Measuring the Economizing Mind in the 1940s and 1950s

The Mosteller-Nogee and Davidson-Suppes-Siegel Experiments

to Measure the Utility of Money

Ivan Moscati*

Forthcoming in *History of Political Economy*, 48, annual supplement, 2016.

September 2015

* Correspondence may be addressed to Ivan Moscati, Department of Economics, University of Insubria, Via Monte Generoso 71, 21100 Varese, Italy; e-mail: ivan.moscati@uninsubria.it. I am grateful to Jeff Biddle, Simon Cook, Catherine Herfeld and the other participants in the 2015 HOPE conference for helpful discussions and suggestions on earlier drafts of the paper. This paper has also benefited from insightful comments made by Colin Camerer, the participants in a session at the 2015 ASSA meeting, and an anonymous referee. I also thank Fondazione Cariplo for financial support, and Dina Melnichuk for research assistance in Friedman's archives at the Hoover Institution Library. Any errors are mine.

Abstract

The paper studies the origin, content and impact of two experiments to measure the utility of money. The first experiment was performed between 1948 and 1949 by F. Mosteller and P. Nogee, and grew directly out of Mosteller's discussions with M. Friedman and L.J. Savage. The second was carried out in 1954 by D. Davidson, P. Suppes, and S. Siegel. Both experiments relied on expected utility theory (EUT), and both groups of experimenters concluded that their findings supported the measurability of utility as well as EUT.

For a number of reasons, the two experiments provide a case study that illuminates the interaction between economics and psychology in the 1940s and 1950s. First, their designs exhibit a tension between the economic image of human agency associated with EUT, and insights from experimental psychology research that were in contrast with EUT's assumptions. Second, both experiments were performed by psychologists and other non-economists, and the paper reconstructs how their authors became interested in measuring the utility of money. Third, the paper shows that the psychological insights contained in the designs of the two experiments found some application between 1955 and 1965, but were quickly forgotten afterwards. Only in the 1970s, when robust experimental evidence against EUT accumulated, were economists compelled to re-consider those psychological insights.

Keywords

Experimental economics; Experimental psychology; Utility measurement; Expected Utility Theory; Subjective probability.

JEL codes

B21: History of Economic Thought since 1925; Microeconomics.
B31: History of Economic Thought: Individuals.
C91: Design of Experiments; Laboratory, Individual Behavior.
D81: Criteria for Decision-Making under Risk and Uncertainty.

This paper explores the relationships between economics and psychology in the 1940s and 1950s by investigating the origin, the content, and the influence on economic analysis of two experiments on individual decision making. Both experiments had the same goal and overall structure: they aimed at measuring the utility of money of a number of individuals on the basis of their preferences between gambles where small amounts of real money were at stake. Both experiments relied on the theory of choices involving risk that in those years was emerging as the dominant one in economics, namely expected utility theory (EUT), with the utility measures generated by both experiments providing an indirect test of EUT's validity.

The first experiment was performed between 1948 and 1949 at Harvard University by Frederick Mosteller, a statistician who in the late 1940s had become interested in experimental psychology, and Philip Nogee, then a Harvard Ph.D. student in psychology. The Mosteller-Nogee experiment grew directly out of Mosteller's discussions with economist Milton Friedman and statistician Leonard Jimmie Savage. The second experiment was carried out at Stanford University in spring 1954 by Donald Davidson and Patrick Suppes, two analytical philosophers, and Sidney Siegel, who was then completing his Ph.D. in psychology at Stanford with an experimental dissertation on the psychological determinants of authoritarianism. Davidson, Suppes, and Siegel took into account some aspects of the psychology of decision that had been neglected by Mosteller and Nogee in order to neutralize them and thus obtain more reliable utility measures. Both groups of experimenters concluded that their findings supported the experimental measurability of utility as well as EUT.

Although these two experiments do not tell the whole history of the interdisciplinary interaction between economics and psychology in the 1940s and 1950s, they provide a case study that, for a number of reasons, allows us to illuminate a significant part of that interaction. First, the designs of the two experiments exhibit a tension between the economic image of human agency associated with EUT, and insights from experimental psychology research carried on since the 1920s. For instance, EUT assumes that individuals' preferences over gambles are fixed and do not fluctuate: if an individual states that he prefers gamble A to gamble B, he won't reverse his preference if asked to compare the two gambles again. However, experimental psychologists such as Louis Leon Thurstone (1927) had called attention to the fact that, because of judgment errors, distraction, or variation in sensibility, individuals' comparative judgments do fluctuate. Mosteller and Nogee incorporated Thurstone's psychological insight into the design of their experiment. Or, to take another example, EUT assumes that individuals do not distort objective probabilities and therefore evaluate the event "heads" in tossing a fair coin as equally likely to the event "tails." Yet, experimental psychologists such as Malcolm Preston and Philip Baratta (1948), Richard Griffith (1949) and Ward Edwards (1953, 1954a, 1954b) had shown that in fact individuals do distort objective probabilities, even simple ones. In designing their experiment, Davidson, Suppes, and Siegel took account of this problem and contrived a cunning device to overcome it.

The second reason to focus on the two experiments is sociological in nature. In the 1940s and 1950s, and in fact until recent times, economists typically lacked the expertise to perform experimental research. Therefore economic experiments were often carried out by experimental psychologists and occasionally by other non-economists. This was also the case with our two experiments, and the paper reconstructs how their authors, who were not economists, became interested in measuring an archetypal economic object such as the utility of money using a theory of decision making that was cultivated chiefly by economists. As we will see, the relationship between economists and psychologists did not fit a simple model of division of labor, in which the economists supplied the theory and the psychologists provided the experimental technology. The psychologists actively contributed to identifying and also eliminating the "disturbing causes" that could spoil the significance of the experimental measurements of utility.

Third, the limited influence of the two experiments on economics during the late 1950s and 1960s shows, among other things, that in that period economists were not interested in integrating into decision theory the psychological insights contained in the experiments. To do so would clearly have been a difficult task, and since the two experiments suggested that EUT performed pretty well as a predictor of choice under risk, it could be set aside as non-urgent. Only in the 1970s, when robust experimental evidence against EUT accumulated, were economists compelled to re-consider the psychological phenomena discussed by Mosteller, Nogee, Davidson, Suppes and Siegel.

1. Setting the Stage

1.1. Thurstone's Probabilistic Method of Measuring Sensations

Since the rise of psychometrics in the second half of the nineteenth century, psychologists have tried to measure sensations and intellectual abilities in an experimental way (Michell 1999). In the 1920s, the American psychologist Louis Leon Thurstone put forward a probabilistic method of measuring sensations known as the "method of comparative judgment." In Thurstone's approach, a single subject is confronted with pairs of stimuli, e.g. pairs of lights, and asked to compare them with respect to some dimension, e.g. brightness. Because of judgment errors, distraction, or variation in sensibility, the subject's comparative judgment "is not fixed. It fluctuates" (Thurstone 1927, 274). As a consequence, when the subject is confronted more than once with the same pair of stimuli A and B, sometimes he will rank A over B, and sometimes B over A. Thurstone used the frequency of the comparative judgments to rank the stimuli: the light that is perceived as brighter more than fifty per cent of the time is taken to be brighter. As we will see, Mosteller and Nogee employed Thurstone's probabilistic approach to give an empirical content to the economic notion of indifference.

1.2. Friedman and Wallis' Critique of Thurstone

Since the so-called "marginal revolution" of the 1870s, the economic theory of decision-making has been largely based on the concept of utility. However, the fact that utility cannot be observed and measured in a straightforward way ensured that this concept soon became the subject of controversy (Moscati 2013 and 2015a). In 1930 Thurstone made a foray into economics, performing a pioneering experiment aimed at measuring the indifference curves of an individual (Thurstone 1931; for a discussion, Moscati 2007). In 1942 Milton Friedman and Allen Wallis, the latter a statistician who had studied at Chicago with Friedman, criticized Thurstone because in his study the experimental subject had to choose from fictional rather than actual commodities: "For a satisfactory experiment it is essential that the subject give actual reactions to actual stimuli. […] Questionnaires or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement" (Wallis and Friedman 1942, 179-180). This criticism influenced the design of the Mosteller-Nogee experiment.

1.3. Measuring Utility through EUT

A new approach to utility measurement was suggested by John von Neumann and Oskar Morgenstern in their *Theory of Games and Economic Behavior* (1944). They introduced a set of axioms on the individual's preferences that impose, among other things, the following constraints. First, the individual harbors well-defined preferences over all gambles, independently of whether these gambles are simple or complex; accordingly, for any pair of gambles A and B, he either prefers A to B, B to A, or judges them indifferent. Second, the individual must understand correctly the objective probabilities of uncertain events, even complicated ones such as 16.67. Third, he cannot distort objective probabilities by reinterpreting them according to some subjective disposition. Finally, the individual does not obtain utility from the act of gambling, but only from the gamble's payoffs. Von Neumann and Morgenstern proved that an individual satisfying these and other axioms – let us call him an "EUT decision maker" – prefers the gamble associated with the highest expected utility.

