Reconciling Introspective Utility with Revealed Preference: Experimental Arguments Based on Prospect Theory*

Mohammed Abdellaoui[a], Carolina Barrios[a],

Peter P. Wakker[b,**]

[a] *GRID, CNRS, ENSAM & ESTP, 30 Avenue du Président Wilson 94230 Cachan, France*

[b] *Dept. of Economics, University of Amsterdam, Roetersstraat 11, Amsterdam, 1018 WB, The Netherlands, 31-(0)20.525.42.29 (O), 31-(0)20.525.41.26 (S), 31 (0)20 525.52.83 (F), P.P.Wakker@UvA.NL*

April, 2005

Abstract

In an experiment, choice-based utility of money is derived from choices under risk, and choiceless utility from introspective strength-of-preference judgments. The well-known inconsistencies of risky utility that result if the data are analyzed in terms of expected utility are resolved if the data are analyzed in terms of prospect theory. One consistent cardinal utility index for risky choice then results. Remarkably, this cardinal index also agrees well with the choiceless utilities. This finding suggests a relation between a choice-based and a choiceless concept. Such a relation would imply that introspective judgments can provide useful data for economics, and can reinforce the revealed-preference paradigm. Implications for the classical debate on ordinal versus cardinal utility are discussed.

---

* Mark Blaug, Denis Bouyssou, Itzhak Gilboa, Jean-Yves Jaffray, Edi Karni, Veronika Köbberling, Mark Machina, Chris Starmer, and Stefan Trautmann made helpful comments.

** Corresponding author: Peter P. Wakker, CREED, Dept. of Economics, University of Amsterdam, Roetersstraat 11, Amsterdam, 1018 WB, The Netherlands; 31-(0)20.525.42.29 (O), 31-(0)20.525.41.26 (S), 31-(0)20 525.52.83 (F); P.P.Wakker@uva.nl; http://www1.fee.uva.nl/creed/wakker.

1. Introduction

Utility has been a controversial concept throughout the history of economics, with interpretations shifting over time. Since the beginning of the twentieth century, after what has become known as the ordinal revolution, utility has been taken as an ordinal concept, based solely on observable choice, in mainstream economics (Pareto 1906). Ordinalism has dominated economics ever since (Hicks & Allen 1934).

Based on the many anomalies of observed choice that have been discovered in the twenthieth century, several authors have argued that a reinterpretation of utility broader than purely ordinal is relevant for mainstream economics. One of the earliest proponents was van Praag (1968), who used subjective questions to measure welfare. Recently, Kahneman (1994) initiated a stream of papers arguing for the relevance of experienced utility in economics. Such a broader reinterpretation was also advocated by a founder of the Econometric Institute of the Erasmus University, Jan Tinbergen (1991), who wrote in a special issue of the Journal of Econometrics on the measurement of utility and welfare:

> The author believes in the measurability of welfare (also called satisfaction or utility). Measurements have been made in the United States (D.W. Jorgenson and collaborators), France (Maurice Allais), and The Netherlands (Bernard M.S. Van Praag and collaborators). The Israeli sociologists S. Levy and L. Guttman have shown that numerous noneconomic variables are among the determinants of welfare ... (p. 7).

This paper presents an investigation into broader interpretations of the utility of money, using an experimental approach. We will compare experimental measurements of choice-based and choiceless utilities, and investigate their relations. Our main finding will be that there are no systematic differences between the different measurements. This finding suggests that choiceless empirical inputs can be useful for the study and prediction of observable choice. Let us emphasize that we make this suggestion only for choiceless empirical inputs that can be firmly related to observable choice. These choiceless inputs should reinforce, rather than renounce, the achievements of the ordinal revolution.

