

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: <http://www.nber.org/books/kmen80-1>

Publication Date: 1980

Chapter Title: Token Economy and Animal Models for the Experimental Analysis of Economic Behavior

Chapter Author: John H. Kagel, Raymond C. Battalio

Chapter URL: <http://www.nber.org/chapters/c11711>

Chapter pages in book: (p. 379 - 401)

Token Economy and Animal Models for the Experimental Analysis of Economic Behavior

JOHN H. KAGEL

and

RAYMOND C. BATTALIO

DEPARTMENT OF ECONOMICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS

Introduction

During the past several years our research interest in individual economic behavior has resulted in conducting experiments in token economies with human subjects and in operant conditioning chambers using laboratory animals. In the context of the present volume these experimental methodologies may be viewed as an econometric research strategy—a method for gathering data to analyse statistically economic theories and relationships. The primary distinguishing characteristic of this research strategy lies in the use of data collected by the investigator, rather than relying on data collected as an input to, or by-product of, government administrative procedures. This has two advantages. First, collecting one's own data frequently provides the least cost method, sometimes the only method, for obtaining observations adequate to investigating certain hypotheses. Second, using experimental methods the investigator has the ability to manipulate environmental conditions to provide as precise and demanding tests of the hypotheses in

question as ethical and legal restrictions, and the technologies at his disposal, allow. In short, experimental methods complement existing econometric methods for evaluating economic theories by providing an observational basis for examining aspects of those theories which could not be done using routinely available data or which would be considerably more expensive to evaluate using nonexperimental methods.

This paper sets out to explicate, by example, this use of token economy and laboratory animal experiments. Emphasis will be placed on methodological issues which arise in using these technologies in economic research rather than on any comprehensive review of experimental results. The first, and major, part of the paper concerns the use of these methodologies in studying theories of individual subject behavior. Following this, we indicate how these research methods can be used to study directly aggregate economic behavior and policy orientated research questions.

1. Experimental Studies of Individual Behavior

The Slutsky–Hicks theory of commodity demand and labor supply behavior is integral to much of economic analysis. While this is a theory of individual subject behavior, the theory has only infrequently been investigated using individual subject data as the unit of analysis. A proximate cause of the lack of empirical research in this area is that the vast majority of existing observations of economic activity consist of cross-sectional data or aggregate time series data. Token economy experiments and studies of laboratory animal behavior provide unique opportunities to collect individual subject data suitable to investigating these theories.¹

Figure 1 illustrates this point. Each panel shows the labor supply curve for an *individual worker*, a rat working for alcohol in an operant conditioning chamber (a) and a human working for alcohol in a simplified token economy (b). Although economic textbooks and journals are replete with hypothetical illustrations of backward bending labor supply functions, the data in Fig. 1 are, to our knowledge, the first exemplification of such a relationship using data for individual workers to appear in literature.

¹ Experimentation in token economies must satisfy established ethical and legal standards for studies with human subjects (Rivlin & Timpane, 1975). Universities and funding agencies have institutional structures to enforce these standards. This will restrict certain economic experiments to nonhuman populations (Kagel *et al.*, 1975). Research on economic processes in token systems also has an important role to play in enhancing the therapeutic effectiveness of these systems (Winkler, 1971).

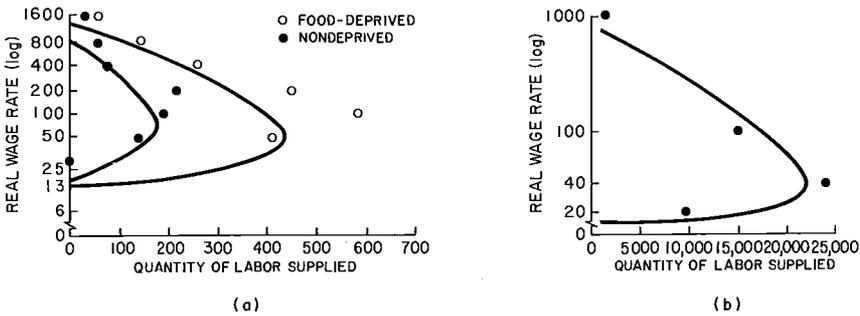


Fig. 1. Labor supply behavior of a rat (a) and a human (b) working for alcohol. Data for the rat are from Meisch & Thompson (1973). Human data are from Bieglow and Leibson (1972). In both (a) and (b) curves were sketched freehand to indicate their characteristic shape. Wage rates have been indexed with a value of 100 assigned to the midpoint of the ratio requirements imposed.

1.1. ON BRINGING BEHAVIOR INTO THE LABORATORY

An immediate question which arises with respect to the data reported in Fig. 1 is whether or not it legitimately represents a labor supply function; i.e., how does one translate the behavior of everyday life into the laboratory? In answer to this we note, as a general principle, that it is not necessary, nor is it desirable in many cases, to establish a direct physical analog between behavior in the laboratory and outside it. Rather, what is required is to develop a process within the laboratory with all the essential properties of the economic behavior under investigation and then to study the effects of alterations in parameter values on that behavior (Sidman, 1960). Technically the establishment of such a laboratory environment corresponds to the development of a formal empirical interpretation system (Basmann, 1975) or a definite set of semantic rules (Brunner, 1967) for the uninterpreted theory in question. Any "pure" economic theory permits a number of alternative interpretation systems, no one of which is *a priori* correct. The sole restriction is that the interpretation system does not introduce logical inconsistencies into the theory.² Changes in the interpretation system may alter test outcomes

² See Brunner (1967) for an example of where the semantic rules themselves introduce inconsistencies into the theory. In this context it should be mentioned that the problems encountered in developing a "pure" economic theory to the point where an economically efficient empirical interpretation system can be operationalized and competing hypotheses tested against each other are not unique to using nonexperimental data, but frequently are the most difficult problems in designing an experiment also. The fact is that many extant presentations of "pure" economic theories have been left very crudely formulated or have been formulated in such a way that they are virtually immune from falsification (Basmann, 1975). For citation of similar problems in behavioral biology, see Wilson (1975, pp. 28-29).

or estimated parameter values in the same way that modifications of the assumptions of the "pure" theory may alter its deductive implications. It is, in fact, the task of the decentralized research process to explore the effects of alterations in these rules, without which the "pure" or uninterpreted theory says nothing about the world.

