



Australian School of Business Working Paper

Never Stand Still

Australian School of Business

Australian School of Business Research Paper No. 2013 ECON 34

Episodes from the Early History of Experimentation in Economics

Andreas Ortmann

This paper can be downloaded without charge from
The Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=2368596>

Episodes from the Early History of Experimentation in Economics

by

Andreas Ortmann, School of Economics, Australian School of Business,
University of New South Wales, Sydney, NSW 2052, Australia¹

December 12, 2013

Introduction

The present article builds on a background paper that was commissioned for a “Witness Seminar” in 2010 that had a dozen prominent experimental economists – witnesses, indeed -- discuss the origin and evolution of experimental economics.² Rather than providing a history of the experimental method in the behavioral science (with particular emphasis on those practices that informed experimental practices in economics; see Ortmann 2009), I was asked to provide exhibits from the early years of experimental economics. . I was asked to refrain from interpretation and evaluation: “any suggestion of a linear history (as for example when, how and why the experimental methods in economics departed from those in psychology) should be avoided. ... it is absolutely crucial to the success of the Witness Seminar to have the paper written open-ended, highlighting the questions at the time about specific episodes in experimenting.” (Letter Harro Maas and Andrej Svorencik 12/12/2009)

The five episodes that I present below are meant to be “museum pieces”; their purpose was to trigger memories and initiate discussions. For each museum piece I sketched its context, then I summarized it (making occasionally, and quite intentionally, heavy use of quotations), then I highlighted the methodological questions that the particular episode illustrated.

Episode One: The Wallis - Friedman (1942) critique of the Thurstone (1931) experiment

Context

The Wallis – Friedman critique of Thurstone’s (1931) study about the experimental constructability of indifference maps has received prominent play in several places (e.g., MacCrimmon & Toda 1969; Castro & Weingarten 1970; Kagel 1972; Battalio et al. 1974; Kagel & Battalio 1980, see fn 8; Roth 1995; Moscati 2007; Lenfant 2009³).

¹ I dedicate this manuscript to Raymond C. Battalio who surely would have been among the eminent invitees to the Witness Seminar had he not passed away, at age 66, Dec 1 2004 (<http://econweb.tamu.edu/battalio.asp>). He was an audacious pioneer of experimental economics. He was also a wonderful teacher and mentor. I am grateful for all he taught me.

² The version of this article presented at the Witness Seminar is available upon request from a.ortmann@unsw.edu.au. A version of the present manuscript will be published in a focus section of *The Journal of the History of Economic Thought* that will be devoted to a Witness Seminar on the History of Experimental Economics that was held in May 2010 at the KNAW in Amsterdam.

³ Interestingly, Lenfant documents that three reviews of Thurstone’s article that preceded the Wallis-Friedman chapter gained little traction.

Summary

Thurstone was a professor of psychology at the University of Chicago and his experimental study was motivated through “numerous conversations about psychophysics with my friend Professor Henry Schultz” (Thurstone 1931, p. 139). Schultz’s major interest was the measurement of utility and demand functions (Schultz 1933, 1938). Thurstone (1931, p. 139) – acknowledging his own limited knowledge of economics -- credits Schultz with the problem formulation and the suggestion to apply the experimental method to this problem in economic theory.

Lenfant (2009), based on considerable sleuthing, suggests how the Thurstone study fit into Schultz’s overall research agenda and how Wallis and Friedman got into the game. The bottom-line is: they met as graduate students of Schultz in 1934 and eventually overlapped for years at a time at The National Resources Committee and the Statistical Research Group (Lenfant 2009, p. 19). Lenfant conjectures that Friedman’s interest in the Thurstone study may have been the result of his significant contribution to chapters 18 and 19 of Schultz (1938) and accompanying discussions that Schultz and Friedman must have had in this context. Be that as it may, Wallis & Friedman (1942) ended up as a joint contribution to a volume in memory of Henry Schultz. The alphabetical inversion of names on the chapter is unusual and probably reflects the relative contributions; it is in this context noteworthy that Wallis graduated in psychology and economics from Chicago and Columbia (Lenfant 2009, p. 19).

Thurstone (1931) tried to trace out through questionnaires the indifference maps for hats, shoes, and coats of a group of girls.⁴ Specifically, he offered his subjects, hypothetically, various bundles of commodities (e.g., hats vs pairs of shoes and hats vs overcoats) and then constructed from their responses indifference curves. He even estimated parameters that he used for out-of-sample prediction (e.g. the subjective trade-off between shoes and overcoats).

Although Thurstone, the psychologist, cared about subtleties such as experimenter expectancy effects (e.g., Rosenthal & Rubin 1978, Rosenthal 1992; Ortmann 2005; Zizzo 2010), Wallis & Friedman (1942) critiqued his experiment on several grounds, foreshadowing what nowadays is often referred to as the artificiality critique (e.g., Schram 2006). To wit:

“For a satisfactory experiment it is essential that the subject gives actual reactions to actual stimuli . . . Questionnaires or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react. The reactions of people to variations in economic stimuli work themselves out through a process of successive approximation over a period of time. The initial response indicates only the first step in a trial-and error-adjustment.” (Wallis & Friedman 1942, pp. 179-80)

“If a realistic experimental situation were devised, it would, consequently be necessary to wait a considerable time after the initial application of the stimulus before recording the reaction. Even an experiment of restricted scope would have to continue for so long a period that it would be exceedingly difficult to keep ‘other things the same’.” (Wallis & Friedman 1942, p. 180)

⁴ In his article, Thurstone presented only the data of one of his subjects (Thurstone 1931, p. 154, p. 166) stressing that “the same procedures have been tried on several other subjects with similar results” (p. 166)

Wallis and Friedman proposed an experiment in which children, day after day, would be offered various combinations of candy and ice cream, of which they each would choose one to consume. Wallis & Friedman had several objection to their own proposal (e.g., related to the stability of preferences across time, and preferences for variety) and concluded, that “it is probably not possible to design a satisfactory experiment for deriving indifference curves from economic stimuli” (Wallis & Friedman 1942, p. 181) be it for the simple reason that it would be difficult to keep other things the same.⁵

Interestingly, the related experimental work on transitivity by mathematician/mathematical economist May (1953, 1954), economist Papandreou (1953, Papandreou et al. 1954), and sociologist Rose (1957), extensively and superbly discussed in Moscati (2007, pp. 376 – 384), were also based on hypothetical choices and subject to the Wallis-Friedman critique. One can speculate whether these papers did not place in good journals for that reason (with the notable exception of Rose’s) but Moscati argues that they were nonetheless (ironically, again with the notable exception of Rose’s) influential in triggering a debate of the transitivity axiom. Moscati’s case seems persuasive. (One could, of course, ask whether these papers might have been even more influential had they not been subject to the Wallis-Friedman critique.)

The methodological questions this museum piece highlights:

This museum piece addresses the hypothetical nature of the stimuli and subjects’ unwillingness to state, or – because of the artificiality of the experimental situation -- their inability to know their true reactions, the unrepresentativeness of subjects’ response, and the fact that preferences in economic situations are unlikely to be stable.

