights or sounds), such as cards with varying instructions posted in front of the classroom in the Ulman and Sulzer-Azaroff (1975) experiment, then interference from this source is unlikely and counterbalancing is not necessary. Because the basis of the ATD is stimulus discrimination, it is crucial that discriminations are formed. Thus, if S^Ds must be counterbalanced, this can occur only after discriminations have been made clearly, as in Agras et al. (1969). Using much shorter periods of time, successful discriminations were also made in the McCullough et al. (1974) data (see Figure 1), but if no differences had appeared in these data, one could not say if the treatments did not differ from each other or simply that the discriminations were not made. It seems best to be conservative in this instance to ensure that proper discriminations are made, but if proper discriminations are made, treatments that are topographically very similar (e.g., 5-min versus 10-min time-outs) can be compared. As noted above, this problem can be avoided by using instructions as discriminative stimuli in applied research (Kazdin & Hartmann, in press). If one is comparing 5- versus 10-min time-out periods for disruptive behavior in children, there is no reason why one would not describe the treatments as each became appropriate, particularly because someone would most likely describe these conditions in the normal use of this procedure in the natural environment.

Finally, the comparison of two treatments is not the only question answerable with an ATD. As Kazdin and Hartmann (in press), among others, point out, a comparison of the effectiveness of different therapists or of different times of treatment administration could also be carried out. This could be accomplished by collapsing data points across interventions and examining the therapeutic effects of two quickly alternating therapists. This comparison could be made more elegantly by having two or more therapists alternate quickly in the administration of a single treatment.

REFERENCE NOTES

1. Ulman, J. D., & Sulzer-Azaroff, B. *Multielement baseline design in applied behavior analysis*. Symposium presentation at the annual meeting of the American Psychological Association, Montreal, August 1973.
2. Abel, G. G., Blanchard, E. B., Barlow, D. H., & Flanagan, B. *A controlled behavioral treatment of a sadistic rapist*. Paper presentation at the ninth annual convention of the Association for Advancement of Behavior Therapy, San Francisco, December 1975.

REFERENCES

- Agras, W. S., Leitenberg, H., & Barlow, D. H. Social reinforcement in the modification of agoraphobia. *Archives of General Psychiatry*, 1968, **19**, 423-427.
- Agras, W. S., Leitenberg, H., Barlow, D. H., & Thomson, L. E. Instructions and reinforcement in the modification of neurotic behavior. *American Journal of Psychiatry*, 1969, **125**, 1435-1439.
- Agras, W. S., Leitenberg, H., Barlow, D. H., Curtis, N., Edwards, J., & Wright, D. The role of relaxation in systematic desensitization. *Archives of General Psychiatry*, 1971, **25**, 511-514.
- Azrin, N. H., & Holz, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Croft, 1966.
- Barlow, D. H., & Hersen, M. Single case experimental designs: Uses in applied clinical research. *Archives of General Psychiatry*, 1973, **29**, 319-325.
- Benjamin, L. S. A special Latin square for the use of each subject "as his own control." *Psychometrika*, 1965, **30**, 499-513.
- Browning, R. M. A same-subject design for simultaneous comparison of three reinforcement contingencies. *Behaviour Research & Therapy*, 1967, **5**, 237-243.
- Browning, R. M., & Stover, D. O. *Behavior modification in child treatment: An experimental and clinical approach*. Chicago: Aldine-Atherton, 1971.
- Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963.
- Cochran, W. G., & Cox, G. M. *Experimental designs* (2nd ed.). New York: John Wiley, 1957.
- Corte, H. E., Wolf, M. M., & Locke, B. J. A comparison of procedures for eliminating self-injurious behavior of retarded adolescents. *Journal of Applied Behavior Analysis*, 1971, **4**, 201-215.
- Doke, L. A., & Risley, T. R. The organization of day-care environments: Required vs. optional ac-