Von Neumann and Morgenstern also indicated a handy way in which EUT could be used to measure utility experimentally. If an EUT decision maker is found to be indifferent between, say, a gamble yielding \$500 with probability 0.4 and \$1,000 with probability 0.6 – such a gamble can be written as [\$500, 0.4; \$1,000, 0.6] – and \$600 for sure, then we can infer that for him u(\$600) is equal to 0.4×u(\$500)+0.6×u(\$1,000), where u is the individual's utility function. The EUT axioms also imply that the function u is cardinal in nature, i.e., that it is unique up to linearly increasing transformations, so that only two points of it are arbitrary.[1] Thus we can state that u(\$500)=0 and u(\$1,000)=1, and establish that for the EUT decision maker u(\$600)=0.4×0+0.6×1=0.6.

Beginning in the mid-1940s, the plausibility of von Neumann and Morgenstern's axioms and the validity of EUT became the subject of an intense debate in which Friedman and Savage (1948, 1952), Jacob Marschak (1950, 1951), Paul Samuelson (1950, 1952), Kenneth Arrow (1951), Harry Markowitz (1952), Armen Alchian (1953), Robert Strotz (1953), Maurice Allais (1953), Daniel Ellsberg (1954) and other major economists of the period took part.[2] The Mosteller-Nogee experiment played some role in this debate.

---

[1] That u is unique up to linearly increasing transformations means that any utility function u′ obtained by multiplying u by a positive number α and then adding any number β, that is, any u′ such that $u' = \alpha u + \beta$, with $\alpha > 0$, still represents the individual's preferences. The two arbitrary points of a cardinal function reflect the fact that α and β are arbitrary.

[2] For reviews of the debate on EUT in economics, see Fishburn 1989, Fishburn and Wakker 1995, Giocoli 2003, Mongin 2009 and 2014, Heukelom 2014, Moscati 2015b. Although economists investigated and applied EUT more intensively than scholars in other disciplines, the theory was also discussed and used by

One of the early articles in support of EUT was coauthored by Friedman and Savage (1948). The authors here clarified how EUT can be fruitfully applied to the analysis of gambling and insurance, introduced concepts such as risk aversion that quickly became central to the theory of risky decisions, and suggested how the experimental measures of utility obtained on the basis of EUT could be used to test the validity of EUT. If, to continue the previous numerical example, in a further experiment the EUT decision maker is found to be indifferent between gamble [$10,000, 0.2; $500, 0.8] and $1,000 for sure, then he should also be indifferent between gamble [$10,000, 0.12; $500, 0.88] and $600 for sure.[3] If this is not the case, noted Friedman and Savage, "the supposition that individuals seek to maximize expected utility would be contradicted" (1948, 304). The strategy suggested by Friedman and Savage to test EUT was promptly implemented by their friend Mosteller.

2. Friedman, Savage, and the genesis of the Mosteller-Nogee Experiment

Frederick Mosteller (1916-2006) first studied mathematics and statistics at the Carnegie Institute of Technology, and in 1939 began his Ph.D. at Princeton's Department of Mathematics. At a tea in Princeton in 1941 he met Savage, who had studied at the University of Michigan and was spending the academic year 1941-1942 at the Institute for Advanced Study, also located in Princeton, as a post-doctoral student and von Neumann's assistant. Savage and Mosteller quickly became buddies and ended up teaching together an

philosophers, psychologists, mathematicians and other non-economists. On the reception of EUT outside economics, see Edwards 1961, Luce and Suppes 1965, Erikson and others 2013, and Heukelom 2014.

3 If the EUT decision maker is indifferent between [$10,000, 0.2; $500, 0.8] and $1,000 for sure, then u($1,000)=0.2×u($10,000)+0.8×u($500); but u($1,000)=1 and u($500)=0, and therefore u($10,000)=1/0.2=5. The expected utility of [$10,000, 0.12; $500, 0.88] is 0.12×5+0.88×0=0.6, which is exactly the utility of $600 for sure.

undergraduate course in algebra and trigonometry at Princeton University (Mosteller 2010). In 1942, however, Savage left Princeton for Cornell University.

In 1944 Mosteller interrupted his studies and moved to New York to work at the Statistical Research Group (SRG) at Columbia University, a wartime think-tank providing statistical analysis for the U.S. Army and directed by Allen Wallis (Wallis 1980). At SRG, Mosteller found Savage, who had joined the SRG some months earlier. Friedman had been at SRG since March 1943, and had also become acquainted with Savage. At SRG, Mosteller, Savage, and Friedman worked together on various different projects: a paper on sequential statistical estimators (Girshick, Mosteller and Savage 1946), a paper designing a metallurgical experiment to find the alloy that maximizes the time to rupture under a given stress (Friedman and Savage 1947),[4] and a book on acceptance sampling (Freeman, Friedman, Mosteller, and Wallis 1948).

When the war ended and SRG was dismantled, Mosteller, Savage and Friedman took different paths. Mosteller returned to Princeton and completed his Ph.D. with a dissertation in statistics. In 1946 he joined the faculty of Harvard's Department of Social Relations, then chaired by sociologist Talcott Parsons. The Department hosted disciplines as diverse as psychology, sociology, and anthropology, and was endowed with a laboratory for psychological experiments. As for Friedman, in 1946 he returned to the University of Chicago, where he began teaching a course in price theory that revived his interest in utility analysis. Savage also joined the University of Chicago in 1946. There, he and Friedman began collaborating on the EUT article mentioned above (Friedman and Savage 1948).

Meanwhile, working in the interdisciplinary environment of Harvard's Department of Social Relations, Mosteller became increasingly involved in psychology. Beginning in 1947-

[4] Friedman and Savage did not conduct any actual experiment but only discussed how, based on statistical theory, such an experiment should be carried out.

1948, he co-authored with some Harvard colleagues various papers aimed at measuring visual stimuli and perspective illusions and pain sensations (Bruner, Postman, and Mosteller 1950, Keats, Beecher, and Mosteller 1950). In 1948-1949, Mosteller also taught a course on psychometric methods. Given this background, it is not surprising to find that in late 1947 or early 1948 he conceived of an experiment to measure utility:

> Von Neumann and Morgenstern's book, *Theory of Games and Economic Behavior*, led to much research on ideas of utility, and I thought it would be worth actually trying to measure utility in real people to see what would happen. (Mosteller 2010, 196)

Beginning in February 1948, Mosteller discussed the design of the experiment in a number of letters with Friedman and Savage. Wallis also contributed to the discussion, albeit in a minor way.[5] In a letter dated February 27, 1948, and addressed to both Friedman and Savage, Mosteller discussed which payoffs and winning probabilities the gambles in the experiments should have ("I have thought over Jimmy's idea that we should maintain true odds at 1:1 […] but I don't see where it gets me"), and how to phrase the instructions for the experimental subjects ("One phrase I have considered is 'Try to make as much money as you can'. But this has drawbacks").[6] Friedman and Savage gave prompt feedback to Mosteller, and supported his experimental project.[7]

In late February and early March 1948, Mosteller conducted a couple of pilot studies for the experiment, and it is probably at this point that he co-opted into the project Philip Nogee (1916-1980), then a Harvard Ph.D. student in clinical psychology whom Mosteller was supervising. According to Mosteller (2010, 196), Nogee was "very quantitatively inclined and

[5] See Milton Friedman Papers, Box 30, Folder 37, and Box 39, Folder 10.

[6] Friedman Papers, Box 39, Folder 10.

[7] See in particular Friedman to Mosteller, March 3, 1948, and Savage to Mosteller, March 28, 1948; both letters are in Friedman Papers, Box 39, Folder 10. On Friedman's stance toward experimental research in the late 1940s and 1950s, see Moscati 2007.

eager to participate in designing and executing such an experiment." In his correspondence with Friedman and Savage, Mosteller did not mention Nogee by name but, beginning in March 1948, when referring to the experiment he often used the first-person plural.[8] The hidden figure within Mosteller's "we" was certainly Nogee.

Mosteller and Nogee conducted the actual experiment a year after the pilot study, i.e., from February through May 1949, and devoted the second part of 1949 and early months of 1950 to analyzing the experimental data. From time to time, Friedman asked Mosteller about the progress of the project.[9] The article was eventually published in the October 1951 issue of the *Journal of Political Economy* under the title "An Experimental Measurement of Utility." By that time, Nogee had left Harvard to become an assistant professor at Boston University.

3. Mosteller and Nogee's Experiment

Mosteller and Nogee (henceforth MN) presented their article as almost an outgrowth of Friedman and Savage's 1948 paper, and addressed two main questions: whether utility can be measured, at least "in a laboratory situation," and whether utility, if measurable, "can be used to predict behavior" (1951, 371-372). I first give an overview of the MN experiment and its main results, and then discuss the tension between the economic and psychological elements of its design. To appreciate that tension, going into some of the experiment's technical details is necessary.