Episode Two: Morgenstern (1954) on experiment and large scale computation in economics

Context

“The possibilities of controlled direct experiments in the economy as a whole are very numerous – contrary to a widespread belief of the opposite. Indeed, they are only limited by the amounts of money one wishes to devote to them and by restrictions of ethics, common decency, political prejudices and the like – all of them very sound restrictions. However, even within these restrictions a larger monetary effort could provide significant quantities of new information not available so far.”
(Morgenstern 1954, p. 515)

⁵ Later authors (namely, Rousseas & Hart 1951; MacCrimmon & Toda 1969; see Moscati 2007, pp. 373 – 376 regarding the former and pp. 387 – 391 regarding the latter) begged to differ. Moscati (2007) documents that the Rousseas & Hart article, save a couple of citations, had little traction. The MacCrimmon & Toda article did better, as it managed to attract in the following two decades about a dozen citations in the journals of economics, business, and finance listed in JSTOR. The extended discussion of methodological issues that might mar an experiment’s validity (e.g., MacCrimmon & Toda 1969, pp. 434 – 436 and pp. 447 – 450), is a very explicit response to various criticisms leveled against precursor experiments in this area. Indeed, these pages demonstrate convincingly the seriousness with which methodological issues were being discussed at the tail end of the sixties. This statement is also true for Jim Friedman (1969) which was published in *The Review of Economic Studies*, as part of a *symposium on experimental economics*.

Innocenti & Zappia (2005, pp. 84 – 85) point out that in his earlier book on the reliability of economic data Morgenstern (1950) listed among the many sources of errors in economic statistics the lack of verification through experiments. In fact, it was the first of the sources that he listed. Morgenstern (1954) is a reversal of sorts because there he makes the persuasive case – to which later authors referred (e.g., Kagel & Winkler 1972; Kagel & Battalio 1980) – that experiments were a natural for economics. His revised assessment was quite likely a consequence of Morgenstern attending the University of Michigan Summer 1952 Seminar at Santa Monica which brought together a diverse group of researchers. This gathering strikes me as the most important event for experimental economics during the 1950s, an assessment shared by others (e.g., Heukelom 2010). In Thrall et al (1954, p. 331), Morgenstern is listed as having contributed a paper titled “Experiment and Computation in Economics” which seems to have been a precursor of Morgenstern (1954) which was written in 1953 (Morgenstern 1954, p. 496).

It is noteworthy, particularly in light of Naylor’s revisitation⁶ of Castro & Weingarten (1970), to note that to Morgenstern computations could be substitutes for experiments. Foreshadowing the purposes of experiments later identified by Witness Seminar participants (e.g., Roth 1995, he states:

“We distinguish two types of experiments: (1) Experiments of the first kind are those where new properties of a system are to be discovered by its manipulation on the basis of a theory of the system; (2) Experiments of the second kind do not primarily rely on a theory but aim at the discovery of new, individual facts. The distinction is not sharp, since the results of the experiments of the second type are eventually incorporated into a theory whereby they receive their standing. We can now state a general thesis: Every computation is equivalent to an experiment of the first kind and vice versa. The equivalence rests on the fact that each experiment (certainly each of the first kind) can be conceived of as being – or using – an analogue computing machine.” (pp. 499-500)

This thesis is then expanded on a dozen pages. In the following summary I shall focus on the section in Morgenstern (1954) which follows those pages and which deals with experiments in today’s meaning of the word.

Summary

⁶ Castro & Weingarten (1970) set out to explain “the manner in which experimental techniques may be systematically applied to the analysis of basic economic behavior.” (p. 598) Both Kagel (1972) and Naylor (1972) reacted explicitly to the Castro & Weingarten article, the former pointing out the “advantages and potentialities of designed experimentation in token economies” (p. 779) by way of several examples, while the latter argued “[the] paper by Castro and Weingarten is both misleading and incomplete. First, the literature in economics abounds with articles on experimental economics. Second, the authors’ survey of the literature in experimental economics omits some of the most important contributions in this field. Third, the authors completely ignore an important new field of experimental economics – namely, computer simulation experiments with models of economic systems. My paper contains an extensive survey of the literature on real-world experiments in economics as well as computer simulation experiments.” (p. 347) Among the articles Naylor discussed were a complete list of those motivated by Thurstone (1931) including MacCrimmon & Toda (1969), multiple references to Siegel’s work, more than a dozen articles on market experiments (missing, though, the work of the “Frankfurt school”; see Sadrieh & Weimann 2008), half a dozen articles dealing with large scale social experiments, and a long section on computer simulation experiments.

Morgenstern (1954) acknowledges “first the occasional appearance of strictly planned experiments and second the ability to compute on a large scale (with the aid of electronic computers) by making use of currently available theory. ... During the current decade still further possibilities will undoubtedly be explored of which those connected with experiment and computation appear to be especially promising. Their particular appeal lies – at least to my mind – in the combination of a profound study of the data and their new processing, with a rigor of the theoretical reasoning that can compare favorably with that of the natural sciences.” (pp. 484-5) Clearly the natural sciences (explicitly physics and astronomy) were Morgenstern’s template. He called them “the advanced empirical sciences” (p. 485) and left little doubt that economics, to his mind, was an empirical backwater.

Acknowledging that making experiment and large-scale computation standard tools in the profession’s tool-box would not be easy because economists had to acquire new skills and to become acquainted with new ideas and techniques, Morgenstern left no doubt that the conventional wisdom regarding the impossibility of experimentation in economics was wrong: “I do believe that there exists great opportunity for direct experiments now and in the future. I am thinking of the actual, physical, experiment, i.e., one in which physical reality is being subjected to desired conditions, as distinguished from the so-called ‘thought experiment’.” (p. 486); he labels thought experiments “indirect” experiments and the actual physical experiments “direct” experiments. His concern is mainly with the latter and our ability “in the real, physical, world under controlled conditions [to change] those variables that economists deem significant in their science and upon which one may be able to operate.” (p. 487)

After a lengthy section in which Morgenstern summarizes the development towards computation and computability in economics (touching on issues such as the solvability and stability of systems of many equations, the lack of experimental and other empirical determinations of initial conditions and parameters, knowledge requirements, and computation techniques), he turns his attention to “the direct experiment and measurement” (pp. 506 – 511) and “experimental possibilities in economics” (pp. 511 – 520: direct experiments, pp. 520 – 538), stressing that “the range of direct experiments thus far performed in economics or feasible in the future is very considerable. The frequently encountered opinion that direct experiments (...) for all practical purposes are impossible, cannot be maintained. On the contrary, the possibilities are numerous and depend to a large extent merely on the (monetary) means to be utilized.” (p. 512)

Arguing that “a history of the chief economic experiments performed would be of high value and very instructive” (p. 512), he then summarizes those he knows. He talks about von Thuenen’s agricultural experiments⁷, arguing that “Thuenen’s combination of experimental and theoretical effort has never been matched or surpassed in economics” (p. 513), the routine marketing experiments done by individual business organizations (p. 513)⁸, arguing that they represent a lost

⁷ He “carried out extensive and costly experiments on his large estate, kept precise records, designed new methods for experimentation, all this lasting many years.” (p. 512)

⁸ “Good examples are offered by a company planning the sale of new product and testing in similar, but separate towns, different advertising campaigns using various packages, prices, timing, etc. ... These experiments, so far as I can see, are totally neglected in economics where they are, or rather will be, of the greatest significance, even in their present form. ... Steps should be taken for a close cooperation with business firms; the benefits would be mutual and of the highest significance.” (p. 513)

opportunity for significant advances, and he recommends “a fine experiment” (Brunk & Federer 1953), and some literature quoted in that article as templates for good experimentation.