The procedures underlying the data reported in Fig. 1 provide an example of such an empirical interpretation system. In both experiments value was induced on an arbitrary job task—lever pressing—through giving subjects with strong preferences for alcohol access to limited amounts of this commodity as a result of performing the job task a predetermined number of times. This experimentally induced constraint on behavior can be written as

$$C = rW, \quad (1)$$

where the amount of work performed (number of lever presses) is denoted by W , r is the experimenter determined wage rate (the inverse of the number of lever presses required per ounce of alcohol consumed), and C is total alcohol consumption or total income earnings in this world with one economic good.

Since the length of the experimental sessions was time determined, subjects faced a second constraint

$$T = L + \hat{W}, \quad (2)$$

where T is total time of the experimental session, \hat{W} is time spent lever pressing, and L is time spent in "leisure" activities (time not spent lever pressing). Since the amount of time taken to complete a level press can for all practical purposes be treated as a constant α ,

$$\hat{W} = \alpha W \quad (3)$$

Assuming subjects' preferences between income and leisure can be represented by a quasi-concave utility function with the usual properties, the constraints (1)–(3) establish a basis for investigating labor supply behavior in the laboratory. Typical labor supply functions resulting from such investigations are reported in Fig. 1. Figure 2 shows the relationship between the demand for income and effort price ($1/r$) corresponding to each of the labor supply functions in Fig. 1. The inverse relationship between income and effort price shown is a well established law of animal behavior (Battalio, Kagel, & Green, 1979; Allison, 1979).³

³ See Allison (1979) and Staddon (1979) for additional results taken from the literature in experimental psychology.

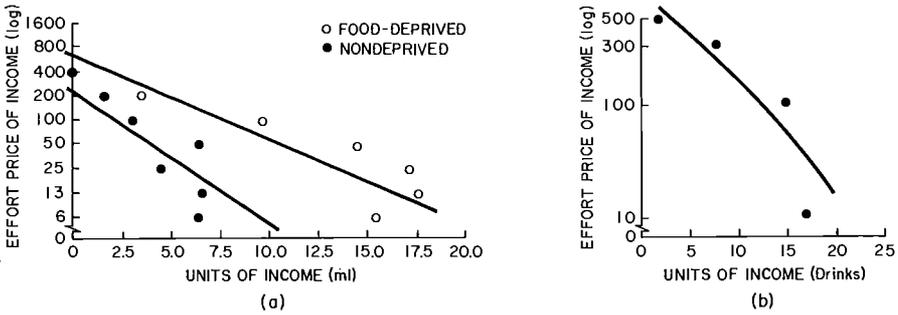


Fig. 2. Demand for income as a function of its effort price for the rat reported in Fig. 1a and the human reported in Fig. 1b. Curves sketched by freehand.

It is quite clear that the interpretation system developed above characterizes a highly simplified labor-leisure choice process—a one-commodity, one-job world. Although there may be no exact physical analog to these procedures outside the laboratory, they contain all the essential characteristics needed to define a logically consistent empirical interpretation system for a simple Slutsky-Hicks labor supply model. As such, they can be employed to test important properties of the model such as the negativity of the Slutsky substitution effect, the common assumption that leisure (not working) is a superior good throughout the choice set, and Robbins' (1930) suggestion that the demand for income will be inverse to its effort price. (Some initial outcomes of these investigations and their relationship to more familiar econometric studies using aggregate per capita or cross-sectional data are discussed in the following sections.)

The process of developing an experimental analog to an economic theory is similar to the development of the theory itself; start with something simple and then add more complex characteristics to the system and see how behavior is altered.⁴ While a fairly obvious basis exists for adding additional dimensions to the labor-leisure choice problem for humans, a similar basis exists for subhumans also. In both cases a number of alternative or multiple job tasks can be introduced into the situation, including ones with nonpecuniary returns (nonzero response rates in the absence of explicit payoffs), see Battalio *et al.* (1979). In addition, highly preferred (or multiple) commodities can be introduced, including nonfood items, and one can extend the analysis to wage rate schedules other than the simple piecework schedules underlying the data in Figs. 1 and 2 (Battalio *et al.*, 1979).

⁴ Budgetary considerations also dictate proceeding in this way since increasing the complexity of these experiments would increase both their time and dollar costs.

1.2. METHODOLOGICAL ASIDE: WHY STUDY THE ECONOMIC BEHAVIOR OF NONHUMANS?

The application of experimental methods to the study of nonhuman economic behavior is predicated on a number of factors. It has precedence in comparative studies of animal behavior which are well established fields in psychology and biology. One important outcome of these research programs has been to indicate that the laws and principles of economic behavior would be virtually unique if they did not apply, with some variation of course, to nonhumans as well as humans. Thus we proceed under the assumption that there is behavioral, as well as physiological, continuity across species, and if we identify genuine instances where this continuity breaks down, we shall have obtained a great deal of insight into the functional, or evolutionary, basis of the behavior in question.

This continuity in behavioral processes across species enables us to exploit the fact that the economic cost of experiments with nonhumans is considerably lower than the cost of comparable research with humans. Figure 1 provides a partial indication of this, where less than half as many data points are found for the human subject as compared to those for the rat.⁵ Further, experiments using animals permit a degree of control and manipulation of experimental conditions that may be necessary for investigating some hypotheses but which are unethical or illegal when applied to humans. For example, tests of the hypothesis put forward by Stigler & Becker (1977) that differences in "tastes" between individuals at a point in time are a function of differences in behavioral histories (analyzable within a neo-classical economic framework), rather than differences in genetic makeup, require enforced separation of parents from offspring. Such studies can only be performed using laboratory animals.

One might still argue against the relevance of subhumans as models for the experimental analysis of economic theories on the grounds that these theories postulate an element of calculating optimization that only humans can perform. We do not argue that humans can consciously calculate. However, it is not clear that one can demonstrate, observationally, that such calculations actually underlie the behavior that the Slutsky-Hicks theory attempts to characterize or that subhumans do not possess similar calculating abilities, albeit on a much more primitive level. Further, a number of prominent economic theorists have recognized that the ability to calculate consciously is irrelevant to characterizing behavior as a solution to a constrained

⁵ Although exact figures are not available, a conservative estimate would place the cost of experiments using human subjects at four times that of comparable experiments using nonhumans.

optimization problem

... it is possible to formulate our conditions of equilibrium as those of an extremum problem, even though it is admittedly not a case of an individual's behaving in a maximizing manner, just as it is often possible in classical dynamics to express the path of a particle as one which maximizes (minimizes) some quantity despite the fact that the particle is obviously not acting consciously or purposively (Samuelson, 1947, p. 23).

The only question that one can answer observationally is under what conditions do the predictions of our theories seem to accord with the "facts" of behavior.⁶ We turn to this use of token economy and laboratory animal experiments in the following sections.