---

[8] Mosteller to Friedman, March 8, 1948: "We have run off one session"; Mosteller to Savage, April 5, 1948: "I think we are closing in pretty sharply on the utility curve with a 5¢." See Friedman Papers, Box 39, Folder 10.

[9] Friedman to Mosteller, December 13, 1949: "How is the gambling experiment?" Friedman to Mosteller, February 22, 1950: "I am extremely curious how the experiment has been coming out. Does it live up to its premise of last spring? Your paper should be fascinating." See Friedman Papers, Box 39, Folder 37.

3.1. Overview

MN carried out their experiment at the Laboratory of Social Relations at Harvard University. The experimental subjects were ten Harvard undergraduate students and five Massachusetts National Guardsmen. Each had to choose, a number of times, whether to participate in a gamble where they could win or lose small amounts of actual money, or refuse the gamble. Subjects played with $ 1.00, which they received at the beginning of each experimental session.

The experiment had two parts. In the first, the subjects faced "simple-hands" gambles with the following structure. The experimenters showed a card with five numbers called a "hand," as in the game of poker dice, e.g., 66431. The card also indicated an amount of money – call it $M – that a subject would win if, by rolling five dice, he "beat" the displayed "hand," whereby the strength of the hands is calculated as in poker. Thus, for instance, hand 22263 ("three of a kind") beats 66431 ("pair"). The amounts of money used by MN ranged from a minimum of 2.5¢ to a maximum of $5.07. If the rolled five dice did not beat the displayed hand, the subject lost 5¢. The subjects were informed about the actual odds for each hand, i.e., about the ratio between the probability of winning over that of losing.

By confronting the experimental subjects with different gambles of this form, MN identified a number of monetary amounts $M and odds *o* for which the experimental subjects were indifferent between participating or not in the gamble. Based on this information and the assumption that u($0)=0 and u(–5¢)=–1, MN measured the utility that seven amounts of money had for their experimental subjects. In particular, MN identified the amounts of money corresponding to the utility-levels 0.5, 1, 2, 5, 10, 20, and 101, thus identifying seven points of the utility function for money of each subject. By connecting these seven points by straight lines, they drew an approximate graph of the utility curves for money of the subjects. For instance, Figure 3b of their article, which is reproduced in Figure 1 below, shows the estimated utility curve for money of experimental subject B-IV.

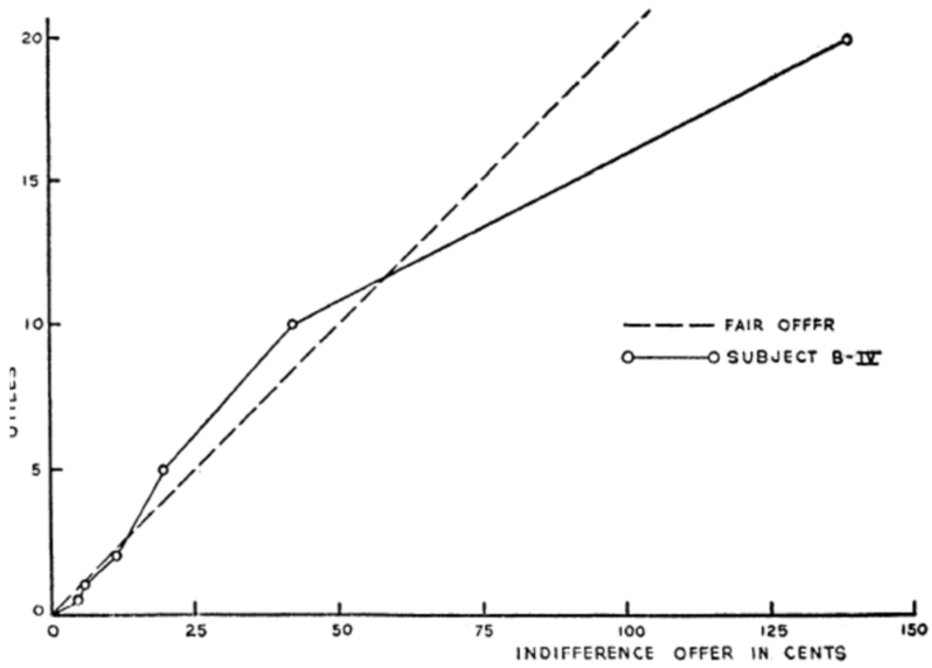

Figure 1. The continuous broken line shows the utility curve for money of subject B-IV. The utility-level 101 is not shown. The dotted straight line is the hypothetical utility function of an individual neutral to risk. *Source*: Mosteller and Nogee, 1951, 387.

In the second part of the experiment, MN used the utility functions elicited in the first part to predict whether the subjects would accept or reject more complicated gambles called "doublet hands." In a doublet-hand situation, subjects were faced with two hands, such as 66431and 22263, the odds of beating only the weaker hand (66431), the odds of beating both hands, and the associated winnings. If, by rolling the five dice, the subject did not beat either of the hands, he lost 5¢.

If the utility function obtained in the first part of the experiment from simple hands could be used to predict the subjects' choices over doublet hands, then EUT and the utility measurements based on it would be validated.

3.2. Findings

In the first part of their experiment, MN were able to elicit the utility curves for money of all but one of the experimental subjects.[10] MN's first conclusion was therefore that "it is feasible to measure utility experimentally" (403). In particular, MN found that while the Harvard students tended to have concave utility curves for money, i.e. to be risk adverse, the Massachusetts National Guardsmen tended to have convex utility curves for money, i.e. to be risk seeking.

The findings of the second part of the experiment were less clear-cut but, nevertheless, MN assessed them in a favorable way: predictions of choice behavior over doublet hands based on simple hands "are not so good as might be hoped, but their general direction is correct" (399). Thus, MN concluded that "the notion that people behave in such a way as to maximize their expected utility is not unreasonable."

3.3. Design

Many elements in the design of the MN experiment rely on the economic image of human agency embodied in the EUT decision maker. I begin discussing these elements (items i-v), and then move to the parts of the experimental design that were shaped by psychological considerations (items vi-vii).

[10] The outlier was a Harvard undergraduate whose behavior was "so erratic that no utility curve at all could be found for him" (Mosteller and Nogee, 1951, 385).

i) *Real Money*. In the psychometric experiments carried out by psychologists, subjects had to respond to stimuli, e.g. had to say which light they perceived as brighter, but did not receive any reward or penalty depending on their responses. In contrast, MN introduced actual monetary rewards into their experiment, and devoted part of their article to arguing that these rewards represented nontrivial incentives for their subjects (376, 402-403).

Neither in the correspondence between Mosteller, Friedman, Savage and Wallis that I perused, nor in the MN article, is there is any reference to Friedman and Wallis' 1942 critique of Thurstone's experiment (see section 1.2). However, MN's commitment to real and nontrivial monetary incentives implicitly agrees with Friedman and Wallis' claim that experiments in which subjects give "conjectural responses to hypothetical stimuli" are of little interest for economists.[11]

ii) *Objective Probabilities*. MN informed their experimental subjects about the actual statistical odds, i.e., about the objective probabilities, of winning and losing. Around the time MN performed their experiment, empirical studies on betting behavior carried out by psychologists Malcolm Preston and Philip Baratta (1948) and Richard Griffith (1949) suggested that bettors harbor "psychological probabilities" that do not coincide with the corresponding mathematical probabilities. In particular, these studies indicated that bettors overvalue low mathematical probabilities and underestimate high ones. Around the same time, the idea of a subjective approach to probability was promoted by the rediscovery of Frank Ramsey's essay "Truth and Probability" ([1926] 1950).

[11] The issue of whether monetary incentives are really necessary to motivate experimental subjects is still a subject of controversy between economists, who typically argue that monetary incentives are indeed necessary, and psychologists, who generally do not think so. For recent discussions, see Guala 2005, and Fréchette and Schotter 2015.

Although MN cited Ramsey a number of times, and discussed at length the issue of whether subjective (or psychological) and objective (or mathematical) probabilities coincide, their experimental design is based on objective probabilities only and, therefore, on the implicit assumption that experimental subjects do not distort objective probabilities.

iii) *Complicated Probabilities*. MN presented their experimental subjects with gambles having quite complicated odds. The winning odds for the simple hands the subjects faced in the first part of the experiment were: 1:0.50, 1:1.01, 1:2.01, 1:5.00, 1:10.17, 1:20.24, and 1:101.32.[12] The winning odds for the doublet hands used in the second part of the experiment were even more complicated. Although for an EUT decision maker all probability figures are equally comprehensible, from a psychological viewpoint it is natural to argue that subjects may find it hard to understand odds and probability figures like those used by MN.

iv) *Gamble vs. Sure Outcome*. In the MN experiment subjects had to choose between a proper gamble, in which they could win $M or lose 5¢, and the sure outcome associated with the refusal of gambling. In the EUT framework there is no difference between gamble-vs-gamble choices and gamble-vs-sure-outcome choices because EUT rules out the existence of a specific utility or disutility deriving from the very act of gambling. However, if such specific utility for gambling exists – as seems plausible from a psychological viewpoint – such utility would distort utility measures like those obtained by MN. A positive utility for gambling would in fact lead to overestimation of the utility of sure amounts of money, while a negative utility for gambling would have the opposite effect.