He then argues that “the possibilities for experimentation in business are practically inexhaustible: management can experiment with wage rates, hours, pay systems, etc. in normal surroundings far in excess of the ideas management has now, and economics could profit from the results immediately provided the cooperation is established which is indispensable for the progress of both.” (p. 514)

In a similar vein Morgenstern then stresses the ample opportunities for experimentation (“policy measures”) that government has. It does not use them in satisfying ways, though; “as experiments they are unclear, vaguely conceived and inadequately described so that the possibilities of an exploitation of the experiences for scientific purposes are very limited.” (p. 515)

Morgenstern finally dives into his account of designed experiments involving large aggregates, mentioning the sizable Jesuit settlements in Paraguay in the 17th and 18th century, Fourier’s “Phalanges” in France, Owen’s organizations in Scotland, various religious communities in the United States that employed special economic systems, the dated money experiment of the Woergl community in Austria, and the social credit movement in Canada. He sees these attempts as valid templates of designed experiments, essentially real-life laboratories that could speak to various issues (price systems, monetary arrangements, etc.) He even addresses the issue of the remuneration of the participants in these experiments: “Their remuneration introduces complications as well as simplifications. Both have been encountered previously and can be duly considered (cf. the Mosteller-Nogee experiment where essentially the same situation arose.)” (p. 517)

Morgenstern then goes on to discuss the possibilities of token economies of various kinds (more on this below). He singles out Radford’s “most interesting and brilliant account of price fluctuations (mostly in terms of cigarettes) in a British P.O.W. Camp” (p. 517) as a prototype of experimental studies that “may reveal decisive properties of demand, money, preferences, etc. about which we suspect nothing to at present.” (p. 517) He even discusses the problems of experimenter effects (p. 518) arguing that, while important to deal with, they are not an obstacle to the possibility of experimentation.

Moving on to designed experiments not involving large aggregates, and stressing that his knowledge is limited to publications, Morgenstern discusses skeptically Chamberlin (1948) “who used students in a course on economics to construct a market, mostly for pedagogical purposes, although some new results are also claimed.” (p. 519) He next discusses, much more positively, the Mosteller-Nogee (1951) experiment. He sketches its purpose – also relative to Von Neumann & Morgenstern (1944) -- and judges its merits as follows: “This may be the first direct experiment in economics that can compare with those in the physical sciences, including psychology. It is a true experiment and goes in every respect far beyond the questionnaire” (p. 519) The reference to the Mosteller-Nogee experiment, and his assessment, are interesting because they strongly suggests that Morgenstern must have been fully aware of the Wallis-Friedman critique of Thurstone’s experiment.⁹ In fact, throughout his article one gets the impression that Morgenstern has bought

⁹ Mosteller & Nogee (1951) do not cite Wallis & Friedman (1942) but they mentioned prominently and repeatedly (footnotes 1 and 7) that their experimental design benefitted from discussions with Friedman and Savage, as well as Wallis. Throughout their article one finds methodological reflections that are clearly responses to the objections that Wallis & Friedman (1942) enumerates.

into that critique hook, line, and sinker and that he takes the methodological implications of that critique for a given.

Morgenstern also discusses Ward Edwards' related experiment on economic decision making in gambling situations, pointing out that it had been presented at the 1952 Meetings of the Econometric Society. He stresses the relevance of this psychologist's work for economics both because of its relevant subject matter and methodological "neatness" (p. 520). He concludes: "From these experiments – and the many that will undoubtedly follow – will result a theory of utility that of a truly scientific character, removed from the realm of pure speculation." (p. 520)

He concludes his discussion of direct experiments with a reference to "quite extensive experiments" on games of strategy, referencing explicitly the Kalisch et al 1952 Rand Corporation working paper that was published as chapter 19 in Thrall et al. (1954) the same year (Kalisch et al. 1954). Interestingly, he argues "[these experiments] aim at gaining information about tendencies to form coalitions, their stability, preferences for certain types of strategies, etc. As long as these particular games are not specifically identified with typical economic situations we shall not enter upon further discussing, although they are potentially very important." (p. 521)

The methodological questions this museum piece highlights:

Morgenstern's chapter is remarkable in that he bluntly, confidently, and – in light of the considerable evidence that he lays out – convincingly contradicts what he identifies as the prevailing opinion at that time: that controlled direct experiments are not implementable. In contrast, Morgenstern sees them as a necessary step to move away from the shaky empirical foundations that he sees the house of economics built on (e.g., Morgenstern 1954, p. 489). He also seems to have accepted the Wallis – Friedman critique and seems to have methodological conceptions (financial incentives, ethical restrictions – probably implying no deception --, external validity (if necessary through work with token economies; see p. 517) that seem strikingly in line with today's conceptions – at least experimental economists' conceptions (e.g., Hertwig & Ortmann 2001) -- of what constitutes appropriate forms of experimentation.

Episode Three: Thomas Juster (1970) on the possibilities of experimentation and the quality of data-input in the social sciences

Context

Juster's article mirrors Morgenstern's concerns about the quality of data accessible to, and produced by, social scientists, how and why the production of economic knowledge differs between natural sciences and social sciences and what, if anything, can be done about it; it seems not coincidental that Juster refers to the second edition of Morgenstern (1950) on the first page of this article (fn 3; see also fn 11). They were clearly in agreement that "economists possess a very large and often quite useful stock of qualitative knowledge but a remarkably skimpy stock of quantitative knowledge." (Juster 1970, 139) The specific concern of both was the reliance on analysis of existing data, at sharply diminishing returns, and the need for new "experimental sets of microdata" (Juster 1970, p. 138). Morgenstern and Juster had different interpretations of the word "experimental", though. As we will see, for Juster "experimental" had to do with the framing and treatment definitions of survey questions. Incentive-compatible elicitation does not come up as an issue.

Juster addressed his concern in a series of articles before (e.g., Juster 1960, Juster 1961; 1964, 1966; see also Brady 1965 and Namias 1965) and after 1970 (e.g., Juster 1974; see also Juster & Stafford 1991 for an interesting turn to a new but clearly related area).

Some of Juster's work is related to that of George Katona of the Survey Research Center (SRC) at the University of Michigan, the same place where mathematical psychologists Coombs and Edwards resided although the interaction of Katona with Coombs and Edwards seems to have been minimal, or non-existent (Heukelom 2010). The SRC was set up after the Division of Program Surveys of the Department of Agriculture was dissolved and its senior officials (including George Katona) reconstituted themselves as the SRC. The SRC transformed a survey on liquid asset into annual surveys on consumer finances that were administered by the SRC and sponsored by the Federal Reserve Board starting in 1947.

