Review of *Strategies and Tactics of Behavioral Research* (2nd ed.) by J. M. Johnston and H. S. Pennypacker

Joel Greenspoon
University of North Texas

The preface of a revised edition provides interesting reading, especially the justification for the revision. Johnston and Pennypacker provide an excellent rationale for their revised edition of *Strategies and Tactics of Behavioral Research*. The authors indicate that a major objective of the revision was to enhance the readability of the book. Although the content of the second edition is similar to that of the first edition, Johnston and Pennypacker made an effort to present the material in ways to increase the probability of controlling the desired behaviors of the reader. Because students with different histories may use the book in a variety of courses, the authors have tried to present material that will be compatible with these different histories. They deleted *human* from the title, because the strategies and tactics of behavioral research are, or at least should be, applicable to all investigations of behavior.

The table of contents reveals that both editions contain the first four parts. The fifth part of the first edition has been deleted. The second edition is somewhat shorter than the first edition in terms of number of pages. A separate readings supplement included with the second edition results in a somewhat larger text. Many chapter titles have been changed, and some chapters have been deleted from the second edition.

The inclusion of boxes to describe additional points or issues is an excellent idea. The study questions at the end of each chapter are very beneficial to stu-

dents (perhaps to instructors as well). The glossary is a very valuable addition to the second edition, especially for readers with a limited history of behavior analysis and research methodology. The white page background of the new edition provides a better contrast than the ivory background of the first edition, thus providing easier word discrimination.

This book is not designed to be a cookbook on how to do behavioral research. It is designed to introduce the reader to the many issues that must be considered and resolved in the process of developing and conducting experiments. The authors have done an excellent job of describing many of these issues. They have also provided a number of solutions to these issues. They make it clear that the solution to any issue involves considering many different factors. There is rarely just one solution to the issue. The authors have done an excellent job of presenting experimental research as a complicated process that requires considerable work. Research is not a do-it-by-the-numbers activity.

The first section of the book may be considered an introduction to how to do science, especially experimental science. The authors' approach is to examine the contingencies that are involved in doing science. They point out that it is very important for the experimenter's behavior to come under the control of the data, rather than being totally rule governed. The emphasis is on treating behavior as the phenomenon rather than an epiphenomenon. Experimentation should be concerned with environment-behavior interactions. They stress the necessity and importance of a definition of behavior that differentiates it from nonbehavior (a null class) and is amenable to measurement.

A rather long, complex definition of behavior is presented that stresses the in-

The full citation is Johnston, J. M., and Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum. ISBN 0-8058-0905-8. 393 pages, with accompanying set of readings.

Address correspondence concerning this article to Joel Greenspoon, Center for Behavior Analysis, University of North Texas, Denton, TX 76203.

teraction of the organism with the environment, detectable displacement in space through time of some part of the organism, and a measurable change in the environment. However, in discussing their definition of behavior, the authors state, "Because it may not fully describe what behavior *really* [italics added] is, the usefulness of this definition (or any definition of a natural phenomenon) lies in its effectiveness as a kind of interim guide or rule based on what we think we already know" (p. 23). Most scientists would have no trouble accepting the last part of this statement, but may have some trouble with the first part.

This definition of behavior may create some difficulties with feelings and thoughts as behaviors. Johnston and Pennypacker recognize these difficulties, but state that their definition does not exclude feelings and thoughts from the behavioral repertoire. They add, "this requirement for detectability does not really mean that if it cannot be measured it is not behavior" (p. 24). If it cannot be measured, researchers will have some difficulty investigating it, whatever "it" is. The authors say their definition is a working definition for researchers, not theoreticians. If the researchers and theoreticians operate from different definitional frameworks, a theory of behavior based on research findings might be extraordinarily difficult to develop.

With regard to data analysis, they advocate repeated measurements of behavior of the same organism and are averse to lumping together the measures of different subjects. They contend that lumping measurements depletes the purity of the phenomenon and makes it difficult to identify significant relationships.

Hypothesis testing as an experimental procedure is not viewed very favorably, because it may lead to biasing the procedures or design to produce the predicted outcome. This possibility is not limited to hypothesis testing. The absence of a well-developed theory is a greater handicap to meaningful hypothesis testing. Most hypothesis testing in the behavioral sciences is empirically based rather than theory based, and thus is a very weak form of hypothesis testing.