[12] These winning odds correspond, respectively, to winning probabilities 66.67, 49.75, 33.22, 16.67, 8.95, 4.71, and 0.98.

v) *Different Types of Gambles.* MN used the utility measures elicited from choices involving simple hands to predict choices involving doublet hands, which were much more complex. Again, from the economic viewpoint this is perfectly legitimate because an EUT decision maker displays well-defined preferences over all gambles, independently of whether these gambles are simple or complex. From a psychological standpoint, however, using utility measures elicited from simple gambles to predict choices over complex gambles may appear questionable.

Although the above elements of the MN experimental design rely on the image of human agency associated with EUT, other elements suggest a sensitivity to psychological considerations that EUT rules out.

vi) *Understanding Money*

MN's utility measurement was based on the identification of monetary amounts $M and odds *o* such that the experimental subjects were indifferent between participating in a gamble yielding $M with odds *o*, and rejecting it. In principle, the indifferent gamble can be identified in two ways, namely (a) by fixing $M and adjusting *o* until the indifference point is reached, or (b) by fixing *o* and adjusting $M until the indifference point is reached. Within the EUT framework these two methods are equivalent because the EUT decision maker understands amounts of money and odds, i.e., probabilities, equally well. However, from a psychological perspective it can be argued that individuals are more familiar with amounts of money than with probabilities, and that, therefore, they are more capable of adjusting the former than the latter. Based on these considerations, method (b) should be preferred. This is in fact MN's stance:

> The experimenters preferred to […] search for a [monetary amount] A that would bring a balance [i.e., indifference]. […] Most people are more familiar with amounts of money than with probabilities. (Mosteller and Nogee 1951, 373).

vii) *Probabilistic Indifference*. Although the EUT-based method of measuring utility is grounded on the identifications of indifferent gambles, it is difficult to observe indifference in an actual experimental setting. Experimental subjects typically agree participation in a gamble or turn the gamble down, but it is not clear what the behavioral correlate of indifference would be: dithering? That the subject dithers for some sufficiently long time?

To circumvent the problem, MN resorted to Thurstone's probabilistic approach to comparative judgment and defined a subject indifferent if, when confronted with a gamble more than once, he accepted and rejected it equally often: "When […] B and D are chosen *equally often*, i.e., each chosen in half of their simultaneous presentations, the individual is said to be indifferent between B and D" (374).

This definition of indifference was accompanied by an explicit criticism of the non-probabilistic approach to preference built into the EUT axioms. MN argued that "subjects are not so consistent about preference and indifference as Von Neumann and Morgenstern postulated" (404). Rather, gradation of preference is the rule, as "the experience of psychologists with psychological tests has shown" (374).

3.4. Summing Up

Mosteller and Nogee were, respectively, a statistician turned experimental psychologist and a psychologist. Nevertheless, the design of their experiment owes much to the economic image of human agency. Some elements of their design – such as the idea of having experimental subjects adjust amounts of money rather than probabilities, or the probabilistic approach to indifference – do rely on psychological insights, but many other elements are tailored to the EUT decision maker. In particular, the MN design rules out the possibility that actual individuals may distort objective probabilities, misunderstand complicated odds, derive utility from the very act of gambling, or behave differently when faced with different types of gambles. The fact that in the MN design the tension between the economic understanding of human agency associated with EUT and the psychological insights conflicting with it is resolved very much in favor of the former appears to reflect Friedman's and Savage's influence on Mosteller.

4. The Rise of EUT, 1950-1954

In 1950 EUT was still only one among many theories of decision under risk on the economists' table (Arrow 1951); but from the early 1950s its fortunes rapidly improved. Marschak (1950) and others provided axiomatizations of EUT more transparent and compelling than that offered by von Neumann and Morgenstern. Marschak (1951) also articulated a normative argument in favor of EUT according to which the theory indicates how rational individuals should choose.

The end of a first phase of the debate on EUT and the rise of the theory as the mainstream economic model for decision making under risk can be associated with an important international conference that took place in Paris in May 1952. At this conference, the opponents of EUT headed by Allais faced a forceful group of supporters of the theory – collectively labeled by Allais as the "American School" – that included Arrow, Friedman, Marschak, Savage, and Samuelson. Samuelson also organized a symposium on EUT that was published in the October 1952 issue of *Econometrica* and collected a number of articles backing EUT.[13] The publication of Savage's *Foundations of Statistics* (1954) reinforced the

[13] On the Paris conference, see Jallais and Pradier 2005 and Mongin 2014. Samuelson was initially a severe critic of EUT but correspondence and discussion with Savage, Marschak and Friedman led him to endorse the theory; on Samuelson's conversion to EUT, see Moscati 2015b.

dominant status of EUT in economics. Building on Ramsey's subjective approach to probability, Savage showed how to extend EUT to the case in which the probabilities of uncertain outcomes are not objectively given but express the decision maker's subjective beliefs about the likelihood of the outcomes.

The MN experiment played some role in the rise of EUT. Between 1952 and 1954, advocates of EUT argued that the experiment provided reliable empirical support to the theory (Friedman and Savage 1952, Markowitz 1952, Alchian 1953, Strotz 1953, Marschak 1954, Savage 1954). Critics of EUT, by contrast, claimed that the MN results were inconclusive or hardly extensible to more realistic situations (Manne 1952, Allais 1953, Ellsberg 1954). At any rate, the MN experiment showed that EUT has clear empirical implications, can be used to make predictions, and can therefore be falsified by experimental findings. This was not the case for other theories of decision under risk such as Allais'. More importantly, even if inconclusive or hardly extensible to more realistic situations, the results of the MN experiment failed to contradict EUT. Therefore, supporters of the theory such as Friedman and Savage (1952, 466) could claim that the experiment justified some "mild optimism" about the validity of EUT.

The rise of EUT as the mainstream economic model for risky choices in the early 1950s explains why Davidson, Suppes, and Siegel (henceforth DSS), when they entered the economics of decision-making from their home disciplines, focused on EUT rather than other theories of choice under risk. But before we turn to DSS and outline how they embarked on the experimental measurement of utility, a further contextual element should be mentioned.

As illustrated in section 3.3, Mosteller and Nogee had discussed the possibility that subjective and objective probabilities may not coincide. Ward Edwards, another of Mosteller's Ph.D. students in psychology at Harvard, investigated the issue further (on Edwards, see Shanteau, Mellers, and Schum 1999). In his doctoral dissertation and a series of papers derived from it, Edwards (1953, 1954a, 1954b) presented experimental results confirming that bettors harbor subjective probabilities that are at odds with objective probabilities. In particular, Edwards (1953, 363) pointed out that this fact "has serious implications for the utility curves of Mosteller and Nogee and indeed for the whole method of utility measurement proposed by von Neumann and Morgenstern." In fact, if subjective and objective probabilities are different, and the experimenter has no clue about the former, the expected-utility formula presents two unknowns – the utilities of money *and* the subjective probabilities – so that it becomes useless for quantifying the utilities of money in terms of probabilities. In designing their experiment, DSS took account of this problem and contrived a cunning device to overcome it.

5. Suppes, Davidson, Siegel, and the Stanford Value Theory Project

Patrick Suppes (1922-2014) studied physics and meteorology at the University of Chicago (B.S. 1943) and, after serving in the Army Air Force during the war, in 1947 entered Columbia University as a graduate student in philosophy (Suppes 1979). At Columbia he came under the influence of the philosopher and measurement theorist Ernest Nagel, and also took courses in advanced mathematical topics. Around 1948 he was one of a group of Columbia Ph.D. students who organized an informal seminar on von Neumann and Morgenstern's *Theory of Games*. He graduated in June 1950 and, in September of the same year, joined the Department of Philosophy at Stanford University, where he remained for the rest of his working life.

Suppes' early publications had little if anything to do with economics, psychology or the behavioral sciences in general. Rather, Suppes worked on the theory of measurement and the foundations of physics using a strict axiomatic approach. It must be noted, however, that in a footnote in his first article, Suppes (1951, 104, footnote 2) criticized the way von Neumann and Morgenstern defined the relation of indifference in their *Theory of Games*. This passing criticism shows that from the early stages of his scientific career Suppes had been familiar with von Neumann and Morgenstern's axiomatic approach to utility theory.