It was the Federal Reserve Board that commissioned the Smithies Report (1955), a detailed review of the work of the SRC. The report addressed many of the issues that Juster subsequently addressed "experimentally" and, for a public document, it was fairly critical of the operation that Katona and his colleagues ran, as was Juster (e.g., Juster 1961).

Summary

Juster (1964) is a study of the interaction of anticipated and actual purchases of 13 consumer durable products that ranged from new automobiles to garbage disposal units. Remarkably, this study draws on interviews, and re-interviews, of about 20,000 households from the Consumers Union membership. To understand the ambition of this project it is useful to know that the SRC relied initially on 3,500 samples and later on 3,000.

Juster (1966) was a postscript of sorts to Juster (1964).

Juster's "experiments" in survey design were motivated by the insight that consumer purchase intentions were insufficient predictors of purchase rates. The key problem was that purchase intentions are binary and that, while they work reasonably well for intenders (duly controlling for response biases, income, assets, time frame, kind of product, etc.), they tend to fail for nonintenders which for the products under consideration were naturally the vast majority. Juster's results for intenders suggested that buying intentions were reflecting respondents' subjective probability of purchasing a product and that a stated intention was a function of that subjective probability. Intentions surveys could not detect movements in mean probability among nonintenders, and, given their (nonintenders') weight in the overall sample, intentions surveys were therefore bound to run into trouble. Juster suggested that probability statements (which in his view were in any case underlying expressed purchase intentions) "might well be obtainable empirically" (Juster 1966, p. 658) and correspondingly proposed a reframing of the questions asked in consumer surveys.

Contemporary reviewers of Juster (1964) such as Namias (1965) and Brady (1965) acknowledged his contributions to the art of analysis of survey data, research methodology, and the improvement of relevant survey techniques: "Among his contributions is the analysis of the significance of changes in wording, an essential requirement for the proper interpretation of survey data, and for increased accuracy of prediction." (Namias 1965, p. 109) And (Brady 1965, p. 203) adds, "The study is more generally a demonstration of a methodology for the study of formulation of survey questions. Through the discovery of a hypothesis about the nature of the responses to different questions on the same subject, the directions of further work can be guided efficiently with an expectation of progressive improvement in technique."

The quite dramatic improvements in the quality of predictions resulting from well-designed surveys demonstrated in Juster (1964, 1966) clearly informed Juster (1970; see also 1984 for a similar, even starker example). The latter was an attempt at policy intervention squarely aimed at those in power. As mentioned, the issue was the remarkably skimpy stock of quantitative knowledge and the need for new “experimental sets of microdata” (Juster 1970, p. 138). It should be clear though that Juster was not talking about the kind of direct experiments that Morgenstern had talked about, notwithstanding the curious excursion on pp. 142 – 144 which seems heavily influenced by Morgenstern’s ruminations on direct experiments. Compare, for example, Juster’s discussion of “research designs widely used in analysis of managerial decision making and marketing strategy.” (Juster 1970, p. 142)

Juster also addresses the issue how economists got stuck in the rut of a remarkably skimpy stock of quantitative knowledge¹⁰, and how they might manage to get out. He talks about how research economists spend their time, the costs of basic and processed data, the high costs of new data and data input in the social sciences: “For the most part, the data inputs into economic research consist of processed rather than basic data, and economics is probably unique among the sciences in the proportion of professional resources that go into the processing and manipulation of basic data.” (p. 140) He then compares this situation with the situation of the physical sciences: “Empirical research in the physical sciences is based almost entirely on observations generated as an essential part of the research process itself, and a large proportion (probably more than half) of professional skills is devoted to the questions of what observable phenomena are to be measured and how can the measurement be made.” (p. 141)

The methodological questions this museum piece highlights:

The specific concern of both Morgenstern and Juster was the problematic reliance on analysis of existing data, and the need for new experimental sets of microdata. The one obvious methodological insight of relevance to experimental economics one can take home from Juster (1964, 1996) is the impact that wording can have.

Episode Four: Token Economy and Animal Models for the Experimental Analysis of Economic Behavior (Kagel & Battalio 1980)

¹⁰ “The production of new microdata sets is extremely costly. ... The enormous cost of generating new microdata sets thus means that economists tend to use the best existing sets of data even though they may be seriously deficient for research needs, a tendency that can be offset only if the high costs are balanced by equally high returns.” (p. 140)

“Assume that you had a dissertation student writing on the question: ‘Other things equal, are retail prices higher in ghetto areas than suburban areas?’ Further assume that data already available could be used to examine the problem, but that the data were, as usual, seriously deficient ... Further assume that an extra two years would be required to collect and analyze the relevant data, and that additional costs would be incurred to raise the required funds. Given these facts, do you recommend that your dissertation student obtain the relevant observations or use existing data? I suggest that your student’s professional reputation and expected income would be unaffected by the choice of research strategy, and hence that you could not in fairness recommend the investment of an extra couple of years in generating observations. Yet the social value of the two research designs is markedly different: one has the potential for providing a solid basis for policy decisions; the other does not.” (p. 145)

Context

As also evidenced in the frequent references to the work of Skinner, Ayllon & Azrin, Kazdin & Bootzin in their earlier work during the seventies, as well as Battalio & Kagel and their various frequent collaborators, were very familiar with the history of behavior modification.¹¹ Kazdin (1978) has sixty pages of references and twelve hundred names indexed. When Kagel & Winkler (1972) laid out areas of cooperative research between economics and applied behavioral analysis – a prospective collaboration that, for the record, they called *behavioral economics* -- token economies and animal experiments were widely used in the behavioral sciences but not (yet) in economics.

Interestingly, Kagel & Winkler (1972) motivated their proposal for a behavioral economics in ways similar to Morgenstern. In fact, the second edition of Morgenstern (1950), Morgenstern (1954), and Juster (1970) are explicitly referred to. As Kagel & Winkler (1972, p. 337) state succinctly, “there is a fundamental imbalance in behavioral economics between work on a slowly growing but still weak observational foundation and a proliferating super-structure of observationally uninterpreted theories and tedious arithmetic computational techniques. One need not look far for the reinforcement contingencies sustaining this research behavior.” Empirical analysis, it is then argued, is less regarded than formal mathematical reasoning, and creating your own data is more costly than using existing data even if the latter are afflicted with various inaccuracies. Acknowledging that other options exist (and in fact pointing at Morgenstern’s discussion of direct experiments and Juster’s suggestions for experimental control in economic research), Kagel & Winkler (1972) make the argument for both token economies and laboratory studies of “the behavior of animals below the human level.” (p. 339)

Kagel & Battalio (1980) is a snapshot of their accomplishments in the decade afterwards (e.g., Battalio et al. 1973; Battalio et al. 1974; Battalio et al. 1981a; Battalio et al. 1981b; to name a few); this chapter has been chosen as exhibit because the authors discuss (their) work on token economies and laboratory studies with animals below the human level in parallel and because they place emphasis on methodological issues.

McDonough (2003), pointing out that animal experiments are barely mentioned in Kagel & Roth (1995), has pointedly asked, how come that “despite confident conclusions and an admirable publication record in leading economics journals, Kagel and his associates have attracted few, if any colleagues into animal labs? ... This failure appears mysterious.” (p. 401) Indeed, Battalio et al (1991) was the last major publication coming from the Texas A & M rat lab, notwithstanding Kagel’s persuasive plea (Kagel 1987; see Loomes 1988).