Each step in the process of developing an experiment is analyzed carefully. The dangers and pitfalls that an experimenter may encounter are clearly described. Suggested resolutions for the various problems that may arise in each step are provided, but always with the cautionary note that the experimenter must examine the resolution within the context of the problem and the particular situation. High-quality research benefits the individual, the discipline, and the culture.

The second part of the book is concerned with measurement. Some behavior analysts may be uncomfortable with their use of the terms "true" and "accurate" that appear frequently, suggesting that there is a true and accurate measurement. To say that one measure is more accurate than another suggests that it is possible to obtain the "ultimate" measurement, against which other measures can be compared. How to obtain this ultimate measurement is not clearly described. The authors indicate the difficulties in determining the true measure.

The issues of reliability and interobserver agreement are discussed, and reliability measures are seemingly considered to be better. Reliability measures also require interobserver agreement. If a scale generates a number that controls the emission of the same verbal behavior from two or more observers, we tend to say the scale is reliable. However, reliability is predicated on agreement among the two or more observers.

In the case of observers of classroom behavior, we describe the correspondence of the verbal behavior of the observers as interobserver agreement. There is a difference between these two kinds of interobserver agreement in that the behavior of the observer of the scale may be controlled by a more restricted set of variables than is the behavior of the observers in a classroom. The time frame in which the two kinds of observations are made may have a markedly different effect on each. The histories of both sets of observers are a source of behavioral control for both.

The issue of calibration poses similar problems. The researcher assumes that the instrument against which other in-

struments are calibrated provides a true and accurate measure. But the researcher has no way of testing that assumption. The process can lead to an infinite regress as a calibrating instrument is tested against another instrument, which is tested against another instrument, *ad infinitum* or *ad nauseam*.

Johnston and Pennypacker discuss a wide range of issues that the researcher must address in the process of developing an experiment. The issues involved in the selection of the unit of analysis and response classes are described with some specific helpful suggestions. A clear distinction is drawn between the functional and topographical response classes. Johnston and Pennypacker (and probably most behavior analysts) prefer the functional response class. They point out situations in which a topographical response class may be preferred and more appropriate. They warn against group behavior, because behavior is the interaction of the individual organism with the environment. Groups do not behave, and group data may obscure important relationships between the individual and the environment. They also warn against a nonbehavior class, because the organism is always doing something.

They describe the factors that should be considered in the process of selecting a unit of analysis that is most appropriate for the experimental question. Johnston and Pennypacker differentiate property and dimension. Property is a fundamental quality of a phenomenon, and dimension is a quantifiable aspect of the property. For example, weight is a fundamental property, and grams are quantifiable aspects of the property. Some behavior analysts may have some problems with this distinction. The authors call fundamental dimensions *idemnotic* and refer to them as *absolutes*. *Vaganotic* is their term for units that may vary. Thus, *vaganotic* may refer to different units that have the same label (e.g., IQ based on two different tests).

Counting as a measure does not seem to have the same status as measures such as latency and duration. Counting is the means by which latency and duration are obtained. We count the number of units

of time to determine the duration of a behavior. Key peck as a unit may not have the degree of generality across sciences of the second or the meter, but the key peck is just as legitimate a unit as the second, and counting them is just as legitimate as counting seconds. It is possible to develop a rather powerful argument that counting is the basis of all measurement, including the *idemnotic* measurements.

The discussion of observing and recording of behavior is excellent. They touch on many critical issues with respect to both of these very important activities in conducting research. The machine-human transducer issue decidedly favors the machine, although there are situations in which, at least at the present time, the human observer is the only transducer.

The authors emphasize direct measurement, the measurement of the phenomenon that is the focus of the experiment. Direct measurement in the final analysis is much simpler than indirect measurement, because many fewer assumptions are involved. Validity is not an issue with direct measurement. The strategies and tactics involved in the observation and recording of behavior are also described. The advantages and disadvantages of continuous and discontinuous recording are discussed in conjunction with the kinds of situations in which one or the other may be most appropriate.

The authors address a variety of issues about variability in the third section of the book. The advantages of considering variability as extrinsic rather than intrinsic are described in terms of desirability, predictability, and explainability, with reference to other phenomena, both organismic and environmental. The variability attributable to extraneous variables may be reduced through more rigorous control of experimental conditions. Variability attributable to the independent variable is highly desirable. The authors describe various sources and measurements of variability. They tend to favor the steady-state strategy of dealing with variability, because it provides greater opportunity to assess rigorously

and conservatively the effects of the independent variable. The steady state is defined as the absence of trend, and trend is defined as a consistent change in one direction. The advantages, as well as other issues relating to the steady state, are described very well.