1.3. USE OF INDIVIDUAL OBSERVATIONS IN HYPOTHESIS TESTING AND THEORY DEVELOPMENT

A number of well-known problems are encountered in testing the Slutsky-Hicks theory using aggregate per capita or cross-sectional data from national economies. First, it is readily demonstrated that aggregation problems in going from individual to aggregate commodity demand or labor supply functions are such that even if all individuals were to behave in accordance with the Slutsky-Hicks theory, the aggregate data need not (Hicks, 1956, chap. 6). This means that falsification or confirmation of the Slutsky-Hicks restrictions on functions applied to such data cannot logically falsify or verify the theory of individual behavior underlying it without assuming that a rather severe set of additional restrictions are met (Gorman, 1953; Green, 1964; Muellbauer, 1975). Assuming for the moment that these additional restrictions are satisfied so that Slutsky-Hicks properties apply to the aggregate data, important problems in testing the theory still remain. One of the most important of these is that some model specification is needed to make computation possible, thereby making it impossible to be sure that the tests performed are those of the assumptions underlying the general theory rather than tests of the particular utility function used to derive (or implied by) the estimating equations (Phlips, 1975; Basman, Battalio, & Kagel, 1973).

⁶ The problem of sorting out between calculating optimization and "as if" optimization comes down to specifying those predictions of the theory which hold up in the one case but not in the other, and then observing what the animals (both human and subhuman) actually do. We are unaware of any efforts along these lines to date.

Use of individual subject data as the unit of analysis avoids the first of the problems mentioned. Appropriate manipulations of experimental conditions then permit testing necessary conditions which must be satisfied if any member of the class of Slutsky–Hicks utility functions is to characterize the data.^{7,8} Examples follow.

Figure 3 reports some initial tests of the negativity of the Slutsky-substitution effect (equivalently, tests of the weak axiom of revealed preference) in the context of the income–leisure interpretation system developed in Section 1.1.⁹ The point labeled A on budget line K° in each panel shows the income–leisure choice of four pigeons under the same initial (baseline) wage rate schedule. Following the establishment of stable baseline conditions, each bird was given one half of its average baseline income earnings in the form of nonwage payments programmed for delivery throughout the session. At the same time, wage rates were halved so that the new budget line would pass through the original equilibrium point A, resulting in budget line K' in each case. For all subjects the equilibrium income–leisure point, shown by the letter B, occurred somewhere along the dashed portion of K' , as it must if the choice process in question is to be characterized by a member of the class of Slutsky–Hicks theories. Tests in addition to those reported in Fig. 3 have been conducted at four different baseline wage rates and with different levels of nonwage income. These tests continue to show negativity of the Slutsky-substitution effect at all but the very highest wage rate where there appears to be either a great deal of instability in the labor supply schedule itself or some systematic alteration of subjects preferences occurring as a result of experiencing the wage rate in question (unpublished data).¹⁰

⁷ These tests do not involve any approximations to some unspecified utility function, as do some of the proposed tests of Slutsky–Hicks restrictions reported in the literature. See Basmann *et al.*, (1973) and Philips (1974, Chap. 2) for discussion of the problems such approximations introduce into the analysis.

⁸ Experiments in preference theory using individual subject data as the unit of observation are not new to economics (May, 1954; Rousseas & Hart, 1951; Thurstone, 1931). What is new about our experiments is that the technologies employed result in the commodities and/or jobs in the choice set being an integral part of the ongoing activities of subjects for reasonably long periods of time. This automatically induces nontrivial values on the outcomes of individual responses to the experimental contingencies, an important element in effectively designing economic experiments (Siegel, 1961; Smith, 1976) and a serious deficiency in earlier experimental studies (see McCrimmon & Toda, 1969, for citations to this earlier literature).

⁹ These experiments are being conducted in collaboration with Professor Len Green of the Department of Psychology at Washington University.

¹⁰ This is no easy point to sort out. It takes approximately 3 to 3½ months to obtain each data point in this region of the choice space while maintaining subjects at a constant (80% of normal) body weight.

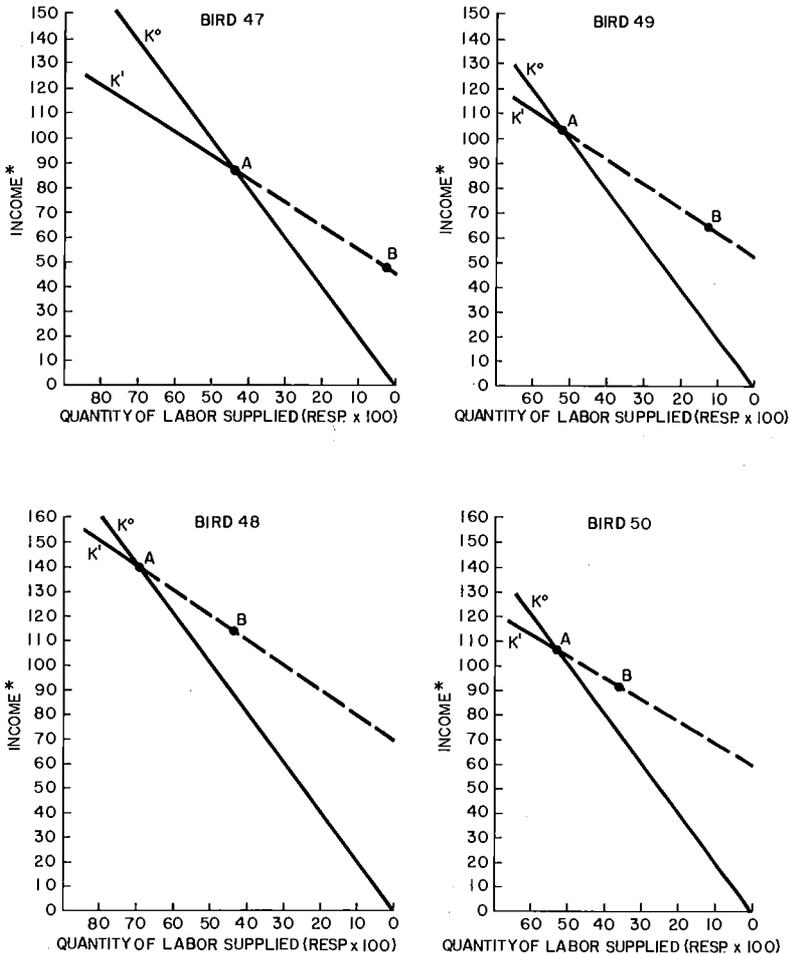


Fig. 3. Effects of income compensated wage rate changes on labor supply. Closed circles are mean choices over the last 5 days of an experimental condition. Standard errors of the mean are too small to graph, averaging between 1–2% of mean income. (*, 3-second accesses to food hopper) [Redrawn from Battalio, Kagel, & Green (1979, Fig. 6).]

