REVISTA MEXICANA DE ANÁLISIS DE LA CONDUCTA 1998, **24**, 137-151 NÚMERO 2 (SEPTIEMBRE)
MEXICAN JOURNAL OF BEHAVIOR ANALYSIS NUMBER 2 (SEPTEMBER)

UTILITY OF GROUP METHODOLOGY IN BEHAVIOR ANALYSIS AND DEVELOPMENTAL PSYCHOLOGY

LA UTILIDAD DE LA METODOLOGÍA DE GRUPOS EN EL ANÁLISIS DEL COMPORTAMIENTO Y EN LA PSICOLOGÍA DEL DESARROLLO

HAYNE W. REESE[1]
WEST VIRGINIA UNIVERSITY

ABSTRACT

Many independent variables related to development are better approached with group than single-subject methods, either because experimental manipulation would be unethical or because they are only indexes of the real causal variables. Both kinds of variables can be studied in correlational group research, but correlational research deals with effect-effect relations, not functional relations, which are the subject-matter of behavior analysis. Understanding is the goal of behavior analysis; it is *tested by* rather than *equivalent to* the demonstration of control, and the demonstration can come from experimental group research in addition to single-subject research. Deviations of individual performance from group means is not necessarily a problem; variability can be dealt within experimental group research; between-group experiments can demonstrate functional relations; and statistical inference done properly is objective.

Key words: single-subject design, group design, variability, between-group experiments, developmental psychology

RESUMEN

Muchas de las variables independientes relacionados con el desarrollo se analizan mejor utilizando diseños de grupo, en lugar de métodos de un sujeto único, ya

[1] Much of the discussion in the second and third sections of this article was presented in, respectively, Reese (1996) and my paper in M. Peláez-Nogueras (Chair), *On prediction, control, understanding and scientific explanations*, symposium conducted at the meeting of the Association for Behavior Analysis, Chicago, IL, May 1997. Address reprint requests to Hayne W. Reese, Department of Psychology, West Virginia University, P.O. Box 6040, Morgantown, WV 26506-6040, USA.

sea porque la manipulaciones experimentales no serían éticas, o porque tan sólo son índices de las auténticas variables causales. Ambos tipos de variables se pueden analizar en estudios de grupos correlacionales, pero éstos tratan con relaciones de efecto-efecto y no con relaciones funcionales, las cuales son el objeto de estudio en el análisis del comportamiento. La meta del análisis del comportamiento es el comprender y esto se comprueba con, pero no es equivalente a, la demostración de control. La demostración de control puede derivarse tanto de la investigación experimental que utiliza grupos, como de la investigación que utiliza sujetos únicos. Desviaciones del promedio de grupo en las ejecuciones individuales no representan necesariamente un problema. La variabilidad de los datos conductuales puede ser tratada mediante diseños experimentales de grupos, dado que pueden mostrar relaciones funcionales y a que la inferencia estadística que los acompaña es objetiva.

Palabras clave: diseños de sujetos únicos, diseños de grupo, variabilidad, experimentos entre grupos, psicología del desarrollo

---

In this article I argue that research on psychological development usually involves independent variables that are more easily studied with group-research methods than with single-subject methods. I also argue that the use of single-subject methods is not essential for the behavior analysis of development and that the use of group methods can be a legitimate strategy. The arguments are given in the first two sections of the article; in the section after these, real and alleged limitations of group designs are discussed. As will be seen, group designs do not necessarily have most of the limitations usually attributed to them.

Variables Affecting Development

Typical Developmental Variables

Most of the research in developmental psychology involves "developmental" variables that cannot be experimentally manipulated but that can be assessed and varied between groups. Examples of developmental variables include age, schooling, sex, socioeconomic status, and social or cultural group. Baron (1990) and Neuringer (1991) said that group designs are necessary for studying the effects of developmental variables such as these. Responding specifically to Baron, Sidman (1990) said that in principle, single-subject designs could still be used if the researcher is ingenious enough. I agree, except that the researcher needs not only ingenuity but also, as indicated in the next subsection, much more knowledge than is currently available about most developmental variables.