The authors stress the important role that variability plays in research. Variability has much to do with the way in which the experimenter designs an experiment. The experimenter minimizes the undesirable effects of variability attributable to extraneous variables and maximizes the opportunity for the effects of the independent variable(s) to be effective.

The strategic and tactical issues involved in arranging experimental conditions are described very thoroughly and effectively. The authors' very conservative approach to experimental design is evident. The conservative approach is desirable for students being introduced to research. Johnston and Pennypacker note that experienced researchers may proceed somewhat differently. They also stress the importance of replication and describe several different types of replication. The functions of replication are discussed, with emphasis on replication for reliability or generality. This issue is pertinent to the type of replication that may be most appropriate. The number of subjects relates to replication, and the authors indicate that number *qua* number is not as critical as other factors to be considered in the replication.

The tactical issues in arranging experimental conditions are designed to implement the strategy that has been adopted. A major tactical issue involves arranging the conditions so that the effects of the independent variable can be most readily discovered by minimizing the effects of extraneous variables. They recommend caution in attributing causal properties to the independent variable. Tactically, the reliability–generality issue of replication involves the sequencing of the control and experimental conditions. If replication of control and experimental conditions is highly similar, the replica-

tion is designed for reliability. If the replication involves systematic changes in the conditions, the replication is designed for generality. The authors make this distinction very clear.

Johnston and Pennypacker provide a very interesting and important discussion of the arrangement of comparisons. The problems that may arise with the ABAB design are clearly described. Their stress on ways to control or evaluate the effects of extraneous variables seems well justified.

The advantages and problems associated with the single subject–single response–single setting design are clearly described. The advantages and problems associated with the multiple subject–multiple response class–multiple setting designs are also described in detail. The complexities and intricacies of these designs are described with emphasis on the requirements for making comparisons across subjects, response classes, and settings. One point that is frequently made is that it is always desirable to maximize the number of comparisons of the same condition, whether with single or multiple subjects, single or multiple response classes, or single or multiple settings.

The final section of the book is concerned with interpretation. Basic to interpretation is data analysis. Johnston and Pennypacker discuss data analysis within the context of stimulus control. That is, the data should control the analysis so that support for interpretation will be maximized. They advise sticking closely to the original data and minimizing transformations. If data transformations are performed, they should serve a specific, clearly described function. Summarizing data within or across subjects should be oriented toward providing information to answer the experimental question.

The authors prefer graphic analysis to statistical analysis. They describe problems, pitfalls, and suggested solutions to graphic analysis. They also recognize that there are situations in which descriptive statistical analyses may be desirable and appropriate. Inferential statistics have

little or no utility as a form of data analysis for the kind of experimental methodology described in this book.

After the data analyses are completed, the interpretation of the experiment is considered. Johnston and Pennypacker emphasize that interpretation should not be limited to the data and the data analysis. Instead, the interpretation should include all of the components of the experiment. Although this is a reasonable suggestion, it is frequently overlooked by many experimenters. Many factors may affect one's interpretation of an experiment. The theoretical biases of the experimenter, the related literature, extraexperimental contingencies, the experimental question, and specific components of the experiment itself all affect the interpretation. The experimental question should function as an important source of control over the interpretation.

The generality of the interpretation is also subject to many sources of control. Generality of interpretation may extend across several dimensions (e.g., subjects, response classes, settings, and species). The generality of variables involved may be extended across a number of dimensions, including methods and procedures.

Evidence for generality may arise from a number of sources. The multiplicity of conditions and the number of subjects may support the generality of the interpretation. Evaluating an interpretation is

very difficult and can vary from one researcher to another. The experimenter who provides the greatest amount of detail about the experiment to support the interpretation will probably generate more acceptance than one who provides very skimpy details.

The redundancy in the book is probably necessary to introduce the novice to the vagaries and complexities of conducting research within a behavior-analytic framework. Because the beginning student may have had little exposure to behavior-analytic research, the repetition of the explanations of various aspects of methodology may serve a valuable function.

An overall assessment of this book is that it provides the beginning researcher with an excellent introduction to the trials, tribulations, and rewards of conducting experiments. The authors make it abundantly clear that high-quality research makes great demands on the experimenter. They also make it clear that as the researcher acquires more and more experience in conducting experiments, he or she will quickly recognize the problems and develop his or her own solutions to these problems. Despite philosophical differences that the reader may have, I believe that most experimenters can become better experimenters by familiarizing themselves with the content of this book.