¹¹ Rutherford (2009) and Lemov (2005) have provided us with highly readable accounts of the history of the many players and experiments in this space. Rutherford attempts to sketch out the huge influence that behaviorism had and, notwithstanding the naysayers, allegedly still has both within the academy and outside of it. In many respects, Rutherford’s book complements Lemov’s masterful study of the motivations (attempts at engineering human and social behavior on an ambitious scale) and drivers (general funding through various covert (e.g., the Human Ecology and related satellite projects) and not-so-covert state agencies (e.g., the CIA). In Lemov’s book Skinner takes second place to influential (at their time) researchers such as Jacques Loeb, John B. Watson, Beardsley Rummler, Clark Hull, O. Hobart Mowrer, Neal Miller, John Dollard, George Peter Murdock, Louis Jolyon (“Jolly”) West, Harold George Wolff, and Ewen Cameron. Both authors overlook the influence that their protagonists had on experimental economics, or at least that part of experimental economics that was attracted by token economies and animal experiments.

As a topic of interest to economists, token economies went into a tailspin already by the end of the 1970s, notwithstanding Kazdin's claim that the future was still bright. Rutherford (2009, p. 65, pp. 74 – 77, and chapter 4; see also Kazdin 1978, pp. 346 – 72 and Kazdin 1982, pp. 437 - 41) attributes this development to a number of practical, ethical, political and legal issues. Surely, cost must have played a role, too.

Summary

Kagel & Battalio (1980) lead into their article with a figure that presents data on the labor supply behavior of a rat working for alcohol in an operant conditioning chamber and humans working for alcohol in a simplified token economy. They stress that this was apparently “the first exemplification of such a relationship using data for individual workers to appear in literature.” (p. 380)

After discussing how to bring behavior into the laboratory (and hence addressing the issue of external validity) and why, and based on what assumptions, the small worlds of the human and animal lab can be used to test properties of the Slutsky-Hicks labor supply model, they make a methodological detour addressing the issue of why the economic behavior of nonhumans is indeed worth studying.

They argue “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. ... 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 neoclassical framework), rather than differences in genetic make-up, require enforced separation of parents from offspring. Such studies can only be performed using laboratory animals.” (p. 384)

Having discussed further the use of individual observations in hypothesis testing and theory development (and in passing – fn 8 – having made clear that token economies and laboratory animal studies almost automatically address the Wallis-Friedman critique), Kagel and Battalio address internal and external validity as criteria for experimental results, the former one addressing the issue whether indeed treatments have made a difference, the latter addressing the issue of generalizability: “To what populations, settings, and variables can the effects reported be generalized?” (p. 390)

With the subjects typically employed in token economies and laboratory animal studies being “unrepresentative”, the issue of the generalizability to “typical” behavior in national economies becomes important. Kagel & Battalio argue that trade-offs have to be made and point at their studies of consumer demand behavior in a token economy with long-term psychiatric patients (Battalio et al. 1973, 1974) which could not have been done with other human communities due to the available budget. In the following section they give additional examples, arguing that one of them was “the closest approximation to date of Oskar Morgenstern's (1954) notion of establishing communities for the explicit purpose of conducting economic experiments.” (p. 394) They also

point to the disciplining and confidence-building role of systematic and direct replication, preferably through other researchers.

A couple of pages later, they implicitly address the Wallis-Friedman critique of Thurstone's experiment, arguing that token economies and laboratory animal experiments address all concerns: "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 economics experiments (Siegel, 1961; Smith, 1976) and a serious deficiency in earlier experimental studies (see MacCrimmon & Toda, 1969, for citations to this earlier literature)." (Kagel & Battalio 1980, p. 386).

As regards internal validity, Kagel & Battalio discuss the importance of within-subject replication (ABA designs) and warn that one ought not to take for granted reversibility to the baseline A condition. The occasional failure of that reversibility makes necessary refinements in experimental methods and theory.

The methodological questions this museum piece highlights:

Token economies and laboratory animal settings, at least to the extent that they can be cleanly designed and implemented, are microcosms that address the methodological criticism that Wallis & Friedman directed at the Thurstone experiment. In fact, the arguments in favor of these two then new (for economics but by no means for biology and psychology) technologies in the experimental analysis of economic behavior seemed very persuasive indeed.

Episode Five: Siegel's work on guessing sequences

Context

"I have not the slightest doubt that if Sid Siegel had lived, say another 25–30 years, the development of experimental economics would have been much advanced in time. He was just getting started, was a fountain of ideas, a powerhouse of energy, and had unsurpassed technique and mastery of experimental science. Twenty-five years later I asked Amos Tversky, 'Whatever happened to the tradition of Sidney Siegel in psychology?' His answer: "YOU'RE IT!" (Smith, as quoted in Hertwig & Ortmann 2001, p. 442)

"Sid was more than a master experimentalist, he also used theory and statistics with skill in the design and analysis of experiments. I am persuaded that if Sid had lived he would not only have been the deserving Nobel Laureate who was well out in front of the rest of us, but also the timetable for the recognition of experimental economics would have been moved up perhaps several years." (Smith 2008, p. 198)

Simon (1959, e.g., pp. 258 – 9) discussed empirical studies on decision making under uncertainty, noting that the new axiomatic foundations of utility theory "have led to a rash of choice experiments. An experimenter who wants to measure utilities, not merely in principle but in fact, faces innumerable difficulties. Because of these difficulties, most experiments have been limited to

confronting the subjects with alternative lottery tickets, at various odds, for small amounts of money. The weight of evidence is that, under these conditions, most persons choose in a way that is reasonably consistent with the axioms of the theory – they behave as though they were maximizing the expected value of utility and as though the utilities of several alternatives can be measured. [Here footnote that refers to Edwards 1954¹² and Davidson and Suppes 1957.]” (Simon 1959, p. 258)

This assessment, and also his assessment of the literature on transitivity in the following footnote, presented a problem of sorts for Simon (1958) and his campaign against classic theory (e.g., Simon 1947, 1955). His way out was squarely aimed at the external validity of the available evidence:

“When these experiments are extended to more ‘realistic choices – choices that are more obviously relevant to real-life situations – difficulties multiply. In a few extensions that have been made, it is not at all clear that the subjects behave in accordance with the utility axioms. There is some indication that when the situation is very simple and transparent, so that the subject can easily see and remember when he is being consistent, he behaves like a utility maximizer. But as the choices become a little more complicated – choices, for example, among phonograph records instead of sums of money – he becomes much less consistent. [References to Davidson and Suppes (1957) and May (1954), with the footnote referring to Rose (1957) and the published version of Papandreou (1953); see Moscati 2007, p. 379.]”