Suppes' knowledge of game and decision analysis increased in the early 1950s under the influence of J.C.C. "Chen" McKinsey, Suppes' postdoctoral tutor at Stanford. McKinsey (1908-1953) was a logician who had worked intensively on game theory at the RAND Corporation, a think tank created by the U.S. Air Force in 1946 and located in Santa Monica, California. In 1951 McKinsey joined Stanford's Philosophy Department, having being forced to leave RAND because his homosexuality was considered a security risk (Nasar 1998). At the time, he was completing his *Introduction to the Theory of Games* (McKinsey 1952), which became the first textbook in game theory. Suppes' familiarity with game theory and decision analysis was further enhanced by his summer research position in the early 1950s, working with David Blackwell and Meyer A. Girshick while they were writing their book *Theory of Games and Statistical Decisions* (1954), in which the tools of decision and game theory were employed to evaluate statistical procedures.

In 1953 John Goheen, the chair of Stanford's Department of Philosophy, obtained a grant from the Ford Foundation for a study on "Value, Decision and Rationality." Around the same time, Goheen negotiated contracts with two military agencies, namely the Office of Naval Research and the Office of Ordnance Research of the U.S. army, for work on the theory of decisions involving risk. Goheen entrusted McKinsey and Suppes with the project, which was renamed the "Stanford Value Theory Project" (Suppes 1979, Isaac 2013).[14]

[14] On the reasons why in the 1950s the Ford Foundation and military agencies such as the Office of Naval Research were interested in funding research on decision making, see Pooley and Solovey 2010, Erikson and others 2013, and Herfeld, this volume.

McKinsey and Suppes co-opted into the enterprise Donald Davidson, another philosopher who had joined the Philosophy Department in January 1951. Davidson (1917-2003) had studied at Harvard University (B.A. 1939, Ph.D. 1949), where he was influenced by logician and analytical philosopher W.V.O. Quine. Today, Davidson is best known for his influential work on the philosophy of mind and action, the philosophy of language, and epistemology. However, he published these works only from the early 1960s on. In the 1950s he was still very busy with teaching and did not as yet have a clear philosophical project. As he explained in a later interview: "Suppes and McKinsey took me under their wing […] because they thought this guy [i.e., Davidson] really ought to get some stuff out" (Davidson in Lepore 2004). Most of the research connected with the Stanford Value Theory Project was conducted between 1953 and 1955, and appeared in print between 1955 and 1957. However, McKinsey contributed only to the first part of the project because in October 1953 he committed suicide.

The final output of the Stanford Value Theory Project consisted of three articles, which were theoretical in nature,[15] and a book presenting the results of the experiment to measure the utility of money. Davidson and Suppes began thinking about the experiment in November 1953. However, neither had any previous experience in experimental investigation, and they therefore brought Sidney Siegel into the project. Siegel (1916-1961), then completing his Ph.D. in psychology at Stanford, had begun his doctoral studies in 1951, at the age of thirty-five, after having taken a rather singular biographical and educational path (Engvall Siegel 1964). In the doctoral dissertation he completed in fall 1953, Siegel presented a possible measure of authoritarianism based on experimental techniques (Siegel 1954).

[15] Davidson, McKinsey and Suppes (1955) introduced into decision analysis the so-called "money pump argument", which shows that an individual with intransitive preferences can be exploited and induced to pay money for nothing. Suppes and his doctoral student Muriel Winet (1956) put forward an axiomatization of cardinal utility based on the notion of utility differences. Davidson and Suppes (1956) developed an axiomatization of subjective EUT in which the set of alternatives over which the individual's preferences are defined is not infinite, as it is in the Ramsey-Savage framework, but finite.

DSS conducted their experiment to measure the utility of money in spring 1954; their experimental subjects were nineteen male students hired through the Stanford University Student Employment Service.[16] DSS first presented their experimental results in a Stanford Value Theory Project report published in August 1955 (Davidson, Siegel, and Suppes 1955), and two years later in the book *Decision Making: An Experimental Approach* (1957). As Ellsberg (1958, 1009) noticed in reviewing the book, it is not so much a systematic introduction to decision making as "a long article, dealing fairly technically with problems connected with this particular set of experiments."

On the cover of the book, the work is presented as co-authored by Davidson and Suppes "in collaboration with Sidney Siegel," who, meanwhile had moved to Pennsylvania State University. In practice it is difficult to disentangle the individual contributions to the experiment, and therefore in the following I will consider it as a joint product.

6. The Davidson-Suppes-Siegel experiment

The primary aim of DSS was "to develop a psychometric technique for measuring utility" in an interval scale, i.e., in a cardinal way, within the expected-utility framework (1957, 25). In particular, they were "originally inspired by the desire to see whether it was possible to improve on the Mosteller and Nogee's results" (20).

[16] Early in 1955, and without Siegel's collaboration, Davidson and Suppes performed a second experiment. This involved seven students in music at Stanford University, and aimed at measuring their utility for LP records of classical music on the basis of their choices between gambles having the records as prizes. The significance of this experiment, however, was marred by the high number of intransitive choices observed by Suppes and Davidson. The intransitives, which were probably due to the fact that most students perceived the LP records as too similar, made it tricky to identify even ordinal utility functions for the records. See Davidson and Suppes, 1956; and Davidson, Suppes, and Siegel 1957, 84-103.

As already mentioned, the general structure of the MN and DSS experiments is analogous. However, DDS modified a number of the elements of the MN design in order to make the decision tasks faced by their experimental subjects psychologically more friendly than the tasks faced by the MN subjects. However, and unlike the majority of current behavioral economists, DSS did not exploit their psychological insights to argue that the standard economic theory of decision making was flawed. Rather, they used psychological considerations to avoid confusions or biases in their subjects, thereby generating utility measures as similar as possible to those that would have been obtained if the choices had been made by EUT decision makers. To understand how DSS managed to do this, we have to enter into the technical details of their experimental design.

6.1. Design

i) *Real Money*. Like MN, DSS used real monetary payoffs. At the beginning of each experimental session, subjects received $2.00 to gamble, and the gambles' payoffs ranged from –35¢ to +50¢ (fractions of cents were not allowed). The use of real money was motivated by philosophical preoccupations very much in accord with the sort of preoccupations expressed by Friedman and Wallis in their critique of Thurstone's experiment. DSS (1957, 7) in fact followed the behaviorist approach to human agency advocated by Bertrand Russell (1921) and Ramsey ([1926] 1950), downplaying the relevance of introspection for decision analysis, and arguing that "it is with actual decision-making behavior that decision theory is concerned." Accordingly, DSS required that the decisions made in the experimental situation were "real in the most importance sense," that is, they required that the announcement of preferences between gambles made by a subjected was "followed by […] paying the subject (or collecting from him) the appropriate amount of money" (6-7).

ii) *Subjective Probabilities*. While the MN experiment was based on objective probabilities, DSS explicitly advocated a subjective approach to probability, and followed the Ramsey-Savage subjective version of EUT. Accordingly, they considered gambles of the form "x cents of dollar if event E occurs, –y cents of dollars if event E does not occur" – for brevity, [x¢, E; –y¢, not-E] – whereby each subject was supposed to assign his subjective probabilities Π(E) and Π(not-E) to the two events. As explained in section 4, however, if subjective probabilities are unknown, the expected-utility formula cannot be used to measure the utility of money.

DSS cited the articles by Preston-Baratta and Edwards showing that individuals often distort objective probabilities, and therefore were well aware that the problem could not be solved by choosing events with apparently straightforward objective probabilities, such as "heads" or "tails" in tossing a coin. Moreover, in the pilots of their experiment, DSS (1957, 51) found that their subjects often preferred gamble [x¢, heads; –x¢, tails] to gamble [–x¢, heads; x¢, tails], so showing that they considered "heads" more probable than "tails." DSS eventually solved the problem by constructing special dice carrying a nonsense syllable, such as ZOJ, on three faces, and another nonsense syllable, such as ZEJ, on the other three faces.[17] If we indicate as ZOJ the event that "the syllable ZOJ comes up when tossing a die," and as ZEJ the event that "the syllable ZEJ comes up," DSS found that their experimental subjects were indifferent between [x¢, ZOJ; –x¢, ZEJ] and [–x¢, ZOJ; x¢, ZEJ], so showing that they actually believed that the mutually exclusive events ZOJ and ZEJ were equally likely, that is, that Π(ZOJ)= Π(ZEJ)=0.5.