Additional tests of the Slutsky–Hicks theory are reported by Allison, Miller, & Wozny (1979, Experiment 3), where the authors examine whether or not the labor supply schedule is affected by changes in nominal wage rates that leave the real wage rate unaffected. Using rats as subjects, three nominal wage rates were studied—5 lever presses for 5 licks at a water tube, 10 lever

presses for 10 licks, and 20 lever presses for 20 licks—all of which leave the wage rate unaffected. When travel time between the lever and water tube was negligible, Allison *et al.* report no significant differences in quantity of labor supplied (response rates) with changes in the ratio of lever presses to licks.

Having established that some of the Slutsky–Hicks properties hold for individual labor supply functions under a given interpretation system provides a basis upon which to conduct additional studies. These fall into several broad categories:

- (1) tests of additional restrictions which must hold if members of the class of Slutsky–Hicks theories are to organize the data;
- (2) establishment of lawful relationships between variables which are not deducible on the basis of the Slutsky–Hicks restrictions alone (Battalio *et al.*, 1979);
- (3) identification of particular utility functions for which simple changes in parameter values will organize large bodies of experimental data across different species (Staddon, 1979; Rachlin, 1978);
- (4) exploration of the effects of alterations in the interpretation system, i.e., alterations in experimental procedures, on behavior; and
- (5) contrasts of the economic analysis of the experimental outcomes with competing psychological explanations in an effort to distinguish between explanatory frameworks (Rachlin & Burkhard, 1978; Staddon & Motheral, 1978; Heyman & Luce, 1979).

In the process of exploring experimental analogs to the Slutsky–Hicks theory a number of experimental results have been identified which current theoretical formulations fail to handle. For example, in studying consumer demand behavior we have adhered to ABA types of experimental designs (A is the baseline or some initial set of experimental conditions, B a change in these conditions, and A a return to baseline conditions) which permit a comparison of steady state consumption patterns between baseline periods. While consumption patterns are typically quite similar between baseline conditions as compared to the intervening B condition, frequently they do not revert completely, suggesting that habits developed during the B period carried over into the following A period (Battalio *et al.*, 1973; Kagel *et al.*, 1975). An example of this is reported in Fig. 4. Although we have not had to search far for economic theories which suggest such irreversibilities in consumption patterns (Marshall, 1920, Appendix H; Georgescu-Roegen, 1966, Chap. 2), they are by no means universal (Kagel *et al.*, 1975). Further, we have not, at present, identified the set of initial conditions which will reliably generate irreversibilities, and the theoret-

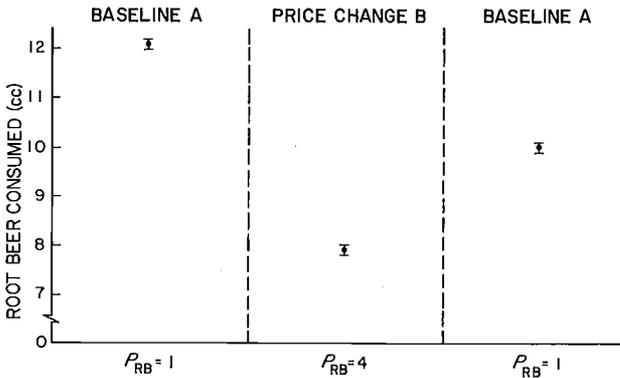


Fig. 4. Effects of income compensated price changes on root beer consumption for a rat choosing between root beer and Tom Collins mix. P_{RB} gives the relative price of root beer. Closed circles are mean consumption values for the last 10 days of each condition, with brackets showing standard error of the mean. Consumption between A periods is significantly different at the .01 level using a two-tailed t -test ($t = -3.13$, degrees of freedom = 18). [Data from Kagel *et al.* (1975).]

ical literature provides few hints as to what these conditions might be. In addition, looking at the day-to-day adjustments in consumption patterns following a change in experimental conditions we find that concepts of habit formation (Phlips, 1974) do not consistently organize these adjustment paths. In fact, at times just the opposite seems to occur, with the animal initially consuming more of the commodity whose relative price has been lowered than in the steady state pattern of consumption (unpublished data). It remains to be seen what concepts in the literature, if any, organize these adjustment paths.

Finally, at times we have found behavior which economic theory simply does not address. For example, in our first studies of consumer demand behavior with food and water as commodities we found that sharp rotations in the budget line through an initial equilibrium consumption bundle could result in severe disruptions in subjects' behavior. This disruption was evidenced by thirsty and hungry rats consistently spending less than their income allowance and experiencing relatively large weight losses which would inhibit their ability to survive (Kagel *et al.*, 1975). There is nothing in existing theoretical formulations that would lead us to anticipate such a result.

We do not find these failures of behavior to conform to existing principles of economic theory particularly disturbing. Results of this type are bound to occur in the process of experimental investigations and serve to point out existing areas of ignorance concerning the behavioral process under

investigation. Given the large components of introspection and casual empiricism underlying current formulations of economic theories of individual choice behavior, the failure to identify existing areas of ignorance would be remarkable indeed. The use of laboratory experimental methods provides a ready-made basis for collecting the additional data needed to close these gaps in our understanding.

1.4. CRITERIA FOR EVALUATING EXPERIMENTAL RESULTS: INTERNAL VERSUS EXTERNAL VALIDITY

In evaluating experimental results, maintaining a distinction between *internal* and *external validity* is most useful (Campbell, 1957). *Internal validity* concerns the reliability of the results reported, without which any experiment is uninterpretable: Did the experimental treatments in fact make a difference? *External validity* concerns questions of *generalizability*: To what populations, settings, and variables can the effects reported be generalized?

The primary method for establishing internal validity in studies of individual subject behavior is the use of within-subject designs where each subject serves as its own control (Sidman, 1960). The most commonly employed design is of the ABA type, an example of which was reported in Fig. 4. In this design the experimental control of a variable is demonstrated through showing that the behavior can be altered and then returned to its initial level through altering experimental conditions. This approach completely bypasses variability due to intersubject differences and enables one to directly observe individual behavior responses. Following the return to baseline conditions experimental treatments will commonly be extended either through repeating the B condition exactly (direct replication) or introducing a similar but different experimental treatment (systematic replication), e.g., if the B phase involves an increase in the relative price of a given commodity, a systematic replication might involve a decrease in its relative price.