Problems of Typical Developmental Variables

Typical developmental variables are problematic in at least two ways. First, variables such as age, schooling, sex, socioeconomic status, and sociocultural group probably do not have uniform effects even within a relatively homogeneous group. Second, these kinds of variables do not really exist in individual subjects except in trivial senses. For example, a person does not have an *age* except in the trivial sense of having a calendar or a clock marking the time since conception, birth, or some other life event; a person does not have *schooling* except in the trivial sense of having records of time spent in class or in school; and a person does not have a *sex* except in the trivial sense of having a label.

The reason for the second problem is that the kinds of variables under consideration are fictitious because they are index variables rather than causal variables and the causal variables they index are largely unknown. For example, schooling indexes at least the emergence of literacy, enlargements of knowledge, a large increase in control by adults and in the requirement of self-control by children, and a large increase in the use of "decontextualized" language, that is, language that can be understood without reference to the immediate context in which it occurs (e.g., Snow, 1983). Similarly, the labels "female" and "male" index many physiological and psychological variables such as (typically) XX versus XY genotype, shift from recessiveness to dominance of sex-linked genes, presence versus absence of menstrual cycles, and differences in typical physique, primary and secondary sex characteristics, gender stereotype, job opportunities, and wages. The problem is that when the independent variable is *schooling* or *sex*, for example, the research is incomplete unless the researcher can determine which if any of the variables indexed by schooling or sex influences the dependent variable.

The indexed variables can be identified theoretically or empirically, if they can be identified at all, but either way, traditional behavior analysis requires manipulating them experimentally in a single subject in order to demonstrate that they control behavior. A problem for developmental research is that although the experimental demonstration of control in a single subject is an excellent research strategy when it works, some of the suspected causal variables relevant to development cannot be manipulated experimentally--at least, not ethically.

For example, Hart and Risley (1995) found that socioeconomic status was the major predictor of vocabulary development in groups of young children. By means of regression analyses, they ruled out several potential causal variables, such as comprehensibility of parents' speech, number of persons present, race, sex of child, and time of day for data collection (pp. 61-62).

Further regression analyses indicated that the major variables were the amount and quality of cumulative experiences involving speech during the first three years of age (chap. 7). Another possibly relevant variable, which they did not assess, is the use of decontextualized speech. The use of decontextualized speech is more frequent in middle-than working-class children (Snow, 1983).

Variables that are correlated with age, schooling, and so forth, can be said to be indexed by that variable; they may be causal or they may themselves be index variables. For example, at about 6 years of age, physique becomes less like that of an infant and more like that of an adult, and "maturity" of physique at this age is correlated with success in school (Simon, 1959). Physique seems usually to be an index variable, but it can be a causal variable determining, for example, the probability of success in certain sports or the probability of being labeled stereotypically as "lean and mean" or "fat and jolly," which in turn would be variables indexed by physique (Sheldon, Stevens, & Tucker, 1940, chap. 7; Skinner, 1953, pp. 25-26).

Relevant Research Strategies

If the variables indexed by typical developmental variables were known, single-subject methods might well be preferred for developmental research. However, a reasonable strategy at present would be to do group research to obtain relevant descriptive information, in effect starting with Baer's (1973) second phase of developmental research (naturalistic research) and then going to his first phase (laboratory research). Hart and Risley's (1995) research strategy could be interpreted in this way, especially because they devoted their last chapter to implications of their group findings for practical intervention.

Goals of Behavior Analysis

The goals of behavior analysis are usually said to be the prediction and control of behavior, with heavy emphasis on control leading to insistence on single-subject research. Alternatively, Morris, Todd, and Midgley (1993) and I (Reese, 1996), among others, have argued that the goal is understanding, which makes prediction and control *tools* rather than *goals*, even though as Morris et al. pointed out and as Skinner (1953, p. 199) implied, understanding involves the discovery of controlling variables. "Controlling" variables are the same as "causal" variables, as they were called in the preceding sections of the present article, but many behavior analysts prefer the adjective "controlling." I suspect that the preference reflects in part some squeamishness about the word "cause" (e.g., Skinner, 1953, p. 23) and in part the emphasis in behavior analysis on "control."