Expected utility provides a firm basis for rational decisions and for Bayesian statistics (Kahneman & Tversky 1979, p. 277; Savage 1954; Zellner 1971). It is also used as a basis for most descriptive economic measurements of utility today, in which risk attitudes are to be captured entirely in terms of utility curvature. This approach is so widespread that it has been ingrained in standard economic terminology, with utility curvature usually described as "risk aversion" or even, in econometric studies, as "individual preference." Many empirical studies have, however, revealed descriptive difficulties of expected utility (Starmer 2000). Descriptive improvements have been developed, such as prospect theory (Kahneman & Tversky 1979, Tversky & Kahneman 1992). Our analysis will first show, in agreement with previous findings (Herschey & Schoemaker 1985), that utility measurement under expected utility leads to inconsistencies, which may explain why there haven't been many estimations of utility yet (Gregory, Lamarche, & Smith 2002, p. 227). We next show that, by means of prospect theory, the inconsistencies can be resolved, and a consistent economic concept of utility can be restored.

Outline of the Paper

Section 2 briefly describes the history of utility in economics up to 1950, focusing on the rise of ordinalism and ending with von Neumann and Morgenstern's (1944) contribution. This history was described before by Stigler (1950), Blaug (1962), and others. Because of new developments in utility theory during the last decades, an update of the history is called for. It is provided in Section 3. Two developments are distinguished. One took place in mainstream economics, where many empirical problems of revealed preference were discovered, leading Kahneman and others to propose new interpretations of utility (Subsection 3.1). The other development took place in decision theory and concerns the distinction between risky and riskless cardinal utility (Subsection 3.2).[1] These developments will lead to the research question of this paper.

Section 4 gives notation and defines expected utility and prospect theory. Section 5 measures choice-based utilities through a recently introduced method, the tradeoff method, which is valid under expected utility but, contrary to classical methods, maintains its validity under prospect theory. Subsequently, choiceless cardinal utility is measured without using any choice making or risk. Remarkably, no significant differences are found between these two measurements of utility. A psychological explanation is given for the plausibility of the equality found. To verify that tradeoff

[1] We use "risky utility" as a shorthand for utility to be used for choices under risk, such as in expected utility.

utilities do reflect choice making, Section 6 compares those utilities with utilities derived from a third, traditional, measurement method, that is also based on choice making, and that uses certainty equivalents of two-outcome prospects with a 1/3 probability for the best outcome. Again, no significant differences are found.

To verify that our design has the statistical power to detect differences, Section 6 also compares the utilities obtained up to that point with utilities derived from a fourth measurement method, again choice-based and again using certainty equivalents, but now of two-outcome prospects with a 2/3 probability for the best outcome. When analyzed through expected utility, the utilities of the fourth method deviate significantly from those found through the other three methods, in agreement with the common findings in the literature (Karmarkar 1978), and falsifying expected utility. The discrepancy is resolved by reanalyzing the data by means of prospect theory. This theory does not affect the first three measurements but it modifies the fourth. After this modification, a complete reconciliation of all measurements obtains, leading to one utility function consistently measured in four different ways.

Section 7 acknowledges and discusses some criticisms that can be raised against our analysis, and compares our findings with other findings in the literature. Motivations and conclusions are in Section 8. Appendix A gives the details of our experimental method for eliciting indifferences, developed to minimize biases. Appendices B and D describe further statistical tests.

Appendix C describes parametric families of utility used in our study. We use two traditional families but also introduce a new one-parameter family, the expopower family, constructed from a more general two-parameter family of Saha (1993). Our family, contrary to existing families, allows for the simultaneous fulfillment of three economic desiderata: concave utility, decreasing absolute risk aversion, and increasing relative risk aversion. There is much interest in such new parametric families of utility. We nevertheless present this material in the appendix because it is more technical than the rest of this paper.

In summary, by using prospect theory and the techniques of modern experimental economics, our paper sheds new light on the measurement, interpretation, and applicability of utility.