- tivities. *Journal of Applied Behavior Analysis*, 1972, 5, 405-420.
- Dunham, P. J. Contrasted conditions of reinforcement: A selective critique. *Psychological Bulletin*, 1968, 69, 295-315.
- Edgington, E. S. Statistical inference and nonrandom samples. *Psychological Bulletin*, 1966, 66, 485-487.
- Edgington, E. S. Statistical inference from $N = 1$ experiments. *Journal of Psychology*, 1967, 65, 195-199.
- Edgington, E. S. *Statistical inference: The distribution-free approach*. New York: McGraw-Hill, 1969.
- Edgington, E. S. $N = 1$ experiments: Hypothesis testing. *The Canadian Psychologist*, 1972, 13, 121-135.
- Edwards, A. L. *Experimental design in psychological research* (3rd ed.). New York: Holt, Rinehart & Winston, 1968.
- Freeman, B. J. Behavioral contrast: Reinforcement frequency or response suppression. *Psychological Bulletin*, 1971, 75, 347-356.
- Greenwald, A. G. Within-subjects designs: To use or not to use? *Psychological Bulletin*, 1976, 83, 314-320.
- Hersen, M., & Barlow, D. H. *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press, 1976.
- Kazdin, A. E. The influence of behavior preceding a reinforced response on behavior change in the classroom. *Journal of Applied Behavior Analysis*, 1977, 10, 299-311.
- Kazdin, A. E., & Geesey, S. Simultaneous-treatment design comparisons of the effects of earning reinforcers for one's peers versus for oneself. *Behavior Therapy*, 1977, 8, 682-693.
- Kazdin, A. E., & Hartmann, D. P. The simultaneous-treatment design. *Behavior Therapy*, in press.
- Leitenberg, H. The use of single-case methodology in psychotherapy research. *Journal of Abnormal Psychology*, 1973, 82, 87-101.
- McCullough, J. P., Cornell, J. E., McDaniel, M. H., & Mueller, R. K. Utilization of the simultaneous treatment design to improve student behavior in a first-grade classroom. *Journal of Consulting and Clinical Psychology*, 1974, 42, 288-292.
- O'Brien, F. Sequential contrast effects with human subjects. *Journal of the Experimental Analysis of Behavior*, 1968, 11, 537-542.
- O'Brien, F., Azrin, N. H., & Henson, K. Increased communications of chronic mental patients by reinforcement and by response priming. *Journal of Applied Behavior Analysis*, 1969, 2, 23-29.
- Powell, J., & Hake, D. F. Positive vs negative reinforcement: A direct comparison of effects on a complex human response. *Psychological Record*, 1971, 21, 191-205.
- Redd, W. H., & Birnbrauer, J. S. Adults as discriminative stimuli for different reinforcement contingencies with retarded children. *Journal of Experimental Child Psychology*, 1969, 7, 440-447.
- Reynolds, G. S. *A primer of operant conditioning*. Glenview, Ill.: Scott, Foresman, 1968.
- Sidman, M. *Tactics of scientific research*. New York: Basic Books, 1960.
- Steinman, W. M. The social control of generalized imitation. *Journal of Applied Behavior Analysis*, 1970, 3, 159-167.
- Ulman, J. D., & Sulzer-Azaroff, B. Multielement baseline design in educational research. In E. Ramp & G. Semb, *Behavior analysis: Areas of research and application*. Englewood Cliffs, N.J.: Prentice-Hall, 1975.
- Waite, W. W., & Osborne, J. G. Sustained behavioral contrast in children. *Journal of the Experimental Analysis of Behavior*, 1972, 18, 113-117.
- White, G. D., Nielsen, G., & Johnson, S. H. Timeout duration and the suppression of deviant behavior in children. *Journal of Applied Behavior Analysis*, 1972, 5, 111-120.
- Zimmerman, J., Overpeck, C., Eisenberg, H., & Glick, B. Operant conditioning in a sheltered workshop. *Rehabilitation Literature*, 1969, 30, 326-334.

Received April 17, 1978

Final acceptance November 15, 1978