Within-subject designs of this type are ideally suited to testing theories of individual consumer demand and labor supply behavior which make directional predictions concerning responses to changes in equilibrium conditions. The key assumption underlying the ABA design, namely, that behavior is reversible when baseline conditions are reinstated, is generally assumed with respect to comparative static propositions in economic theory also. It is worth emphasizing that the use of ABA type designs does not presuppose that responses between baseline periods will be *exactly* the

same, rather, that there is a significant reversal of response patterns with the return to baseline conditions (see Fig. 4). In those cases for which the reversibility assumption does not hold, alternative within-subject designs are available (Sidman, 1960; Kazdin, 1973) or the researcher must use a between-subject design.¹¹

Intersubject replication is also important in establishing internal validity, with the number of replications needed, in part, a function of the extent to which results reported are consistent with earlier experimental findings. Under all circumstances one is likely to encounter smaller subject sample sizes than economists are used to working with. The smaller sample sizes are, in part, a result of the high degree of control of environmental conditions maintained in laboratory experiments which tends to promote uniformity of response patterns, relative to baseline performance levels, if the independent variable being manipulated is in fact controlling behavior (see Fig. 3).¹² Further, in those cases where behavior is not consistently responding to experimental treatments because of some hypothesized inability to control an element of the environment, a more direct solution is to bring this element under control or, if this is too costly, to increase the magnitude of the independent variable under investigation so that the differences between baseline and experimental conditions are now observationally distinguishable. On the other hand, if the hypothesis is that behavior is not under control because of prior behavioral histories or genetic differences, a more direct solution is to find populations to sample with more homogeneous behavioral and/or genetic backgrounds.

In reporting the outcomes of our experiments, questions frequently arise concerning the generalizability of the results reported because they are based on samples from such "unrepresentative" populations relative

¹¹ It has recently been suggested that greatly increased efficiency could be achieved in social policy experiments, such as the New Jersey income maintenance experiment, using within-subject designs (Hall, 1975), while the use of extended baseline (preexperimental) measures of behavior seems preferable to modeling individual response rates as a strategy for reducing between-subject variability in these studies (Cain *et al.*, 1974, p. 206; Hall, 1975). Psychologists commonly adopt both strategies in applied settings (Winett, 1976).

¹² In animal studies using own subject control designs, 2-4 intersubject replications using subjects drawn from the same strain of a given species have proved sufficient, *in practice*, to insure internal validity, *i.e.*, to insure direct replicability of experimental outcomes with a very high probability. This results from the homogeneous genetic and environmental history of the parent population from which subjects are drawn, along with the strong control of nonexperimental factors possible in animal studies. In the absence of cost constraints more direct replications within a given strain would undoubtedly be reported. However, under present cost conditions substantially greater payoffs are to be realized through direct replication using another strain or another species and/or through systematic replications with the same strain.

to our common concerns with behavior in national economies. This is a question regarding the *external validity* of the results reported to which there are a number of general answers. First, questions of external validity are not, of course, confined to experimental research but are endemic to nonexperimental work as well (see Campbell & Stanley (1966) for illustrations of the common nature of the sources of such threats in both experimental and nonexperimental research). Second, questions of external validity, like all questions of inductive inference, are never completely answerable even in theory. However, partial answers are obtained by further empirical research directed at verification of the initial observations within an ever widening set of conditions. This is the role of replication, both systematic and direct, especially by independent research teams, within an experimental science. The discovery and explanation of apparent failures to replicate previous experimental outcomes results in additions to and refinements in experimental methods and/or the concepts and relations in the uninterpreted theory under study. As such, these "failures" are something experimenters seek, rather than shy away from. Third, for economic reasons (questions of costs versus benefits) and for reasons of internal validity, the *sine qua non* of experimentation, behavioral scientists frequently seek out "unrepresentative" populations to work with. For example, economic considerations were in large part responsible for our study of consumer demand behavior in a token economy with long-term psychiatric patients (Battalio *et al.*, 1973, 1974) since the costs of setting up a token system with volunteer households from the community lay outside the budget we had to work with. Considerations of internal validity have also motivated biologists to use protozoa in consumer behavior studies in order to minimize the complex learning and training effects observed for vertebrates with their possible affects on preferences, a confounding factor to be controlled for in studying responses to price and income changes (Covich, 1972).

With respect to specific questions of generalizability associated with using populations from token economies and nonhuman subjects, there are two final points worth making. First, a technological basis exists for replicating token economy experiments across differing token systems with markedly different population characteristics as well as for replication outside the token economy. For example, we have conducted tests of comparable propositions within consumer demand theory using as subjects laboratory animals (Kagel *et al.*, 1975), long-term psychiatric patients (Battalio *et al.*, 1973), and stutterers in a token economy who were temporarily on leave from positions in the national economy (Cenar, 1976), while studies using budget reports represent field trials of these propositions (Koo, 1963; Mossin, 1972). Granted a technological basis for experimentally verifying the generality of results reported, an important factor underlying

the choice of technology for the initial collection of particular kinds of observations is the strictly economic one of costs versus benefits, on which criteria investigations using token economies with institutionalized populations and operant chambers with laboratory animals frequently score quite well. It seems inefficient to forgo such laboratories on the grounds that the results may not generalize to more "normal" populations under more "normal" conditions, particularly when the historical record indicates that the probabilities of such failures are *a priori* considerably less than one. Such laboratories can, at the least, serve as initial screening devices for testing a number of economic hypotheses, the failure of which stimulates the collection of comparable data for more "normal" subjects or the reformulation of the underlying theoretical propositions.

Second, experimental studies of economic behavior of animals and institutionalized populations in token systems expand considerably the scope for comparative analysis in economics and as such are of scientific value in their own right. Within this framework differences between these unusual subject populations and their more "normal" counterparts are as of much interest as are similarities. Understanding the causes and adaptive functions of these differences and similarities can add enormously to our understanding of economic behavior in general.¹³

¹³ Kourilsky & Hirshliefer (1976) have criticized token economies as corresponding too closely to traditional reactive models of economic man. They argue that self-determination and creativity are an essential element of economic decision making lost in more formal models of optimizing behavior and in token economies as well. They also cite studies of "overjustification" effects—the hypothesis that a person's intrinsic interest in an activity may be undermined by inducing him to engage in that activity as an explicit means to some extrinsic goal—as an important potential therapeutic defect in token economies, which they hypothesize to result from subjects' reactions to the manipulative nature of the system. This line of argument has suggested to some that token economies are inappropriate vehicles for the experimental analysis of economic behavior, an implication we strongly disagree with.