The external validity critique is also implicit when later in his paper, Simon addresses binary choice experiments, arguing that much recent discussion about utility had centered around a particularly simple choice experiment: “This experiment, in numerous variants, has been used by both economists and psychologists to test the most diverse kinds of hypotheses. ... How would a utility maximizing subject behave in the binary choice experiment? Suppose that the experimenter

¹² Interestingly, psychologist Ward Edwards was the discussant of the May and Papandreou papers at the Chicago 1952 meeting of the Econometric Society. Edwards had just completed his Ph.D. at Harvard under John G. Beebe-Center and Mosteller, his dissertation being “an experimental study on risky choices that built on the Mosteller-Nogee experiment.” (Moscati 2007, p. 377) Let us add that Ward Edwards, by his own recognizance (Edwards 1981), as a graduate student read an article by psychologist Kurt Lewin and coauthors about aspiration levels and in the same week attended a seminar where Mosteller talked about his work:

“The obvious similarity of the ideas from such dissimilar roots stimulated me, and interacted with knowledge of economic theory obtained because my father was a prominent economist and because I had taken many economics courses as undergraduate. This confluence of ideas (which in fact, as I learned much later, converged earlier in Berlin in 1928) led to my PHD thesis and to my lifelong research interest in decision making. At Johns Hopkins, in my first post-PhD job, I decided that I could not afford my existing state of ignorance about the literature of decision making, then mostly a topic in economics and statistics. Since I had to write the stuff anyhow, I chose to get a publication out of it by writing a review. So I borrowed for three months an office in the library of the economics department, holed up all day, every day, and emerged with ‘The theory of decision making’ in virtually final-draft form. ... The topic of decision making apparently was not salient in psychological thinking as available for research and theorizing; this article made it so. ...

Economists, too, found it useful; they had not considered the possibility that the assertions about individual behavior and rationality that make up the content of microeconomics were readily subject to experimental test, and that some such tests had been performed, with dubious results.” (Edwards 1981, p. 312) Edwards (1954) became indeed a citation classic.

rewarded ‘plus’ one one-third of the trials, determined at random, and ‘minus’ on the remaining two-thirds. Then a subject, provided that he believed the sequence was random and observed that minus was rewarded twice as often as plus, should always, rationally, choose minus. He would find the correct answer two-thirds of the time, and more often than with any other strategy. Unfortunately for the classical theory of utility in its simplest form, few subjects behave in this way. The most commonly observed behavior is what is called *event matching*. [Reference to an example of data consistent with event-matching on p. 283] ... All sorts of explanations have been offered for the event-matching behavior. ... The important conclusion at this point is that even in an extremely simple situation, subjects do not behave in the way predicted by a straightforward application of utility theory.” (Simon 1959, pp. 260-1)

Or did they? Siegel & Goldstein (1959) was published the same year. It was a remarkable article because it demonstrated that rejection or acceptance of the classical theory of utility was importantly conditioned on the way the experiment was conducted.

Summary

In a widely cited and well-known article (e.g., Kagel & Battalio 1980) that summarized his explorations on binary choice games and the alleged evidence against the classical theory of utility, Siegel stated:

“This is a curious result. Since the subject is instructed to do his best to predict correctly which of the two events will occur, and since we may suppose that he is attempting to follow those instructions, should we not expect him to learn to maximize the expected frequency of correct predictions? To do this, he should tend to predict the more frequent event on all trials. We might expect him to come to such stable-state behavior after an initial learning period.” (Siegel 1961, p. 767)

To better understand this result, Siegel & Goldstein (1959) ran three treatments labeled No Payoff, Reward, and Risk, with the second treatment only rewarding good predictions and the third treatment also punishing bad predictions. The results are fairly clear, indicative, and self-explanatory:

TABLE 1
NUMBER OF Ss PREDICTING MORE FREQUENT
EVENT AT VARIOUS PROPORTIONS DURING
FINAL 20 TRIALS OF FIRST
100-TRIAL SERIES

Proportion of Predictions of More Frequent Event	Condition		
	No Payoff (N = 12)	Reward (N = 12)	Risk (N = 12)
.60	2		
.65	3		
.70	3	3	
.75	2	4	
.80	2	3	
.85		1	3
.90		1	1
.95			5
1.00			3

TABLE 2
MEAN PROPORTION OF TIMES THE MORE
FREQUENT EVENT WAS PREDICTED BY A SUB-
GROUP OF FOUR Ss, RANDOMLY SELECTED
FROM EACH PAYOFF GROUP, DURING
FINAL 20 TRIALS OF EACH 100-
TRIAL SERIES

Series	No Payoff	Reward	Risk
1-100	.69	.78	.95
101-200	.74	.85	.95
201-300	.75	.86	.95

Siegel & McMichael Andrews (1962) provided similar evidence for children.

Demonstrating the effects of financial incentives was not the only methodological theme reflected in Siegel's work. A remarkable collection of assessments in his honor (Messick & Brayfield 1964), published a couple of years after his untimely death, features Siegel's key articles but also contributions by his collaborators, students, and wife, psychologist Alberta Engwall Siegel, whose "memoir" touches¹³ on Siegel's non-scientific and scientific accomplishments.

Engwall Siegel identifies as the four major areas of research that Siegel contributed to as "statistics and research decisions", "measurement and decision making", "level of aspiration and bargaining", and "choice behavior".

Engwall Siegel also comments on the tenets that guided Siegel's scientific work (pp. 17 – 22); she stresses that Siegel believed that research ought to be guided by theory, that exploratory studies (or what today might be called fishing expeditions) had no place in research, that data ought to be analyzed only to the extent that a prior hypothesis warranted it, that remaining close to the data was important, and that experimental work was in many situations the way to go because of the control it afforded.

¹³ ... and does so touchingly, yet unsentimentally ...

Engvall Siegel (1964, pp. 19-20) first addresses the artificiality critique (Schram 2006). “Only by involving ourselves in our subjects’ everyday lives in their own milieus can we succeed in studying the important variables, those which really make a difference. The effects an experimenter can produce through his manipulation of some independent variable in an experiment are trivial and insignificant compared to the massive effects produced by the profound manipulations which nature provided. Finally, if social scientists are to succeed in finding answers to the important social questions, they will need to observe the phenomena of interest in their genuine context. Sid’s work exemplifies replies to these arguments. In the first place, the term ‘experimentation’ refers to a design and not a location. The essential features of experimental design are control and comparison, with randomization an essential part of control. Where randomization can be achieved in the field, experiments can be conducted away from any laboratory, ... “ (Engvall Siegel 1964, pp. 19-20)

“Second, experiences in a laboratory need not be artificial and removed from putative ‘real life’. With ingenuity, laboratory experiences may be arranged which are meaningful for the subjects and in which they become personally absorbed. At present, the technique for making laboratory experiences meaningful which has vogue among social psychologists is the technique of deception. ... Sid disliked and avoided deception, principally on ethical grounds, but also because an experiment involving deception creates a climate of suspicion and distrust toward psychological experimentation. In some laboratories of social psychology, the subject enters wondering what lie he is going to be told this time. ... In Sid’s, the subject entered wondering how much money he would make. I say this facetiously to introduce an alternative approach for making laboratory experiences meaningful, the one Sid employed. He built into the experimental situation features which enlisted the subject’s motivation. He believed in ‘the payoff’. ... [follows detailed discussion of papers in the Messick & Brayfield volume] ... ” (Engvall Siegel 1964, p. 20)