2. The History of Ordinal versus Cardinal Utility up to 1950

The first appearances of utility were in Cramer (1728) and Bernoulli (1738), who proposed expected utility as a solution to the St. Petersburg paradox. Utility was presented as a general index of goodness and the authors did not explicitly restrict its meaning to risky decisions. Bentham (1789) gave the first thorough discussion of utility as a central concept in human behavior. Risk was not central in his analysis, although it was mentioned occasionally. In the century that followed, economists used utility as an, in modern terms cardinal, index of goodness. Although there were concerns about the measurement of utility (Cooter and Rappoport 1984), measurability was not a central issue. After the marginal revolution of the 1870s, which showed the importance of comparisons of utility rather than absolute levels of utility, diminishing marginal utility became the central hypothesis. Marshall (1890) pointed out its equivalence to risk aversion, assuming that the expectation of the utility in question governs risky decisions. Table 1 displays the various concepts of utility, discussed hereafter.

	Choice-based	Choiceless
ordinal utility	− Consumer theory	
cardinal utility	− Intertemporal − Welfare − Risk	− Strength of preferences − Experienced (Kahneman)

TABLE 1. Various concepts of utility. The utilities within boxes are commonly required to be restricted to their domains, and not to be applied in other domains. ⇔: A relation between these two is obtained in this paper. It extends vNM (von Neumann-Morgenstern) risky utility beyond risk, and connects an economic, middle-column, concept with a "non-economic," right-column concept.

An important step forward was made at the beginning of the twentieth century, when the views of utility changed profoundly due to the ordinal revolution. Economists became concerned about the empirical observability of utility. Utility was related to observable choice and all associations with introspective psychological judgments were abandoned. This development changed the status of utility from being ad hoc to being empirically well founded. Along with the concern for observability came the understanding of Pareto and others that, if the only purpose of utility is to explain consumer choices, prices, and equilibria, as in the middle cell of Table 1, then utility is ordinal. Any strictly increasing transformation can be applied without affecting the empirical meaning, which implies that utility differences and marginal utility are not meaningful.

Alt (1936), Frisch (1926), and others demonstrated that cardinal utility, which does assign meaning to utility differences, can be formally derived from direct strength-of-preference judgments, such as the judgment that the strength of preference of \$10 over \$0 exceeds that of \$110 over \$100. Such judgments are based on introspection and not on observable choice and are, therefore, considered meaningless by most economists (Samuelson 1938a; Varian 1993 pp. 57–58). Hicks and Allen (1934) strongly argued in favor of an ordinal view of utility, and this became the dominant viewpoint in economics. Similar ideas, in agreement with logical positivism, became popular in psychology, where behaviorism was propagated by Watson (1913), Skinner (1971), and others.

New hope for the existence of cardinal utility was raised by von Neumann and Morgenstern (1944), who derived cardinal utility for decision under risk; earlier presentations of this idea were given by Ramsey (1931) and Zeuthen (1937). After some debates, the consensus became that this risky index is cardinal in the mathematical sense of being unique up to unit and origin, but not cardinal in the sense of being the neo-classical index of goodness that emerged at the end of the 19th century (Friedman and Savage 1948; Baumol 1958 p. 655; Varian 1993).[2] Ordinalism has continued to dominate in mainstream economics ever since.

3. Ordinal versus Cardinal Utility after 1950

This section describes the history of utility in the second half of the twentieth century, which followed after the classic historical review by Stigler (1950) and after von Neumann and Morgenstern's contribution.

[2] For recent deviating viewpoints, see Harsanyi (1978), Loomes and Sugden (1982), Ng (1997), and Rabin (2000, footnote 3). It is remarkable that von Neumann and Morgenstern used their cardinal utility not only to evaluate randomized strategies but also as a unit of exchange between players.

3.1. Ordinal Utility in the Economics Literature after 1950

At the beginning of the ordinal period, promising results were obtained through preference representations and derivations of equilibria (Houthakker 1950; Samuelson 1938b; Savage 1954; Debreu 1959). Soon, however, problems arose (Allais 1953; Ellsberg 1961; Ng 1997 p. 1854; Sen 1974 p. 390; Simon 1955). Cardinal utilities, at least in a mathematical sense, could not be discarded entirely. They were needed, not only for risky decisions such as for mixed strategies in game theory (von Neumann & Morgenstern 1944), but also for intertemporal evaluations (Samuelson 1937), for utilitarian welfare evaluations (Harsanyi 1955), for quality-of-life measurements in health (Gold et al. 1996), and for (−1 times the) loss functions in Bayesian statistics (Zellner 1971). The consensus became that such cardinal indexes are relevant, but should be restricted to the specific domain where they apply, and should not be equated to each other or to neo-classical cardinal utility (Samuelson 1937 p. 160).