Dealing with Variability

Sciences deal with uniformities (e.g., Skinner, 1953, chap. 2), but uniformities are obscured by variability except in research such as Neuringer's (e.g., 1992) in which variability is the dependent variable. Variability is a fact of nature (e.g., Russell, 1929, p. 250; Skinner, 1989); therefore, similarities or classes are not *found* in nature, they are *invented*. They are invented (defined) such that variability is averaged out--the dissimilarities are smoothed out. Consistently with this position, behavior analysis does not deal with repetitions of a given response, it deals with occurrences of an "operant," that is, occurrences of similar responses that define a "response class" (Skinner, 1953, chap. 5; 1989). Consequently, although responses within a class vary, this variability does not matter in behavior analysis. Unwanted variability still occurs, but it can be dealt with in various ways (Sidman, 1960, chap. 5-6 and passim).

Stimuli are also variable and consequently behavior analysis deals not with specific stimuli but with "stimulus classes." A *functional relation* is a class in which stimulus and response classes converge. At issue is whether functional relations, which are classes of classes, can be studied effectively with group as well as single-subject research designs.

Understanding

Most philosophers of science agree that the goal of science is explanation or, as it is also called, understanding. They also agree that understanding comes from theory rather than directly from data and that the way to demonstrate that a phenomenon is truly (correctly) understood is to test predictions derived from a theory. This characterization fits mechanistic sciences, in which confirmation of a prediction is interpreted to indicate *correspondence* between theory and reality, which is the mechanistic truth criterion. However, it does not fit contextualistic sciences and it is therefore irrelevant to behavior analysis if behavior analysis is a contextualistic science.

Is behavior analysis a contextualistic science? Sidman (1990) said:

> I have not seen that contextualism adds anything not already handled by the concept of contingency... I have yet to see a single technique of experimental analysis, a piece of experimental data, or a method of data evaluation whose origination required a new kind of contextual orientation. Nor do I find the "new contextualism" generating practical solutions to any of the many real problems that are inherent in the approaches *Tactics [of Scientific Research]* described. (p. 194)

Although Sidman's point was that contextualism adds nothing new, his remarks actually indicate that behavior analysis is entirely consistent with contextualism. Contextualism could have added something new only if behavior analysis was *not* already thoroughly contextualistic.

The emphasis on control in behavior analysis is consistent with the contextualistic truth criterion: Truth is demonstrated by successful practice, which means success in attaining an end--not just any end but an expected goal that was sought after. In behavior analysis, successful control provides evidence that the goal of understanding has been attained, but a criterion for success is needed because success is always defined in terms of some specific purpose. In other words, unless the purpose of an explanation is specified, the truth of the explanation cannot be demonstrated. The purpose could be a whim, but scientific purposes are generally held to be principled rather than whimsical. Facts about behavior can provide the principles. These facts are not completely dispersed but rather exhibit interrelations that permit organizing them into theories of at least the sort Melvin Marx (1976) called "inductive" (in which the theoretical concepts are directly related to observations rather than several steps removed from observations, as in "deductive" theories). Consequently, understanding is not *equivalent* to control but rather understanding is *tested* by control. However, control provides this test only if the way to control the phenomenon under study is predicted by the theory that provides the understanding.

Control

"Control" is a complicated topic, as Morris et al. (1993) showed, and only one kind--experimental control--is discussed here. "Experimental control" does not mean control of a subject's behavior, it means control of the occurrences of a controlling variable, which is the variable that actually controls the subject's behavior. So why does the *experimenter* have to control the controlling variable? Why not just study natural occurrences of a suspected controlling variable? The received answer is that the study would be correlational and could not demonstrate that the suspected controlling variable was the actual controlling variable.