The criticism that token systems correspond too closely to reactive models of economic man, deficient in some essential respect to behavior in national economies, is a criticism of economic theory rather than the use of token systems to test theories, an issue which V. L. Smith addresses elsewhere in this volume. Further, the planned economic environments we discuss here include a broad range of organizational structures which in some cases permit a great deal of self-determination on subjects part, e.g., in a pilot study preceding the cannabis experiments reported in the text, subjects successfully struck for higher wages midway through the experiment (Kagel, Battalio, & Miles, in press)! In practice there is an inverse correlation between the manipulative nature of token systems, as Kourilsky and Hirshliefer use that term, and the degree of functional autonomy subjects are physically and mentally capable of performing. Finally, recent research (Mynatt *et al.*, 1978; Fisher, 1979) questions the robustness and generalizability of overjustification effects beyond the most limited of conditions and indicates that the phenomena has been identified only once in a functioning token economy, although this has not prevented authors from discussing the practical implications of their findings for these systems.

2. Studies of Aggregate Economic Behavior and More Policy Oriented Research Questions

2.1. ALCOHOL AND DRUG RESEARCH

Token systems and animal models have been used for some time now to address behavioral problems of immediate policy interest in the area of alcohol and drug research (Bigelow & Liebson, 1972; Meisch & Thompson, 1973; Nathan *et al.*, 1972). One of the most ambitious token economy studies of this kind was Miles and Congrieve's use of a "microeconomy" to study the socioeconomic effects of heavy, continuous marihuana smoking (Miles *et al.*, 1974). This controlled economic environment housed volunteer subjects continuously for 98 days in the facilities of the Addiction Research Foundation of Toronto, Canada.

To study the amotivational effects of marihuana on work performance, subjects could weave unlimited numbers of woolen belts on small portable hand looms, receiving \$2.50 (Canadian currency) for each acceptable belt completed. The induced value for this job task was far from trivial, resulting in subjects averaging 8 hours of work a day throughout the study and netting as much as \$5000 by the end of the experiment. The consumer goods provided (free of charge) consisted of assignment, in pairs, to semiprivate rooms with heat, lights, and cleaning services also provided. All other goods had to be purchased from production earnings. A store on the premises offered canteen items, including a full line of alcoholic beverages; recreation facilities were available for small rental fees including a gymnasium, sauna, and swimming; meals were available at the hospital; and staff members were regularly available to purchase retail items outside the hospital at market prices and to escort subjects on field trips. Three experiments with 19-20 subjects each were conducted in this economy, which is the closest approximation to date of Oskar Morgenstern's (1954) notion of establishing communities for the explicit purpose of conducting economic experiments. While relatively expensive and difficult to run, this economy is illustrative of the possibilities in this area (also see Bigelow, Emurian, & Brady (1975) for a similar technology of equal ambition).

Table 1 reports the results of *t* tests for the effects of marihuana on the primary measure of productivity, number of belts produced, in each of the three experiments. As can be seen, no significant differences in production between experimental and control groups were found in any of the experiments. Combined evaluation of the outcome of the experiments using Fisher's *Z* statistic (Winer, 1971) does nothing to alter the acceptance of

TABLE 1
CHANGES IN PRODUCTION RELATIVE TO PREDRUG PERIOD^a

Experiment	Treatment ^b	Number of days of treatment	Number of subjects ^c	Mean change in production (increase)	t statistic (probability $t \neq 0$) ^d
I	Experimental	49	10	91%	1.55
	Control	49	9	60%	(.14)
II	Experimental	53	9	64%	.81
	Control	53	10	36%	(.43)
III	Experimental	56	10	43%	-1.38
	Control	56	10	71%	(.19)

Data from Kagel, Battalio, & Miles (in press).

^a All experiments had a 17 days predrug period. Predrug production levels computed over the last 9 days of the predrug period. For the first week of the predrug period production was close to zero as subjects learned the job task.

^b Experimental subjects required to smoke two (1 gram) cigarettes containing 8 mg of Δ^9 -THC (the major active compound in hashish and marihuana) nightly. See Source for details of experimental design.

^c In both experiments I and II one subject left during the study. Data for these subjects have been excluded from the analysis.

^d Probability levels computed on the basis of a two-tailed alternative to the null hypothesis that the change in Experimental less the change in Control was equal to zero.

the null hypothesis of no difference between groups ($Z = .53$; probability $Z \neq 0 = .59$). Additional analysis of data on hours worked, output per hour, and activities during periods immediately following smoking were used to suggest a general behavioral relationship between marihuana and work performance (Kagel, Battalio, & Miles, in press), the details of which go beyond the scope of the present paper.

Obvious subject preselection biases, the special nature of the job task used in the experiments, the details of the structure of the microeconomy, and a number of other factors limit the ability to generalize the results of these experiments to working conditions in national economic systems. However, the data do falsify the contention that heavy use of marihuana is inherently incompatible with maintaining productivity levels and the capacity to carry out vocational responsibilities; propositions of some concern to both the United States and Canadian Commissions on marihuana and drug abuse. The experimental results also suggest a set of circumstances whereby regular marihuana use would be compatible with no alterations in productivity levels within national economic systems (Kagel *et al.*, in press). These are limited but valuable findings. They are of

interest in their own right and as inputs into designing and evaluating the potential cost effectiveness of longer term and more expensive epidemiological studies within national economies.

Nonhuman subjects have also been extensively used to study the behavioral effects of alcohol and other drugs. The rat data reported in Fig. 1 were generated in one such study. The question addressed there was whether or not alcohol intake would be responsive to changes in costs of consumption, as proved to be the case. The acknowledged fact in biology and psychology that behavioral laws do not stop suddenly at the boundary separating humans from nonhumans provides the foundation for such studies.

2.2. STUDIES OF INCOME DISTRIBUTION

Records of individual subject earnings, maintained in the course of studying consumer demand behavior in the Central Islip token economy (Battalio *et al.*, 1973, 1974) and the effects of marijuana use in the Cannabis economy, provide a basis for studying the distribution of income in these economies. Such distributions are of special interest since they were generated under conditions for which many of the alleged sources of unequal incomes in national economies, such as discrimination, monopoly restrictions, and costs of occupational training and experience, had been eliminated or were greatly reduced (Battalio, Kagel, & Reynolds, 1977). The primary factors responsible for income differences in these economies were differences in tastes for market income versus leisure and differences in abilities working manual job tasks.