[17] According to Alberta Engvall Siegel (1964, 9), Siegel's second wife, it was Siegel who came up with the idea of the ZOJ-ZEJ dice: "A central problem [in the DSS experiment] was identifying an event which had subjective probability .50 for the subject, and Sid devised a zero-association nonsense-syllable die to serve as this event." I have found no evidence confirming or disconfirming this claim. In their autobiographies, Davidson (1999) and Suppes (1979) tend to downplay Siegel's role in the experiment (Davidson does not even mention Siegel), but do not discuss the paternity of the ZOJ-ZEJ dice.

iii) *Simple probabilities*. In using the ZEJ and ZOJ events, DSS not only solved the problem associated with the identification of subjective probabilities, but also avoided the psychologically tricky odds the MN subjects were presented with. If experimental subjects understand fifty-fifty gambles better than gambles with more complex odds, then the utility measures obtained from choices over ZOJ-ZEJ gambles are more reliable than the measures obtained from choices over the simple and doublets hands used by MN.

iv) *Gamble vs. Gamble.* In the MN experiment, subjects had to choose between a proper gamble and a sure amount of money corresponding to the *status quo*. DSS (1957, 23) stressed that, if gambling itself has a negative or positive utility for the subject, this approach "would produce the maximum distortion" in the utility measures.

In order to overcome this problem, in all decision situations but one DSS had their experimental subjects choose between two proper gambles of the form [x¢, ZOJ; –y¢, ZEJ]. In this way, the psychologically plausible utilities for gambling associated with the two bets should cancel out, and therefore the utility measures obtained in the gamble-vs-gamble situation should be more precise than those obtained in the gamble-vs-sure-outcome situation considered by MN.

v) *Same Type of Gambles.* DSS ran a second session of the experiment to check the utility measures obtained in the first session. While MN performed this test by using the utility functions elicited from choices involving simple hands to predict choices involving doublet hands, DSS considered choices over the same type of gambles. Specifically, DSS ran a second experimental session that took place at some period – varying from a few days to several weeks – after the first one. In this second session they re-elicited the utility curves for money of their experimental subjects from choices between [x¢, ZOJ; –y¢, ZEJ] gambles analogous to those used in the first session. Finally, DSS checked whether the utility curves elicited in the second experimental session were like those elicited in the first.

vi) *Understating Money*. In the DSS experiment, the payoffs of all gambles had the same probability, namely 0.5. Accordingly, the identification of indifferent gambles was necessarily based on modification of the monetary payoff at stake in the gambles. In this respect, the DSS experiment was similar to the MN one. In particular, DSS (7-9) ruled out fractions of cents, which are knotty from a psychological viewpoint, and considered only payoffs consisting of integer amounts of cents.

vii) *Approximate Indifference*. Like MN, DSS faced the problem of identifying an empirical correlate of the notion of "indifference." They could not follow MN and use Thurstone's probabilistic definition of indifference because their subjects did not change their minds when confronted more than once with the same pair of gambles. For DSS, this outcome was chiefly due to the fact that their gambles were simpler than those used by MN:

> Once they [the experimental subjects] chose a given option over another, they consistently held to this choice […] The primary reason for this kind of response is no doubt the relative simplicity of the offers. Mosteller and Nogee, using the more complicated game of poker dice to generate chance events, did get a distribution of responses. (1957, 41)

DSS adopted a definition of indifference according to which two gambles A and B are indifferent if A is preferred to B, but, by adding one single cent to one of B's payoffs, the new gamble B′ is preferred to A. For instance, if a subject prefers A=[15¢, ZOJ; –12¢, ZEJ] to B=[11¢, ZOJ; –9¢, ZEJ] but prefers B′=[12¢, ZOJ; –9¢, ZEJ] to A, then A and B are said to be "indifferent." This is an approximate definition of indifference in the sense that it only allows the statement that there exists some amount of money z included between 11¢ and 12¢ such that [z¢, ZOJ; –9¢, ZEJ] is indifferent to A.

DSS assigned utility values –1 and +1 to amounts of money a=–4¢ and b=6¢. Then, for each subject, they looked for four monetary amounts c, d, f, g such that u(c)=–3, u(d)=+3, u(f)=–5 and u(g)=+5. For the reason explained above, however, DSS could determine c, d, f and g only by approximation, and therefore were only able to elicit the bounds of the subjects' utility curves rather than the utility curves themselves.[18]

6.2. Findings

In the first part of their experiment, DSS managed to elicit the bounds of the utility curves of fifteen of their nineteen experimental subjects. The subjects' utility curves were presented in graphs like the one reproduced in Figure 2 below, which refers to Subject 1.

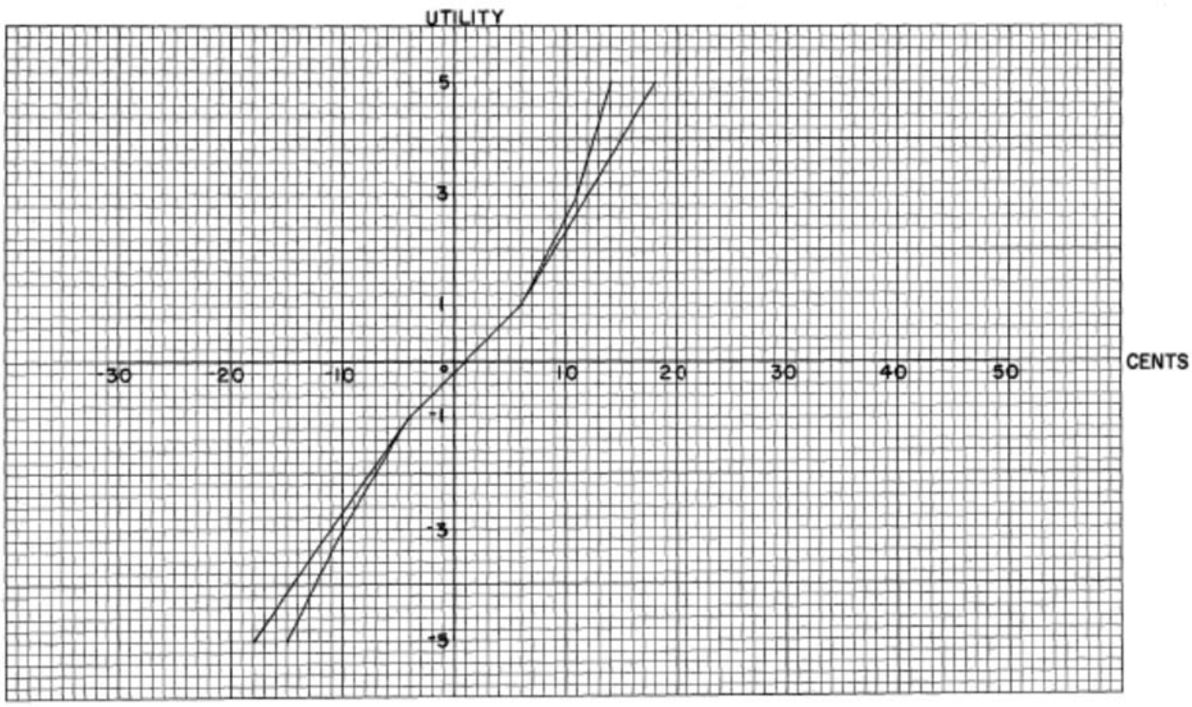

Figure 2. Bounds for Subject 1's utility curve of money. The two continuous lines are the bounds within which the "true" utility curve lies. The bounds are drawn for the utility values –5, –3, –1, +1, +3, +5, which are connected by straight lines. *Source*: Davidson, Siegel and Suppes 1957, 63.

[18] For instance, DDS found that, for Subject 1, the sum c corresponding to utility level –3 lies between –11¢ and –10¢; the sum d corresponding to utility level +3 lies between 11¢ and 12¢; the sum f corresponding to utility level –5 lies between –18¢ and –15¢; and the sum g corresponding to utility level +5 lies between 14¢ and 18¢.

Regarding the fifteen subjects for whom it was possible to elicit utility curves, the main experimental findings obtained by DSS (1957, 62-72) can be summarized as follows:

i) The utility curves of ten subjects displayed a trend similar to the curve of Subject 1's, i.e., convex for wins (love for risk) and concave for losses (risk aversion);

ii) The utility curves of two subjects displayed the opposite trend: they were concave for wins and convex for losses;

iii) The utility curves of the remaining three subjects were fundamentally linear through their length, suggesting that they were risk neutral.

DSS compared these findings with those obtained by MN from choices over simple hands and argued that, despite the differences between the two experiments, the degree of similarity between their results was "fairly striking" (75).

In the second session of the experiment, DSS re-elicited the utility curves of ten of the original fifteen experimental subjects, and found that, for nine of the ten subjects, the utility curves elicited in the two sessions were in fact very similar.[19] Thus, like MN, DSS concluded their experimental study in an optimistic way, i.e., by arguing that measuring utility experimentally appears feasible:

> The chief experimental result may be interpreted as showing that for some individuals and under appropriate circumstances it is possible to measure utility in an interval scale [i.e., in a cardinal way] (19)

This result, in turn, supported the thesis that "an individual makes choices among alternatives involving risk as if he were trying to maximize expected utility" (26).

---

[19] For DSS (1957, 69) the behavior of the remaining subject was explained by the fact that he was a foreign student with some language difficulty.