“The important feature of the payoff is ... that the amount of the payoff to the subject depends directly and differentially on how the subject performs in the experiment; ... Sid’s convictions about decision making convinced him that meaningful observations of social behavior could be made best where payoffs are involved, and he searched for situations in which a payoff could readily be employed. He was suspicious of any study whose measure depended on the subject’s good will or cooperativeness toward the experimenter. As any psychologist would, he always laughed at the ‘gedanken experiments’ of economists in which they try to settle empirical questions by imagining what they would do in the economic situation under control. But he parted company with many psychologists in thinking that self-report measures – adjective checklists, self-rating scales, preferences inventories, attitude questionnaires, personality inventories, and the like – are similarly suspect.” (Engvall Siegel 1964, pp. 20-1)

The methodological questions this museum piece highlights:

Siegel’s work was indeed the precursor of the experimental practices that experimental economists adopted quickly (later codified in Smith 1976, 1982). His methodological stance was affected by the Wallis – Friedman critique (which he did not cite in his major articles but which he surely knew since he referenced Mosteller & Nogee (1951) in Siegel (1957). Essentially, in his own experiments, Siegel (1961) addressed all elements of the criticism that Wallis and Friedman had formulated and additional ones (e.g., concerning the noise that deception might bring into the laboratory).

Concluding remarks

Reading the articles/papers reviewed above was, for the most part, a thoroughly enjoyable journey. Many of them have stood the test of time remarkably well and they remain, even though they may be museum pieces, excellent reads. Throughout we see a concern with the hypothetical and otherwise artificial nature of the stimuli and subjects' unwillingness to state, or – maybe -- their inability to know, their true reactions, the unrepresentativeness of subjects' response, the importance of the wording of instructions and surveys, and the fact that preferences in economic situations are unlikely to be stable. Acknowledging these issues, Morgenstern's chapter is remarkable for its blunt, confident, and ultimately correct assesement of the potential of the experimental enterprise in economics. Decades later the experimental method would become an indispensable tool in the economist's toolbox, a development that was authenticated through the Nobel Memorial Prize in Economic Sciences for Reinhard Selten (whose experimental work was explicitly mention in the citation), Vernon L. Smith, and Alvin E. Roth in 1995, 2002, and 2012, respectively.

REFERENCES

- Battalio, RC, Kagel, J, Kogut, C (1991), Experimental Confirmation of the Existence of a Giffen Good. *American Economic Review* 81, 961-70.
- Battalio, RC, Green, L, Kagel, J (1981), Income-Leisure Tradeoffs of Animal Workers. *American Economic Review* 71, 621-32.
- Battalio, RC, Kagel, JH, Rachlin, H, Green, L (1981), Commodity-Choice Behavior with Pigeons as Subjects. *Journal of Political Economy* 89, 67-91.
- Battalio, RC, Fisher, EB, Kagel, JH, Basmann, RL, Winkler, RC, Krasner, L (1974), An Experimental Investigation of Consumer Behavior in a Controlled Environment. *Journal of Consumer Research* 1, 52-60.
- Battalio, RC, Kagel, JH, Winkler, RC, Fisher, EB, Basmann RL, Krasner, L (1973), A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases. *Western Economic Journal* 11, 411-28.
- Brady, DS (1965), Book review of Juster (1964). *Journal of Marketing Research* 2, 202-4.
- Brunk ME, Federer, WT (1953), Experimental Designs and Probability Sampling in Marketing Research, *Journal of the American Statistical Association* 48, 440-52.
- Carpenter, J, Harrison, GW, List, JA (2005, Eds.), *Field Experiments in Economics*. Greenwich, CT: JAI Press, *Research in Experimental Economics*, Volume 10, 2005.
- Castro, B, Weingarten, K (1970), Towards Experimental Economics. *Journal of Political Economy* 78, 598-60.
- Chamberlin, EH (1948), An experimental imperfect market. *Journal of Political Economy* 56, 95-108.
- Davidson, D, Suppes, P (1957), *Decision Making: An Experimental Approach*. Stanford 1957 (as cited in Simon 1959)
- Edwards, WE (1981), This week's citation classic (W Edwards 1954), *Citation Classics* 6, p. 312
- Edwards, WE (1954), The theory of decision making. *Psychological Bulletin* 51, 380-417.
- Engvall Siegel, A (1964) Sidney Siegel: A Memoir. pp. 1 – 13 in Messick & Brayfield (1964)
- Friedman, JW (1969), On Experimental Research in Oligopoly. *Review of Economic Studies* 36, 399-415.
- Hertwig, R, Ortmann, A (2001), Experimental Practices in Economics: A Challenge for Psychologists? *Behavioral and Brain Sciences* 24, 383 - 403.
- Heukelom, F (2010), Measurement and Decision Making at the University of Michigan in the 1950s and 1960s. *Journal of the History of the Behavioral Sciences* 46, 189-207.

Innocenti, A, Zappia, C (2005), Thought- and performed experiments in Hayek and Morgenstern, pp. 71-97 in P Fontaine & R Leonard (Eds), *The Experiment in the History of Economics*, Routledge, London.

Juster, FT (1984), Basic Research in the Social and Behavioral Sciences. *Science New Series*, Vol. 226, No. 4675, p. 610.

Juster, FT (1974), The Use of Surveys for Policy Research. *American Economic Review* 64, 355-64.

Juster, FT (1970), Microdata, Economic Research, and the Production of Economic Knowledge. *American Economic Review* 60, 138-48.

Juster, FT (1966), Buying Intentions and Purchase Probability: An Experiment in Survey Design. *Journal of the American Statistical Association* 61, 658-96.

Juster, FT (1964), *Anticipations and Purchases: An Analysis of Consumer Behavior*. Princeton: Princeton University Press.

Juster, FT (1961), Book Review: The Powerful Consumer by George Katona. *Journal of Political Economy* 69, 503-4.

Juster, FT (1960), Prediction and Consumer Buying Intentions. *American Economic Review* 50, 604-17.

Juster, FT, Stafford, FP (1991), The Allocation of Time: Empirical Findings, Behavioral Models, and Problems of Measurement. *Journal of Economic Literature* 29, 471-522.

Kagel, JH (1972), Token Economies and Experimental Economics. *Journal of Political Economy* 80, 779-85.

Kagel, JH, Battalio, RC (1980), Token Economy and Animal Models for the Experimental Analysis of Economic Behavior. Pp. 379-401 in Kmenta & Ramsey (1980, Eds).

Kagel, JH, Battalio, RC, Rachlin H, Basmann, RL, Green L, Klemm, WR (1975), Experimental Studies of Consumer Demand Behavior Using Laboratory Animals. *Economic Inquiry* 13, 22-38.

Kagel, JH, Roth AE (1995), *Handbook of Experimental Economics*. Princeton: Princeton University Press.

Kagel, JH (1987), Economics according to the rats (and pigeons too): What have we learned and what can we hope to learn? Pp. 155-192 in Roth (1987, ed)

Kagel, JH, Winkler RC (1972), Behavioral Economics: Areas of Cooperative Research Between Economics and Applied Behavioral Analysis. *Journal of Applied Behavioral Analysis* 5, 335-42.

Kalisch, G, Milnor, JW, Nash, J, Nering, ED (1954), Some Experimental n-person Games. Pp. 301-328 in Thrall et al. (1954).