Actually, this answer is not entirely correct. In principle, correlational research demonstrates causality if all variables that co-occur with the suspected controlling variable are known to have no direct effect on behavior and are known not to influence the effect of the suspected controlling variable. Of course, no one knows that much, so in practice correlational research does not demonstrate causality. Actually, that is not entirely true either. Experimental research can demonstrate control even though experimental research is correlational. It is correlational because the aim is to demonstrate covariation (correlation) between behavior and treatments.

Nevertheless, if "internal validity" (e.g., Perone, 1994) is demonstrated, an experimental design is still preferable to a correlational design in the strict sense. But why use a single-subject experimental design? Why not use a group design with the suspected controlling variable administered to an experimental group and some baseline or placebo variable administered to a control group, or with both variables administered to a single group? A frequent answer is that group curves can obscure the performance of individual subjects, as for example Skinner (1956) and Sidman (1960, pp. 46-51) said and as group researchers also said (e.g., Hayes & Pereboom, 1959; Hilgard & Campbell, 1937; Spence, 1956, pp. 59-61). In fact, group curves are mathematically incapable of representing some kinds of individual curves (Sidman, 1952). However, this answer is often a minor point when it is relevant and it is not always relevant. The issue is discussed later.

Prediction

Role of control. The kind of prediction wanted in behavior analysis involves control (Skinner, 1953, p. 200). Verification of a prediction can be used to define successful understanding. If the prediction is from cause to effect, as Skinner said (1953, p. 199), or as Morris et al. (1993) said, control is successful in a "deep" sense if it leads to discovery of controlling variables. Two questions are whether verification of some kinds of prediction is *not* acceptable evidence for success and whether successful control can be demonstrated only in single-subject research.

Skinner (1953, p. 199) and Morris et al. (1993) said that the kind of prediction involved in testing the truth of understanding is theory-based. The principle is that the relevant kind of prediction involves a scientific theory of some type, but any type of scientific theory will do. It can be an hypothetico-deductive type of theory with intervening variables and hypothetical constructs, or an inductive type, or any other type, so long as it is a *scientific* theory (as defined by, e.g., Marx, 1976; Popper, 1983, p. 288).

Other meanings of "prediction." The word "prediction" has also been used to refer to other kinds of expectation or forecast. Examples are predicting later darkness from present daylight, later summer from present spring, and IQ in adulthood from IQ in childhood. Another example is in a report about a computer program called "MOSAIC-20" that is said to predict "which abusers are most likely to kill their spouses" (Ciabattari, 1997). After the fact, this program showed that O. J. Simpson would have obtained the maximum score predicting the likelihood of homicide. Predictions such as these are based on correlations between effects; they require knowing *that* trends occur rather than knowing *why* they occur. In other words, this kind of prediction does not require theoretical understanding.

This kind of prediction is not directly relevant to behavior analysis because, as Skinner (1953, p. 199) pointed out, it is prediction from one effect to another effect rather than prediction from a cause to an effect. Predictions from effect to effect are not the same as predictions on the basis of a functional analysis; but nevertheless, if the organism's history cannot be assessed, predictions from effect to effect can be useful (Skinner, 1953, pp. 199-200). Even when they are useful, however, such predictions do not further the *control* of behavior (Skinner, 1953, p. 200). Therefore, verifying these predictions cannot further the understanding of behavior, though it can further the effectiveness of dealing with particular individuals in everyday situations. This point is discussed later.

Relevance of "falsificationism." Popper (1983, passim) said that science relies on falsification rather than verification. However, he was talking about logic or what he thought scientists *should* do, not what they actually do. No scientist, except perhaps a masochistic one, hopes for falsification of any predictions except those derived from somebody else's theory. All sane scientists hope for verification of their own predictions. Also, contrary to Lakatos (1978, p. 6), scientists seldom reject their own theory when a prediction is falsified; they try to find the faulty part and fix it. According to Kuhn (1970) and Laudan (1977, pp. 114-118), the history of science fully justifies this procedure.