6.3. Summing Up

In designing their experiment, DSS took into account a number of psychological phenomena that MN had neglected, such as the fact that actual individuals may distort objective probabilities or get confused when facing gambles with complicated odds. If considered from the viewpoint of EUT, these phenomena may be seen as "disturbing causes" that jeopardize the validity of the theory and spoil the significance of the experimental measurements of utility. The very fact of discussing these psychological phenomena in a book addressed not only to psychologists but also, and primarily, to economists was certainly pioneering.

However, DSS took those aspects of the psychology of decision making into account mostly in order to neutralize them. One may say that they took into consideration the psychology of decision only to have their experimental subjects behave like brave EUT decision makers. It is not therefore very surprising that DSS concluded that utility is measurable and EUT validated.

7. Influence in Economics

In the economic literature of the decade 1955-1965, the MN and DSS experiments were cited in connection with issues concerning the measurability of utility and the probabilistic theory of choice. However, their overall impact on the economic analysis of the period was limited.

The most comprehensive discussion of the significance of the utility measurements carried out by MN and DSS can be found in Duncan Luce and Howard Raiffa's *Games and Decisions*, a book that after its publication in 1957 quickly became a key reference point for economists working in decision analysis. Luce and Raiffa (1957, 35) summarized the methods used by MN and DSS to measure utility, praised the DSS experiment as "the most elegant in the area," but pointed out that the MN and DSS utility measures do not appear replicable or applicable outside the laboratory. Despite this limitation, which today we would call an "external validity problem," Luce and Raiffa argued that laboratory attempts to measure utility are worth undertaking in order "to see if under any conditions, however limited, the postulates of the model can be confirmed" (37).

Other commentators, such as Ellsberg (1958) and Arrow (1958), made a similar external-validity point: the utility measurements carried out by MN and DSS appeared hardly extendable to situations different to the very specific ones designed by MN or DSS, e.g., to situations involving larger amounts of money.

Suppes was well aware of this issue. In a paper presented at a conference on game theory held in Princeton in 1961, he acknowledged that experimental studies of utility measurement were as "yet entirely too fragile in relation to the massive claims sometimes made for utility theory" (1962, 62). In particular, he stressed the necessity of measuring utility in situations in which individuals "are making weighty and significant decisions" (62). Suppes even predicted that operations-research people in government and industry were on the verge of undertaking these kind of studies.

This prediction, however, was not fulfilled. In the 1960s economists lost interest in the experimental measurement of utility. Apparently, in that decade only Marschak, in collaboration with psychologist Gordon M. Becker and statistician Morris H. DeGroot, made a further attempt to measure experimentally the utility of money (Becker, DeGroot, and Marschak 1964). However, at that point of his career Marschak was far from being a typical economist (Herfeld, this volume). Since the mid-1950s his research interests had moved away from traditional mathematical economics towards psychology and, more generally, the behavioral sciences. Indeed, the Becker-DeGroot-Marschak article was not published in an economics journal but in the interdisciplinary review *Behavioral Science*.

Marschak and associates presented their experiment as an evolution of the MN experiment and did not cite DSS. In particular, like MN and differently from DSS, Becker, DeGroot and

Marschak assumed that the two Yale students who were their experimental subjects understood well, and did not distort, objective probabilities. Marschak and associates asked the students to state their minimum selling prices for wagers the payoffs of which ranged from 0¢ to +100¢ and, based on EUT, used the selling prices to elicit the students' utility curves for money. The authors concluded that, as the students became more familiar with the experimental task, they behaved more consistently with the EUT model.

More than in the Becker-DeGroot-Marschak study, the psychological insights embodied in the MN and DSS experiments found application in the probabilistic theory of choice, which was advanced in the mid-1950s by Marschak (again) and other economists. The theory is based on a Thurstonian definition of preference – a subject prefers alternative A to alternative B if the probability that he chooses A over B is at least one-half – and its proponents often referred to MN's definition of indifference as a pioneering use of the probabilistic approach in economics (Marschak 1955, Quandt 1956, Debreu 1958, Luce 1959). In association with Davidson (Davidson and Marschak 1959), as well as with Becker and DeGroot (Becker, DeGroot, and Marschak 1963), Marschak also ran experiments to test the probabilistic theory of choice. In order to circumvent subjective distortions of objective probabilities, he and his coauthors used the ZOJ-ZEJ dice introduced by DSS. However, by the mid-1960s the probabilistic approach to choice was substantially abandoned in economics, possibly because its implications for demand and equilibrium analysis were unclear.

With the vanishing of the economists' interest in both the probabilistic theory of choice and the experimental measurement of utility, the psychological phenomena discussed by MN and DSS disappeared from the purview of the mainstream economic theory of decision. Integrating them into decision theory would clearly have been a difficult task, and since the MN, DSS and even the Becker-DeGroot-Marschak experiments suggested that EUT was an acceptable predictor of choice under risk, the difficult task could be set aside as non-urgent.

It was only in the 1970s, when robust experimental evidence against EUT accumulated and theories alternative to EUT began to be advanced, that the psychology of decision re-gained importance for economists. Thus it does not appear merely accidental that, in the celebrated article in which Daniel Kahneman and Amos Tversky (1979, 276) put forward their Prospect Theory, they cited the MN and DSS experiments.[20] For Kahneman, Tversky, and other non-EUT theorists, however, psychological phenomena such as those discussed by MN and DSS ceased to be disturbing causes to be removed, and become fundamental causes to be investigated.

8. Summary and Conclusions

This paper has explored a part of the interdisciplinary interaction between economics and psychology in the 1940s and 1950s by studying the origin, content and impact of two experiments to measure the utility of money. Both experiments were performed by psychologists and other non-economists, and their design contained a number of elements that responded to psychological insights rather than the economic image of human agency associated with EUT. Between 1955 and 1965, these psychological insights found some application in the short-lived theory of probabilistic choice, but were quickly forgotten afterwards.

The story told in this paper, therefore, can be framed as an actual encounter between economics and psychology or, more specifically, between the mainstream economic theory of decision under risk represented by EUT and the experimental tradition in psychology associated with psychometrics. In this meeting the psychologists not only provided the technology to perform the experiments but also identified certain aspects of the psychology of

[20] On Prospect Theory, see the papers by Heukelom and Staddon in this volume.

decision that could spoil the significance of the experiments and suggested how to deal with those aspects.

This meeting, however, remained at a superficial level and did not initiate any substantial modification in the body of economic analysis. The main reason for this outcome appears to have been the difficulty of incorporating the psychological insights contained in the two experiments into EUT without making the latter knotty and unmanageable. Only in the 1970s was economics compelled to re-consider its relationship with psychology and the psychological phenomena discussed by Mosteller, Nogee, Davidson, Suppes and Siegel.

References

Alchian, A. 1953. The Meaning of Utility Measurement. *American Economic Review* 43:26-50.

Allais, M. 1953. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica* 21:503-546.

Arrow, K.J. 1951. Alternative Approaches to the Theory of Choice in Risk-Taking Situations. *Econometrica* 19:404-437.

———. 1958. Utilities, Attitudes, Choices: A Review Note. *Econometrica* 26:1-23.

Becker, G.M., M.H. DeGroot, and J. Marschak. 1963. An Experimental Study of Some Stochastic Models for Wagers. *Behavioral Science* 8:199-202.

———. 1964. Measuring Utility by a Single-Response Sequential Method. *Behavioral Science* 9:226-232.

Blackwell, D., and M.A. Girshick. 1954. *Theory of Games and Statistical Decisions*. New York: Wiley.

Bruner, J.S., L. Postman, and F. Mosteller. A Note on the Measurement of Reversals of Perspective. *Psychometrika* 15:63-72.

Davidson, D. 1999. Intellectual Autobiography. In *The Philosophy of Donald Davidson*, edited by L.E. Hahn, 3-70. Chicago: Open Court.

Davidson, D., and J. Marschak. 1959. Experimental Tests of a Stochastic Decision Theory. In *Measurement: Definitions and Theories*, edited by C. W. Churchman and P. Ratoosh, 233-69. New York: Wiley.

Davidson, D., J.C.C. McKinsey, and P. Suppes. 1955. Outlines of a Formal Theory of Value, I. *Philosophy of Science* 22:140-160.

Davidson, D., S. Siegel, and P. Suppes. 1955. Some Experiments and Related Theory on the Measurement of Utility and Subjective Probability. Technical Report No. 1, August 15, 1955. Stanford University.

Davidson, D., and P. Suppes. 1956. A Finitistic Axiomatization of Subjective Probability and Utility. *Econometrica* 24:264-275.

Davidson, D., P. Suppes, and S. Siegel. 1957. *Decision Making: An Experimental Approach*. Stanford: Stanford University Press.

Debreu, G. 1958. Stochastic Choice and Cardinal Utility. *Econometrica* 26:440-444.

Edwards, W. 1953. Probability-Preferences in Gambling. *American Journal of Psychology* 66:349-364.