- Kazdin, AE (1982), The Token Economy: A Decade Later. *Journal of Applied Behavior Analysis* 15, 431-45.
- Kazdin, AE (1978), *History of Behavior Modification: Experimental Foundations of Contemporary Research*. Baltimore: University Park Press.
- Kmenta, J, Ramsey, JB (1980), *Evaluation of Econometric Models*. New York: Academic Press.
- Lange, O, McIntyre, F, Yntema, TO (1942), *Studies in Mathematical Economics and Econometrics: In Memory of Henry Schultz*. Chicago: Chicago University Press.
- Lemov, R (2005), *World as Laboratory. Experiments with Mice, Mazes, and Men*. New York: Farrar, Straus, and Giroux.
- Lenfant, J-S (2009), Experimental Indifference Curves and the Ordinalist Revolution. Working paper (accessed April 30, 2010: <http://hes-conference2009.com/papers/MON2D-Lenfant.pdf>)
- Loomes, G (1988), Review: Laboratory Experimentation in Economics: Six Points of View by Alvin E. Roth. *Economic Journal* 98, pp. 1206 – 8.
- MacCrimmon, KR, Toda, M (1969), The Experimental Determination of Indifference Curves. *Review of Economic Studies* 36, 1293-1323.
- May, KO (1954), Intransitivity, Utility, and the Aggregation of Preference Patterns. *Econometrica* 22, 1-13.
- May, KO (1953), The Intransitivity of Individual Preferences (Abstract). *Econometrica* 21, 476.
- McDonough, T (2003), Of Rats and Economists. *Journal of the History of Economic Thought* 25, 397-411.
- Messick, S, Brayfield, AH (1964), *Decision and Choice. Contributions of Sidney Siegel*. New York: McGraw-Hill
- Morgenstern, O (1954), Experiment and Large Scale Computation in Economics. Pp. 484-549 in Morgenstern, O (Ed), *Economic Activity Analysis*. New York: John Wiley.
- Morgenstern, O (1950), *On the Accuracy of Economic Observations*. Princeton: Princeton University Press.
- Moscato, I (2007), Early Experiments in Consumer Demand Theory: 1930 – 1970. *History of Political Economy* 39, 359-401.
- Mosteller, F, Nogee, P (1951), An Experimental Measurement of Utility. *Journal of Political Economy* 59, 371-404.
- Namais, J (1965), What Are the Odds? (book review of Juster 1964). *Journal of Marketing* 29, 108-9.
- Naylor, TH (1972), Experimental Economics Revisited. *Journal of Political Economy* 80, 347-52.

- Ortmann, A (2009), Sketching the History of Experimental Methods in the Behavioral Sciences (with particular emphasis on those practices that informed experimental practices in economics). Presentation at the Amsterdam History and Methodology of Economics Group And Cachan History of Social Science Group Fall Workshop Amsterdam School of Economics, December 11, 2009.
- Ortmann (2005) Field Experiments: Some Methodological Caveats, pp. 51-70 in J. Carpenter, G.W. Harrison and J.A. List (Eds).
- Papandreou, AG (1953), An Experimental Test of an Axiom in the Theory of Choice (Abstract). *Econometrica* 21, 477.
- Papandreou, AG et al. (1954), A Test of a Proposition in the Theory of Choice (Abstract). *Econometrica* 23, 333-4.
- Radford, RA (1945), The Economic Organization of P.O.W. Camp. *Economica* 12, 189-201.
- Rose, AM (1957), A Study of Irrational Judgements. *Journal of Political Economy* 65, 394-402.
- Rosenthal, R & Rubin, DB (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences* 3, 377-86.
- Rosenthal, R (1994), Interpersonal Expectancy Effects: A 30-Year Perspective. *Current Directions in Psychological Science* 3, 176-79
- Roth, AE (1995), Introduction to Experimental Economics, pp. 3 – 109 in Kagel & Roth (1995).
- Roth, AE (1993), On the Early History of Experimental Economics. *Journal of the History of Economic Thought* 15, 184-209.
- Roth, AE (1987, ed), *Laboratory Experiments in Economics: Six Points of View*. Cambridge: Cambridge University Press.
- Rousseas, SW, Hart, AG (1951), Experimental verification of a composite indifference curve. *Journal of Political Economy* 59, 288-318.
- Rutherford, A (2009), *Beyond the Box. BF Skinner's Technology of Behavior from Laboratory to Life, 1950's – 1970s*. Toronto: University of Toronto Press.
- Sadrieh, A, Weimann, J (2008), *Experimental Economics in Germany, Austria, and Switzerland. A collection of papers in honor of Reinhard Tietz*. Marburg: Metropolis.
- Schram, Arthur (2006), Artificiality; The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12:2, 225-237. DOI: [10.1080/13501780500086081](https://doi.org/10.1080/13501780500086081)
- Schultz, H (1933), Frisch on the Measurement of Utility. *Journal of Political Economy* 4, 95-116.
- Schultz (1938), *The Theory and Measurement of Demand*. Chicago: University of Chicago Press.

- Siegel, S (1961), Decision Making and Learning under Varying Conditions of Reinforcement. *Annals New York Academy of Sciences* 89, 766-83.
- Siegel, S (1957), Level of Aspiration and Decision Making. *Psychological Review* 64, 253-62.
- Siegel, S, Goldstein, DA (1959), Decision-Making in a Two-Choice Uncertain Outcome Situation. *Journal of Experimental Psychology* 57, 37 – 42.
- Siegel, S, McMichael Andrews, J (1962), Magnitude of Reinforcement and Choice Behavior in Children. *Journal of Experimental Psychology* 63, 337-41.
- Simon, HA (1959), Theories of Decision Making in Economics and Behavioral Science. *American Economic Review* 49, 253-83.
- Simon, HA (1956), A comparison of game theory and learning theory. *Psychometrika* 21, 267-72.
- Simon, HA (1955), *Models of Man*. New York:Macmillan.
- Simon, HA (1947), *Administrative Behavior*. New York: Macmillan.
- Smith, VL (2008), *Discovery. A Memoir*. Bloomington: AuthorHouse 2008.
- Smith, VL (1982). Microeconomic Systems as an Experimental Science. *American Economic Review* 72, 923–55.
- Smith, VL (1976), Experimental Economics: Induced Value Theory. *American Economic Review* 66, 274–9.
- Smithies Report (1955), Consumer Survey Statistics. Report of Consultant Committee on Consumer Survey Statistics. Organized by the Board of Governors of the Federal Reserve System at the request of the Subcommittee on Economics Statistics of the Joint Committee on the Economic Report.
- Thrall, RM, Coombs, CH, Davis, RI (1954, Eds), *Decision Processes*. New York: John Wiley.
- Thurstone, LL (1931), The Indifference Function. *Journal of Social Psychology* 2, 137-67.
- Von Neumann, J, Morgenstern, O (1944), *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Wallis, WA, Friedman, M (1942). The Empirical Derivation of Indifference Functions. Pp. 175 – 189 in Lange, McIntyre, & Yntema (1942).
- Zizzo, DJ (2010), Experimenter Demand Effects in Economic Experiments, *Experimental Economics* 13, 75-98.
- Wordcount: 8713