The most serious blow for the revealed-preference paradigm may have been the discovery of preference reversals, entailing that revealed preferences can depend on economically irrelevant framing aspects even in the simplest choice situations (Grether and Plott 1979; Lichtenstein and Slovic 1971; Camerer 1995). Subsequently, numerous other choice anomalies have been discovered (Kahneman and Tversky 2000). It led Kahneman (1994) to argue that choiceless, "experienced," utility can provide useful information for economics in contexts where such choice anomalies prevail. Many other papers have argued for broader interpretations of utility than purely ordinal, e.g. Broome (1991), Frey and Stutzer (2000), Gilboa and Schmeidler (2001), Kapteyn (1994), Loomes and Sugden (1982), Rabin (2000 footnote 3), Robson (2001 Section III.D), Tinbergen (1991), van Praag (1968, 1991), and Weber (1994 p. 239). A drawback of extending the inputs of utility is, obviously, that predictions of economic decisions then can become difficult. The present paper presents an experimental investigation, based on prospect theory, into broader interpretations of utility, showing that they can positively contribute to economic predictions, rather than complicate them.

3.2. Cardinal Utility in Decision Theory after 1950; Risky versus Riskless Utility

Since the 1970s, several authors in decision theory have conducted empirical studies into the distinction between von Neumann-Morgenstern ("risky") and neo-classical cardinal utility. Contrary to the ordinalists, these authors assumed that choiceless cardinal utility, and thereby marginal utility, is meaningful, and they commonly used strength-of-preference judgments to measure it. As depicted in Table 1, choiceless cardinal utility can also be related to direct experience (Kahneman 1994). Others have related it to just noticeable differences (Allais 1953; Edgeworth 1881), and other psychophysical measurements (Breiter et al. 2001). In this study, we restrict attention to strength of preferences for measuring choiceless utility. In decision theory, such cardinal choiceless utility was usually called riskless utility. The difference between marginal riskless utility and risk attitude has often been emphasized (Camerer 1995 p. 619; Ellingsen 1994; Ellsberg 1954; Samuelson 1950 p. 121), and nonlinear empirical relations between risky and riskless utility have been studied (Bouyssou and Vansnick 1988; Debreu 1976; Pennings and Smidts 2000).

The classical decision-theoretic studies invariably assumed expected utility for analyzing risky decisions. Under this assumption, a difference between marginal utility and risk attitude necessarily implies that the corresponding utility functions must be in different cardinal classes, that is, there must be a nonlinear relation between risky and riskless utility. The main problem in this classical approach may have been the empirical deficiency of expected utility (Camerer 1995). Different methods for measuring risky utility, that should yield the same utilities, exhibited systematic discrepancies (Karmarkar 1978; Hershey and Schoemaker 1985). These were as pronounced as the differences between risky and riskless utility (McCord and de Neufville 1983, p. 295). It led some authors working on risky versus riskless utility to abandon the classical expected-utility approach. For example, Krzysztofowicz and Koch (1989) and McCord and de Neufville (1984) suggested that nonexpected utility theories will better accommodate the discrepancies between marginal utility and risk attitude than nonlinear transformations between risky and riskless utility.

Since the 1980s, many models that deviate from expected utility have been proposed (Camerer 1995; Machina 1982, Starmer 2000). Popular examples are rank-dependent utility (Gilboa 1987; Quiggin 1982; Schmeidler 1989; Yaari 1987) and prospect theory (Tversky and Kahneman 1992). Rank-dependent utility and prospect theory agree on the domain considered in this paper, i.e. two-outcome prospects with known probabilities. These theories assume nonadditive probability weighting. They provide better empirical predictions than expected utility and explain the discrepancies between different utility measurements.