Striking similarities were found in comparing measures of income dispersion in the experimental economies with those of the United States and other market economies. Table 2 reports comparisons based on the Gini coefficient. These results suggest that differences in abilities to perform simple manual job tasks and preferences for market income are *sufficient* to generate income dispersions similar to those observed in national economies, a point which has been seriously questioned in the literature (Fair, 1971; Blinder, 1974). Note that this is not the same thing as asserting that the distribution of income in the United States is created solely by individual differences in ability and effort. Here too there are obvious differences between the institutional structure and population characteristics of the experimental and national economies which prevent such generalizations. But these token economy results falsify the contention the incomes will *necessarily* be more evenly distributed in a world where all sources of income differences, except ability and tastes for leisure, are eliminated. In this respect the analysis serves the traditional function of experimentation in the bio-

TABLE 2

SELECTED GINI RATIOS FOR INCOME DISTRIBUTIONS IN NATIONAL AND EXPERIMENTAL ECONOMIES

Country or experiment	Income concept	Recipient unit	Gini ratio
1. United States	IRS adjusted gross income	Individual tax returns	.46
2. Central Islip ^a	Labor earnings (no personal tax)	Female workers	.41 (max)
3. United States	Pretax money income	Families and unattached indiv.	.40
4. United	Pretax money income	Female-headed families	.40
5. United States	Income after taxes and transfers	Families and unattached indiv.	.36
6. United States	Pretax money income	Families	.35
7. Central Islip ^a	Labor earnings (no personal tax)	Female workers	.34 (min)
8. Cannabis ^a	Pretax labor earnings	All workers	.31
9. United States	Consumption expenditures	Families and unattached indiv.	.31
10. India	Consumption expenditures	Individuals	.31
11. United States	Pretax labor earnings	Household heads, age 25-34	.29
12. Cannabis ^a	Pretax labor earnings	Male workers, age 20-28	.29
13. United States	Pretax wages and salaries	Male workers, full time, all indus.	.28
14. United States	Pretax wages and salaries	Male workers, full time manufact.	.26
15. United Kingdom	Pretax wages and salaries	Male workers, full time, employees	.21

Source: Battlio, R. C., Kagel, J. H., & Reynolds, M. Income distributions in two experimental economies. *Journal of Political Economy*, 1977, 85, 1259-1272. © 1977 by The University of Chicago. All rights reserved.

^a Experimental economies.

logical and behavioral sciences—to isolate basic relationships under controlled conditions.

The successful development of a labor supply model using rats and pigeons as subjects provides a natural basis for extending the token economy studies of income distributions. Maintaining similar rearing and home cage environments in such experiments would control for the effects of previous behavioral history on labor supply, a potential contaminating factor in the token economy studies. Alternatively one could systematically vary rearing and/or home cage environment to determine the effects of different life experiences on performance. Use of nonhuman subjects would also facilitate investigating the effects on the distribution of income of changes in wage rates and in nonwage income levels, to name but a few of the interesting possibilities. Such experiments would also permit determining the effects of such

changes on the ranks of individuals within the distribution, a virtual impossibility using currently available data from national economic systems. Experiments along these lines would involve using substantially larger sample sizes than those commonly employed in studies of individual subject behavior. While such experiments would be relatively expensive, the cost would be but a fraction of the cost of comparable studies using human subjects.

3. Conclusion

This paper has explicated the use of two new (for economics) technologies in the experimental analysis of economic behavior. Hopefully it is clear from the discussion that as experimentalists we are not suggesting that economic research begins and ends with experiments. Rather, the use of experimental methods can be an important element in the research process having important interdependencies with "pure" theory development and nonexperimental field observations as well. In fact the surest sign of a mature and healthy subdiscipline of experimentation within economics will be when practitioners in these other areas begin to incorporate outcomes of experiments as premises in their theories or as observations in need of explanation and replication using data from national economies.

ACKNOWLEDGMENTS

Research was partially supported by National Science Foundation Grants G532057 and SOC 77-27062. The editors and reviewer provided helpful comments on an earlier draft of this paper. Discussions with Ed Fisher, Louis Philips, Dick Winett, and Robin Winkler helped clarify several points in the paper. We alone are responsible for any remaining errors.

REFERENCES

- Allison, J. Demand economics and experimental psychology. *Behavioral Science*, 1979, **24**, 403-415.
- Allison, J., Miller, M., & Wozny, M. Conservation in behavior. *Journal of Experimental Psychology: General*, 1979, **108**, 4-34.
- Basmann, R. L. Modern logic and the suppositious weakness of the empirical foundations of economic science. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, **1975**, **111**, 153-176.
- Basmann, R. L., Battalio, R. C., & Kagel, J. H. Comment on R. P. Bryon's "The restricted Aitken estimation of sets of demand relations." *Econometrica*, 1973, **41**, 365-70.
- Battalio, R. C., Kagel, J. H., & Green, L. Labor supply behavior for animal workers: Towards an experimental analysis. In V. L. Smith, (Ed.), *Research in experimental economics*. Vol. 1. Greenwich, Conn.: JAI Press, 1979. Pp. 231-253.