Predicting an Individual's Behavior

Skinner (1953) said that uniformities are often best found through examination of individuals and furthermore that "A prediction of what the *average* individual will do is often of little or no value in dealing with a particular individual" (p. 19). I would add, however, that dealing effectively with a particular individual in a particular everyday situation does not usually depend on knowing that individual's history of reinforcement--fortunately so because this history is seldom if ever known in sufficient detail to base a prediction on it.

The situation is not hopeless with respect to predicting an individual's behavior, but prediction for the individual will usually and perhaps always be like weather forecasting because many of the variables cannot be controlled. The problem is that weather forecasting is exactly analogous to what Skinner said is not wanted: "No one goes to the circus to see the average dog jump through the hoop significantly oftener than untrained dogs raised under the same circumstances" (Skinner, 1956, p. 228). True, but neither does anyone go to the circus to see a dog behave in a single-subject reversal experiment nor, as Skinner (1956, p. 228) said, "to see an elephant demonstrate a principle of behavior." What one sees at the circus is an animal such as a dog interacting with an apparatus such as a hoop in ways that are consistent with the definition of a response class such as "jumping through the hoop."

Problems of Group Research

Group Curves and Individual Performance

As noted earlier, group curves often do not and sometimes cannot accurately represent the performance of individual subjects, but this point is sometimes minor and it is sometimes irrelevant. The point is not minor in research on schedule control, where the pattern of behavior revealed in a cumulative record is the major issue. It is also not minor in research on stimulus equivalence and other kinds of emergent discriminations, where the desired pattern of behavior is a zero rate of responses in some conditions and a 100% rate in other conditions. In other research, however, the issue is often not *how much* change in behavior occurs across conditions but *whether* change occurs. In this research, if the behavior changes at all in all the subjects, nobody should care whether a group curve accurately represents the *amount* of change in individual subjects.

When group researchers or behavior analysts care about whether individual curves are consistent with the group curve, they can apply any of several available statistical tests. These tests must be used carefully because in essence they involve the "box score" method (Gardner, 1966; Reese, 1991, pp. 213-214), which is not necessarily a good method. The first step in the box-score method is to classify each individual curve as consistent, partially consistent, or inconsistent with the group curve or, more simply, as consistent versus inconsistent. The second step is to determine whether the "consistent" group is significantly larger than the other groups (a chi-square test could be used for this purpose). The final step is to conclude that the group curve is or is not the "true" curve, depending on whether the "consistent" group is or is not significantly larger than the other groups. The problem is that unless all the individual curves are consistent with the group curve, variability is interpreted as error of measurement and any true individual differences are ignored. However, when this problem occurs, it reflects a flaw in group researchers, not in group methodology.

Sidman (1990) identified another problem: The tests can indicate *how many* subjects are likely to exhibit a particular effect, but they cannot indicate *which ones* will exhibit the effect. Sidman's point is correct, but it is also applicable to single-subject studies in which not all subjects exhibit the effect. Except in clinical case history reports, single-subject studies virtually always include more than one subject in order to demonstrate generality (e.g., Perone, 1994), but many single-subject studies do not demonstrate control in every subject and do not indicate why control was not exhibited in the failures.

Some of these studies do not even achieve a level of control that Baer (1990) said is inadequate. As discussant at a symposium, he said that the experimental analysis in a particular study of stimulus equivalence was incomplete because stimulus equivalence was demonstrated in only seven of the eight subjects who were run. Failure to control the behavior of one or a few of the subjects should not lead initially to rejection of the theory underlying a study. Rather, as Baer suggested and as indicated in the discussion of falsificationism herein, such a failure should lead initially to further analyses aimed at identifying the variables that interfered with the manipulated ones.

Control of Variability

Another criticism of group research is that group researchers do not attempt to control variability in each subject experimentally. However, this is again a criticism of what group researchers tend to do, not a criticism of group methodology. Behavior analysts do not necessarily have to give up this kind of control when they use group designs. In the best group research, each subject is run individually.