Several authors have suggested that utility measurement can be improved through prospect theory (Bayoumi and Redelmeier 2000; Bleichrodt, Pinto, and Wakker 2001; Krzysztofowicz and Koch 1989). Before, Fellner (1961 p. 676) suggested the same basic idea. Under prospect theory, aspects of risk attitude not captured by marginal utility can be explained by probability weighting, so that the main reason to distinguish between risky and riskless utility disappears. The experimental findings of this paper will, indeed, find no systematic difference between risky and riskless utility if the data are analyzed in terms of prospect theory.

4. Expected Utility and Prospect Theory

Throughout this paper, U: $\mathbb{R} \to \mathbb{R}$ denotes a utility function of money that is strictly increasing. We examin situations in which U is measurable or cardinal in a mathematical sense, i.e. U is determined up to unit and origin. The same symbol U will be used for utilities measured through strength of preferences as for utilities measured through risky choices under various theories, even though a priori these utilities may be different. The meaning of U will be clear from the context. The different interpretations of U for strength of preference, expected utility, rank-dependent utility, and prospect theory (where the term value function is often used) will be discussed in Section 6.

By (p,x; y) we denote a monetary prospect yielding outcome x with probability p and outcome y otherwise. *Expected utility* (*EU*) assumes that a utility function U exists such that the prospect is evaluated by pU(x) + (1−p)U(y). It is well known that U is cardinal in the mathematical sense of being unique up to unit and origin.[3] Prospect theory assumes that probabilities are weighted nonlinearly, by the *probability weighting function*, denoted w. The *prospect theory* (*PT*) value of a prospect (p,x; y) is w(p)U(x) + $(1-w(p))$U(y), where it is assumed that $x \geq y \geq 0$. EU is the special case where w is the identity. For the prospects considered in this paper, that only yield gain outcomes, original prospect theory (Kahneman & Tversky, 1979, Eq. 2), rank-dependent utility (Quiggin, 1982), and their combination, cumulative

---

[3] It need not be cardinal in the sense of being the neo-classical index of goodness that emerged at the end of the 19th century (Baumol 1958 p. 655).

prospect theory (Tversky & Kahneman, 1992), agree. Gul's (1991) disappointment theory also agrees with these theories on our domain of two-outcome prospects, and, therefore, our conclusions hold under this theory as well. On the domain considered, original prospect theory is not subject to the theoretical problems that have been pointed out for other choices (Handa 1977; Fishburn 1978). The normalization U(0) = 0, necessary in prospect theory when loss outcomes are present, is not required in our domain because it does not affect preferences here.

Similar to the utility function, the function w is subjective and depends on the individual, reflecting sensitivity towards probabilities. Many empirical investigations have studied the shape of w. Figure 1 depicts the prevailing shape (Abdellaoui 2000; Bleichrodt & Pinto 2000; Camerer & Ho 1994; Gonzalez & Wu 1999; Kachelmeier & Shehata 1992; Karni & Safra 1990; Prelec 1998; Quiggin 1982; Tversky & Kahneman 1992; Yaari 1965). For counter-evidence, see Birnbaum & Navarrete (1998) and Harbaugh, Krause, & Vesterlund (2002).

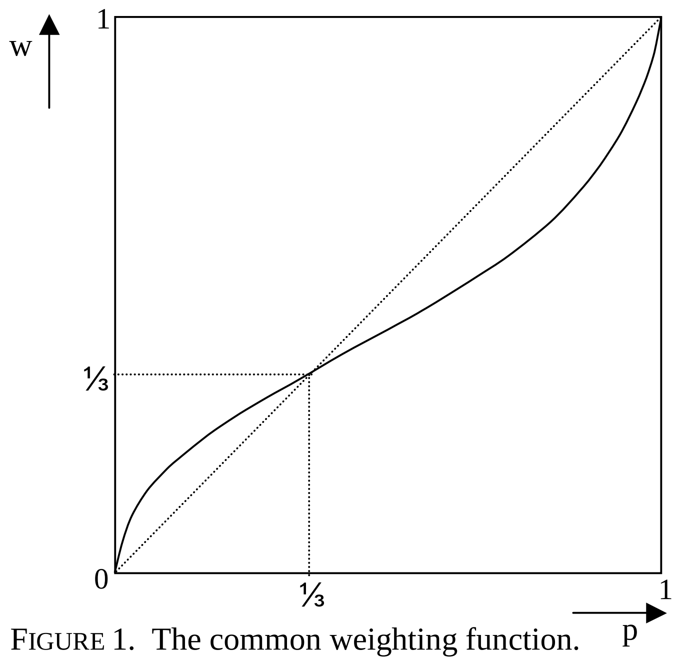

FIGURE 1. The common weighting function.