- Battalio, R. C., Kagel, J. H., & Reynolds, M. Income distributions in two experimental economies. *Journal of Political Economy*, 1977, **85**, 1259-1272.
- Battalio, R. C., Kagel, J. H., Winkler, R. C., Fisher, E. G., Jr., Basmann, R. L., & Krasner, L. A test of consumer demand theory using observations of individual consumer purchases. *Western Economic Journal*, 1973, **11**, 411-428.
- Battalio, R. C., Kagel, J. H., Winkler, R. C., Fisher, E. G., Jr., Basmann, R. L., & Krasner, L. An experimental investigation of consumer behavior in a controlled environment. *Journal of Consumer Research*, 1974, **1**, 52-60.
- Bigelow, G. E., Emurian, H. H., & Brady, J. V. A programmed environment for the experimental analysis of individual and small group behavior. In C. G. Miles (Ed.), *Experimentation in controlled environment*. Toronto: Addiction Research Foundation, 1975. Pp. 133-144.
- Bigelow, G. E., & Liebson, I. Cost factors controlling alcoholic drinking. *The Psychological Record*, 1972, **22**, 305-314.
- Blinder, A. S. *Toward an economic theory of income distribution*. Cambridge Mass.: MIT Press, 1974.
- Brunner, K. A. A case study on the importance of appropriate rules for the competitive market in ideas and beliefs. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 1967, **103**, 173-189.
- Cain, G. C., Nicholson, W., Maller, C., & Wooldridge, J. The labor-supply response of married women, husband present. *The Journal of Human Resources*, 1974, **9**, 201-222.
- Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 1957, **54**, 297-312.
- Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago Ill.: Rand McNally, 1966.
- Cenar, M. Tests of the strong axiom of revealed preference theory and variations in consumer equilibrium demand in a token economy. Unpublished Ph. D. dissertation, Texas A & M University, 1976.
- Covich, A. Ecological economics of seed consumption by *Peromyscus*. *Transactions of the Connecticut Academy of Arts and Sciences*, 1972, **44**, 71-93.
- Fair, R. C. The optimal distribution of income. *Quarterly Journal of Economics*, 1971, **85**, 551-579.
- Fisher, E. B., Jr. Overjustification effects in token economies. *Journal of Applied Behavior Analysis*, 1979, **12**, 407-415.
- Georgescu-Roegen, N. *Analytical economics: Issues and problems*. Cambridge Mass.: Harvard Univ. Press, 1966.
- Gorman, W. M. Community preference fields. *Econometrica*, 1953, **21**, 63-80.
- Green, H.A.J. *Aggregation in economic analysis*. Princeton N.J.: Princeton Univ. Press, 1964.
- Hall, R. E. Effects of the experimental negative income tax on labor supply. In J. A. Peckman and P. M. Timpane (Eds.), *Work incentives and income guarantees*. Washington D.C.: Brookings Institute, 1975. Pp. 115-156.
- Heyman, G. M., & Luce, R. D. Operant matching is not a logical consequence of maximizing reinforcement rates. *Animal Learning and Behavior*, 1979, **7**, 133-140.
- Hicks, J. R. *A revision of demand theory*. London and New York: Oxford Univ. Press (Clarendon), 1956.
- Kagel, J. H., Battalio, R. C., & Miles, C. G. Marihuana and work performance: Results from an experiment. *Journal of Human Resources*, in press.
- Kagel, J. H., Battalio, R. C., Rachlin, H., Green, L., Basmann, R. L., & Klemm, W. R. Experimental studies of consumer demand behavior using laboratory animals. *Economic Inquiry*, 1975, **13**, 22-38.
- Kazdin, A. E. Methodological and assessment considerations in evaluating reinforcement programs in applied settings. *Journal of Applied Behavior Analysis*, 1973, **6**, 517-531.

- Koo, A. Y. C. An empirical test of revealed preference theory. *Econometrica*, 1963, 31, 646-664.
- Kourilsky, M., & Hirshleifer, J. Mini-society versus token economy: An experimental comparison of the effects on learning and autonomy of socially emergent and imposed behavior modification. *Journal of Educational Research*, 1976, 69, 376-381.
- Marshall, A. *Principles of economics*. (8th ed.) London: Macmillan & Co, 1920.
- May, K. O. Intransitivity, utility, and aggregation of preference patterns. *Econometrica*, 1954, 22, 1-13.
- McCrimmon, K. R., & Toda, M. The experimental determination of indifference curves. *Review of Economic Studies*, 1969, 36, 433-451.
- Meisch, R., & Thompson, T. Ethanol as a reinforcer: Effects of fixed-ratio size and food deprivation. *Psychopharmacologia* (Berl.), 1973, 28, 171-183.
- Miles, C. G., Congreve, C. R. S., Gibbons, R. J., Marshman, J., Devenyi, P., & Hicks, R. C. An experimental study of the effects of daily cannabis smoking on behavior patterns. *Acta Pharmacologica et Toxicologica*, Supplement I, 1974, 34, 1-44.
- Morgenstern, O. Experiment and large scale computation in economics. In O. Morgenstern (Ed.), *Economic activity analysis*. New York: Wiley, 1954. Pp. 484-548.
- Mossin, A. A mean demand function and individual demand functions confronted with the weak and the strong axioms of revealed preference: An empirical test. *Econometrica*, 1972, 40, 177-192.
- Muellbauer, J. Aggregation, income distribution and consumer demand. *Review of Economic Studies*, 1975, 62, 525-543.
- Mynatt, C., Oakley, T., Arkkelin, D., Piccione, A., Margolis, R., & Arkkelin, J. An examination of overjustification under conditions of extended observation and multiple reinforcement: Overjustification or boredom? *Cognitive Therapy and Research*, 1978, 2, 171-177.
- Nathan, P. E., Goldman, M. S., Lisman, S. A., & Taylor, A. A. Alcohol and alcoholics: A behavioral approach. *Transactions New York Academy of Science*, 1972, 34, 602-627.
- Philips, L. *Applied consumption analysis*. New York: American Elsevier, 1974.
- Philips, L. Comment. *Annals of Economic and Social Measurement*, 1975, 4, 45-48.
- Rachlin, H. A molar theory of reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, 1978, 30, 345-360.
- Rachlin, H., & Burkhard, B. The temporal triangle: Response substitution in instrumental conditioning. *Psychological Review*, 1978, 85, 22-47.
- Rivlin, A. M., & Timpone, P. M. (Eds.) *Ethical and legal issues of social experimentation*. Washington D. C.: Brookings Institute, 1975.
- Robbins, L. On the elasticity of demand for income in terms of effort. *Economica*, 1930, 29, 123-129.
- Rousseas, S. W., & Hart, A. G. Experimental verification of a composite indifference map. *Journal of Political Economy*, 1951, 59, 288-318.
- Samuelson, P. A. *Foundations of economic analysis*. Cambridge, Mass.: Harvard Univ. Press, 1947.
- Sidman, M. *Tactics of scientific research*. New York: Basic Books, 1960.
- Siegal, S. Decision making and learning under varying conditions of reinforcement. *Annals New York Academy of Sciences*, 1961, 89, 766-783.
- Smith, V. L. Experimental economics: Induced value theory. *American Economic Review*, 1976, 66, 274-279.
- Staddon, J.E.R. Operant behavior as adaptation to constraint. *Journal of Experimental Psychology: General*, 1979, 108, 48-67.
- Staddon, J. E. R., & Motheral, A. On matching and maximizing in operant choice experiments. *Psychological Review*, 1978, 85, 436-444.
- Stigler, G. J., & Becker, G. S. De gustibus non est disputandum. *American Economic Review*, 1977, 67, 76-90.

- Thurstone, L. L. The indifference function. *Journal of Social Psychology*, 1931, 2, 139-167.
- Wilson, E. O. *Sociobiology*. Cambridge, Mass.: Harvard Univ. Press (Belknap), 1975.
- Winer, B. J. *Statistical principles in experimental design*. (2nd ed.) New York: McGraw-Hill, 1971.
- Winett, R. A. Small scale field studies: A microeconomic approach to energy conservation. Paper presented at Southern Economic Association Meetings, Atlanta, Georgia, November 1976.
- Winkler, R. C. The relevance of economic theory and technology to token reinforcement systems. *Behavior, Research and Therapy*, 1971, 9, 81-88.