Regarding group research and the issue of control, I would say that if a phenomenon can be demonstrated in an individual subject, it can be demonstrated in a group. Only relatively powerful controlling variables can be demonstrated to be effective in an individual subject; therefore, if control can be demonstrated in an individual subject, it can especially easily be demonstrated in a group. Single-subject *studies* virtually always include more than one subject in order to demonstrate generality, as mentioned earlier, and a phenomenon demonstrated by visual analysis in such studies is very likely to be significant in a statistical analysis of the data (Perone, 1994). For example, in a study by Sloane, Young, and Marcusen (1977, Exp. 1), noncontingent response costs were demonstrated to affect aggressive behavior in individual subjects and in the group as a whole. In contrast, in a study by Duarte and Baer (1997), effects of sudden versus gradual stimulus change varied across individual subjects and were not significant in the group as a whole.

Perone (1994) argued that single-subject designs are superior to within-subject group designs because, as Sidman (1960, chap. 9) said, single-subject designs require continuing each treatment phase until a steady state is demonstrated. However, steady states are also required in group studies on certain topics. Examples are studies of (a) learning when each subject is given training until reaching a learning criterion, (b) perception when a stimulus is changed only after the subject (usually an infant) looks away from it for a criterion duration, and (c) cognition when the subject receives training in use of a "strategy" (e.g., "mental rotation," a mnemonic technique) until performance reaches an asymptote.

I see no important difference between covariation between behavior and variables manipulated within a subject and covariation between behavior and variables manipulated between subjects, if variability between subjects is taken into account. In fact, as Perone (1994) pointed out, some of the between-subject variability may not need to be taken into account if the subjects are interpreted to be replications in a systematic replication design (Sidman, 1960, chap. 4). The only problem, then, is to avoid sloppiness in doing group studies, but avoiding sloppiness is also a problem in doing single-subject studies.

Some single-subject researchers use single-subject rules loosely. An example is that smoothing a somewhat erratic curve by averaging across time or trials is legitimate if the aim is to make the functional relation clearer--Skinner (1956) and Sidman (1990) endorsed this procedure--but it is not legitimate if the aim is to smooth a highly erratic curve to hide incomplete control of variability (Sidman, 1960, pp. 273-278). Variability can also be obscured by the selection of measures, as Sidman (1990) pointed out, and it is ignored when partial control is accepted, that is, when an effect is not demonstrated in all the subjects.

Demonstrability of an Effect

Sidman (1960) gave yet another reason for not using group designs: If a relation cannot be demonstrated in an individual subject, it does not exist. Specifically, "If it proves impossible to obtain an uncontaminated relation between number of reinforcements and resistance to extinction in a single subject, because of the fact that successive extinctions interfere with each other, then the 'pure' relation simply does not exist" (p. 53). In referring to the *existence* of the relation, Sidman was making an ontological point, but I think the proper point is epistemological--a limit on the kind of knowledge about the relation that can be derived from single-subject research.

Interpreted as epistemological, Sidman's point is correct, but the relation in question could be tested with a standard behavior-analytic design--a multiple-baseline design with a different behavior for each conditioning and extinction cycle or with a different subject for each cycle. Properly done, the multiple-baseline design is a legitimate analogue of the reversal design.

Group research might also yield relevant evidence. If groups of individuals differ in the number of reinforcements of responses comprising a particular operant and differ in no other way that is related to resistance to extinction, and if the resistance to extinction of this operant is found to be greater across groups, the greater the number of reinforcements, then the differences in resistance to extinction are uniquely attributable to the different numbers of reinforcements. In this case, the relation in question is shown to exist in nature because all the individual resistances to extinction fall precisely on the curve representing the relation. In the real world, of course, no groups with the required characteristics have been found, but the differences in number of reinforcements might still be relatively much more important than any other individual differences. Even if not one of the data points falls precisely on the curve for the group as a whole, all the data points could lie fairly close to this curve. The issue then becomes what is meant by "fairly close."