———. 1954a. Probability-Preferences among Bets with Differing Expected Values. *American Journal of Psychology* 67:56-67.

———. 1954b. The Reliability of Probability-Preferences. *American Journal of Psychology* 67:68-95.

———. 1961. Behavioral Decision Theory. *Annual Review of Psychology* 12: 473-498.

Ellsberg, D. 1954. Classic and Current Notions of "Measurable Utility." *Economic Journal* 64:528-556.

———. 1958. Review of *Decision Making* by D. Davidson, P. Suppes, S. Siegel. *American Economic Review* 48:1009-1011.

Engvall Siegel, A. 1964. Sidney Siegel: A memoir. In *Decision and Choice. Contributions of Sidney Siegel*, edited by S. Messick and A.H. Brayfield, 1-24. New York: McGraw-Hill.

Erikson, P. and others. 2013. *How Reason Almost Lost Its Mind*. Chicago: University of Chicago Press.

Fishburn, P.C. 1989. Retrospective on the Utility Theory of von Neumann and Morgenstern. *Journal of Risk and Uncertainty* 2: 127-158.

Fishburn, P.C. and P. Wakker. 1995. The Invention of the Independence Condition for Preferences. *Management Science* 41:1130-1144.

Freeman, H.A., M. Friedman, F. Mosteller, and W. A.Wallis, eds. 1948. *Sampling Inspection.* New York: McGraw-Hill.

Fréchette, G.R, and A. Schotter. 2015. *Handbook of Experimental Economic Methodology*. New York: Oxford University Press.

Friedman, M. Papers. Hoover Institution Archives. Hoover Institution, Stanford, California.

Friedman, M., and L.J. Savage. 1947. Planning Experiments Seeking Maxima. In *Techniques of Statistical Analysis*, edited by C. Eisenhart, M.W. Hastay, and W.A. Wallis, 363-372. New York: McGraw-Hill.

———. 1948. The Utility Analysis of Choices Involving Risk. *Journal of Political Economy* 56:279-304.

———. 1952. The Expected-Utility Hypothesis and the Measurability of Utility. *Journal of Political Economy* 60:463-474.

Giocoli, N. 2003. *Modeling Rational Agents*. Cheltenham: Elgar.

Girshick, M.A., F. Mosteller, and L.J. Savage. 1946. Unbiased Estimates for Certain Binomial Sampling Problems with Applications. *Annals of Mathematical Statistics* 17:13-23.

Griffith, R.M. 1949. Odds Adjustments by American Horse-Race Bettors. *American Journal of Psychology* 62:290-294.

Guala, F. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Heukelom, F. 2014. *Behavioral Economics. A History*. New York: Cambridge University Press.

Isaac, J. 2013. Donald Davidson and the Analytic Revolution in American Philosophy, 1940-1970. *Historical Journal* 56:757-779.

Jallais, S., and P.-C. Pradier. 2005. The Allais Paradox and Its Immediate Consequences for Expected Utility Theory. In *The Experiment in the History of Economics*, edited by P. Fontaine and R. Leonard, 25-49. New York: Routledge.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47:263-292.

Keats, A.S., H.K. Beecher, and F. Mosteller. 1950. Measurement of Pathological Pain in Distinction to Experimental Pain. *Journal of Applied Physiology* 3:35-44.

Lepore E. 2004. Interview with Donald Davidson. In D. Davidson, *Problems of Rationality*, 231-266. Oxford: Oxford University Press.

Luce, R.D. 1959. *Individual Choice Behavior*. New York: Wiley.

Luce, R.D., and H. Raiffa. 1957. *Games and Decisions*. New York: Wiley.

Luce, R.D., and P. Suppes. 1965. Preference, Utility, and Subjective Probability. In *Handbook of Mathematical Psychology*, volume 3, edited by R. D. Luce, R. R. Bush, and E. Galanter, 249-410. New York: Wiley.

Manne, A.S. 1952. The Strong Independence Assumption – Gasoline Blends and Probability Mixtures. *Econometrica* 20:665-668.

Markowitz, H. 1952. The Utility of Wealth. *Journal of Political Economy* 60:151-158.

Marschak, J. 1950. Rational Behavior, Uncertain Prospects, and Measurable Utility. *Econometrica* 18:111-141.

———. 1951. Why Should Statisticians and Businessmen Maximize Moral Expectation? In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, 493-506. Berkeley: University of California Press.

———. 1954. Probability in the Social Sciences. In *Mathematical Thinking in the Social Sciences*, edited by P.F. Lazarsfeld, 166-215. Glencoe: Free Press.

———. 1955. Norms and Habits of Decision Making Under Certainty. In *Mathematical Models of Human Behavior*, 45-53. Stamford, Conn.: Dunlap and Associates.

McKinsey, J.C.C. 1952. *Introduction to the Theory of Games*. New York: McGraw-Hill.

Michell, J. 1999. *Measurement in Psychology*. Cambridge: Cambridge University Press.

Mongin P. 2009. Duhemian Themes in Expected Utility Theory. In A. Brenner and J. Gayon (eds.), *French Studies in the Philosophy of Science*, 303-357. Springer: New York.

———. 2014. Le paradoxe d'Allais: comment lui rendre sa signification perdue. *Revue économique* 65:743-779.

Moscati, I. 2007. Early Experiments in Consumer Demand Theory: 1930-1970. *History of Political Economy* 39:359-401.

———. 2013. Were Jevons, Menger and Walras Really Cardinalists? On the Notion of Measurement in Utility Theory, Psychology, Mathematics and Other Disciplines, 1870-1910. *History of Political Economy* 45: 373-414.

———. 2015a. Austrian Debates on Utility Measurement, from Menger to Hayek. In *Hayek: A Collaborative Biography, Part IV*, edited R. Leeson, 137-179. New York: Palgrave Macmillan.

———. 2015b. How Samuelson Was Converted to Expected Utility Theory. Paper presented at the meeting of the History of Economics Society, Michigan State University, June 2015.

Mosteller, F. 2010. *The Pleasures of Statistics*. New York: Springer.

Mosteller, F., and P. Nogee. 1951. An Experimental Measurement of Utility. *Journal of Political Economy* 59:371-404.

Nasar, S. 1998. *A Beautiful Mind*. New York: Simon and Schuster.

Neumann, J. von, and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Pooley J., and M. Solovey. 2010. Marginal to the Revolution: The Curious Relationship between Economics and the Behavioral Sciences Movement in Mid-Twentieth-Century America. *History of Political Economy* 42 (annual supplement):199-233.

Preston, M.G., and P. Baratta. 1948. An Experimental Study of the Auction-Value of an Uncertain Outcome. *American Journal of Psychology* 61:183-193.

Quandt, R.E. 1956. A Probabilistic Theory of Consumer Behavior. *Quarterly Journal of Economics* 70:507-536.

Ramsey, F.P. (1926) 1950. Truth and Probability. In *Foundations of Mathematics and Other Logical Essays*, edited by R.B. Braithwaite, 156-198. New York: Humanities Press.

Russell, B. 1921. *The Analysis of Mind*. London: Allen and Unwin.

Samuelson, P.A. 1950. Probability and the Attempts to Measure Utility. *Economic Review* (*Keizai Kenkyu*) 1:167-173.

———. 1952. Probability, Utility, and the Independence Axiom. *Econometrica* 20:670-678.

Savage, L.J. 1954. *The Foundations of Statistics*. New York: Dover.

Shanteau, J., B.A. Mellers, and D.A. Schum, eds. 1999. *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Boston: Kluwer.

Siegel, S. 1954. Certain Determinants and Correlates of Authoritarianism. *Genetic Psychology Monographs*, 49:187-254.

Strotz, R.H. 1953. Cardinal Utility. *American Economic Review*, *Papers and Proceedings*. 43, 384-397.

Suppes, P. 1951. A Set of Independent Axioms for Extensive Quantities. *Portugaliae Mathematica* 10:163-172.

———. 1962. Recent Developments in Utility Theory. In *Recent Advances in Game Theory*, edited by O. Morgenstern and A.W. Tucker, 61-72. Princeton: Princeton University Press.

———. 1979. Self-profile. In *Patrick Suppes*, edited by R.J. Bogdan, 3-56. Dordrecht: Reidel.

Suppes, P. and M. Winet. 1955. An Axiomatization of Utility Based on the Notion of Utility Differences. *Management Science* 1:259-270.

Thurstone, L.L. 1927. A Law of Comparative Judgment. *Psychological Review* 34:273-286.

———. 1931. The Indifference Function. *Journal of Social Psychology* 2:139-167.

Wallis, W.A. 1980. The Statistical Research Group, 1942-1945. *Journal of the American Statistical Association* 75:320-330.

Wallis, W.A., and M. Friedman. 1942. The Empirical Derivation of Indifference Functions. In *Studies in Mathematical Economics and Econometrics*, edited by O. Lange, F. McIntyre, and T.O. Yntema, 175-189. Chicago: University of Chicago Press.