I think everyone would agree that if the group curve is drawn to occupy a full 8 1/2 by 11 inch page and deviations of the data points from the curve can be seen only with a magnifying glass, the data points are "fairly close" to the curve. Agreement would decrease, the more easily the deviations could be seen with the naked eye. Statistical tests such as the test for departure from linearity can provide an objective answer to the question of whether the deviations are relatively small. Saying that the deviations are "relatively small" is the same as saying that: (a) the data points are "fairly close" to the group curve, (b) the group curve represents the data points "reasonably accurately," and (c) the group curve fits the data points "reasonably well."

The Nature of Evidence

Sidman (1990) quoted a statement from Mecca Chiesa's doctoral dissertation to the effect that inferences from "accumulated observation" can be changed only by new evidence, but that group researchers can change their inferences by the simple procedure of changing the level of confidence that defines "significant." Sidman said that this difference gives an advantage to inductive inference over statistical inference. However, two considerations undermine Chiesa's point. First, the value of "accumulated observation" is as great for group researchers as for single-subject researchers. Second, if any group researchers use the trick of messing with the level of confidence, it is a flaw in their behavior and the behavior of journal editors rather than a flaw in group methodology as such.

Editors who are careful and knowledgeable do not allow the use of a level of confidence--or level of significance or alpha, as it is also called--other than the conventional .05 unless a convincing rationale is explicitly given. For example, an alpha greater than .05 might be appropriate if the cost of a Type I error (rejecting a true null hypothesis) is far less than the cost of a Type II error (not rejecting a false null hypothesis); and an alpha less than .05 is likely to be appropriate if the sample size is so large that the statistical test has enough power to make psychologically negligible effects statistically significant, or if the number of comparisons to be made is so large that the probability of a Type I error for the study as a whole is larger than .05. A related point is that if alpha is .05, a strictly honest researcher will call probabilities greater than .05, such as .0501, "nonsignificant" and will use the phrase "marginally significant" only to refer to a finding that is significant at precisely the preselected alpha level, because alpha determines the *margin* separating significant from nonsignificant. A strict editor will require the use of these conventions, but many editors are not strict in this way.

Of course, setting alpha at .05 is arbitrary rather than sacred, but once it is set, it cannot be legitimately adjusted to suit the data. That is, if alpha is set at the conventional .05, it cannot be legitimately adjusted downward in order to make a desired effect "significant" nor upward to make an undesired effect "nonsignificant." Otherwise, statistical inference would lose its objectivity and become subjective.

CONCLUSION

In conclusion, many of the flaws attributed to group methodology by behavior analysts are actually flaws in group researchers who misuse the methodology. Perhaps group methodology induces these flaws in group researchers, along the lines of James Gibson's concept of "affordance" (summarized by Gibson, 1982). Maybe the situation is like one in cognitive psychology: Actually running a computer program dictated by an information-processing theory keeps the theory honest by demonstrating that the machine actually generates the output the theory is claimed to predict. Maybe single-subject research could be this kind of fail-safe mechanism for group researchers, as Baer (1988) suggested. In any case, if carefully and properly used, group methodology can further behavior analysts' understanding of behavior. All the criticisms of group methodology can be challenged, and none should be accepted without deep study. They are best interpreted as warnings about how group methodology can be misused rather than as indications of inherent flaws in this methodology.

REFERENCES

Baer, D. M. (1973). The control of developmental process: Why wait? In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues* (pp. 185-193). New York: Academic Press.

Baer, D. M. (1988). Guest editor's note. *Journal of Experimental Child Psychology, 46*, 287-288.

Baer, D. M. (1990, May). [Discussion.] In T. Verhave (Chair), *Extending the boundaries of equivalence classes*. Symposium conducted at the meeting of the Association for Behavior Analysis, Nashville, TN.

Baron, A. (1990). Experimental designs. *The Behavior Analyst, 13*, 167-171.

Ciabattari, J. (1997, May 18). Could a new computer program have saved Nicole Brown Simpson? *Parade Magazine*, p. 16.

Duarte, A. M. M., & Baer, D. M. (1997). Overselectivity in the naming of suddenly and gradually constructed faces. In D. M. Baer & E. M. Pinkston (Eds.), *Environment and behavior* (pp. 210-218). Boulder, CO: Westview Press.

Gardner, R. A. (1966). On box score methodology as illustrated by three reviews of overtraining reversal effects. *Psychological Bulletin, 66*, 416-418.

Gibson, E. J. (1982). The concept of affordances in development: The renascence of functionalism. In W. A. Collins (Ed.), *The concept of development: The Minnesota Symposia on Child Psychology* (Vol. 15, pp. 55-81). Hillsdale, NJ: Erlbaum.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Hayes, K. J., & Pereboom, A. C. (1959). Artifacts in criterion-reference learning curves. *Psychological Review, 66*, 23-26.

Hilgard, E. R., & Campbell, A. A. (1937). Vincent curves of conditioning. *Journal of Experimental Psychology, 21*, 310-319.

Kuhn, T. S. (1970). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 1-23). London: Cambridge University Press.

Lakatos, I. (1978). *The methodology of scientific research programmes*. In. I. Lakatos, *Philosophical papers* (Vol. 1: J. Worrall & G. Currie, Eds.). Cambridge: Cambridge University Press.

Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth*. Berkeley: University of California Press.

Marx, M. H. (1976). Formal theory. In M. H. Marx & F. E. Goodson (Eds.), *Theories in contemporary psychology* (2nd ed., pp. 234-260). New York: Macmillan.

Morris, E. K., Todd, J. T., & Midgley, B. D. (1993). The prediction and control of behavior: Watson, Skinner, and beyond. *Mexican Journal of Behavior Analysis, 19* (Monographic Issue 3), 103-131.

Neuringer, A. (1991). Humble behaviorism. *The Behavior Analyst, 14*, 1-13.

Neuringer, A. (1992). Choosing to vary and repeat. *Psychological Science, 3*, 246-250.

Perone, M. (1994). Single-subject designs and developmental psychology. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 95-118). Hillsdale, NJ: Erlbaum.
Popper, K. R. (1983). *Realism and the aim of science* (W. W. Bartley, III, Ed.). Totowa, NJ: Rowman and Littlefield.
Reese, H. W. (1991). Recommendations for graduate training in child psychology. In J. H. Cantor, C. C. Spiker, & L. P. Lipsitt (Eds.), *Child behavior and development: Training for diversity* (pp. 195-226). Norwood, NJ: Ablex.
Reese, H. W. (1996, October). *Prediction, control, and understanding in behavioral sciences*. Paper presented at the Third International Congress on Behaviorism and the Sciences of Behavior, Yokohama, Japan.
Russell, B. (1929). *Our knowledge of the external world* (2nd ed.). New York: Norton.
Sheldon, W. H., Stevens, S. S., & Tucker, W. B. (1940). *The varieties of human physique: An introduction to constitutional psychology*. New York: Harper & Brothers.
Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin, 49*, 263-269.
Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.
Sidman, M. (1990). *Tactics: In reply . . . . The Behavior Analyst, 13*, 187-197.
Simon, M. D. (1959). Body configuration and school readiness. *Child Development, 30*, 493-512.
Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
Skinner, B. F. (1956). A case history in scientific method. *American Psychologist, 11*, 221-233.
Skinner, B. F. (1989). The behavior of the listener. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 85-96). New York: Plenum.
Sloane, H. N., Jr., Young, K. R., & Marcusen, T. (1977). Response cost and human aggressive behavior. In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, method, and application* (pp. 531-542). Hillsdale, NJ: Erlbaum.
Snow, C. E. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review, 53*, 165-189.
Spence, K. W. (1956). *Behavior theory and conditioning*. New Haven, CT: Yale University Press.