Under expected utility, all risk aversion has to be captured through concave utility whereas under the descriptively more realistic prospect theory, part of the observed risk aversion is due to probability weighting. This suggests that classical estimations of utility are overly concave. A theoretical justification for this claim was provided by Rabin (2000). Our paper will provide data that supports Rabin’s claims, and will show that prospect theory can explain these data.

5. An Experimental Comparison of Choice-Based and Choiceless Utilities

This section presents the first two measurement methods, the, choice-based, tradeoff method and the, choiceless, strength-of-preference method.

Participants and Stimuli. We recruited 50 students from the department of economics of the Ecole Normale Supérieure of Cachan. Each participant was paid FF 150 ($1 ≈ FF 6). No performance-based payments could be used for reasons discussed in Section 7. Each participant was interviewed individually by means of a computer program, in the presence of the experimenter. The participants were familiar with probabilities and expectations but had not received a training in decision theory before the experiment. Prior to the experimental questions, the participants were familiarized with the stimuli through some practice questions. Three participants were discarded because they gave erratic answers and apparently did not understand the instructions; N = 47 participants remained.

Our choice-based method concerns risky choices. Only degenerate or two-outcome prospects were used. They were displayed as pie charts on a computer screen, where different colors were used to designate different areas; see Appendix A. The units of payment in the prospects were French Francs. At the beginning of the experiment, a random device repeatedly picked random points from the pie charts so as to familiarize the participants with the representation of probabilities used in this experiment.

The measurements of this paper are based on indifferences. It is well known that observations of indifferences are prone to many biases, in particular if derived from direct matching. Indifferences derived from choices seem to be less prone to biases (Bostic, Herrnstein, & Luce 1990; Tversky & Kahneman 1992 p. 306). We developed software for carefully observing indifferences while avoiding biases. Appendix A gives details. We assessed three to six points for fitting the utility functions; using such numbers of points was recommended by von Winterfeldt & Edwards (1986, p. 254).

We used a within-subject design, with all measurements carried out for all individuals. All statistical analyses are based on within-subject differences. The tradeoff method was always carried out before the other methods because its answers served as inputs in further elicitations, so as to simplify the comparisons. The order of the other methods was counterbalanced so as to minimize systematic memory effects, which is especially important for the strength of preference measurements.

Measurement methods. For the tradeoff method (*TO method*), we used "gauge outcomes" R and r with R = FF 2000 > r = FF 1000. An outcome t_0 was set at FF 5000 (FF 1 ≈ \$0.17). For each participant, the outcome $t_1 > t_0$ was assessed such that $(⅓,t_1; r) \sim (⅓,t_0; R)$. Next, $t_2 > t_1$ was assessed such that $(⅓,t_2; r) \sim (⅓,t_1; R)$, …, and, finally, $t_6 > t_5$ was assessed such that $(⅓,t_6; r) \sim (⅓,t_5; R)$. Under prospect theory, the indifferences imply the five equalities $U(t_6) - U(t_5) = \cdots = U(t_1) - U(t_0)$, independently of how the participant transforms probabilities (Wakker & Deneffe 1996). Because EU is a special case of PT with a linear weighting function, the five equalities also hold under EU. Setting, as throughout this paper, $U(t_0) = 0$ and $U(t_6) = 1$, we obtain the following equalities.

$$U(t_i) = \frac{i}{6} \text{ for all i.} \tag{5.1}$$

The TO observations can be interpreted as direct observations of the inverse utility function, with $t_i = U^{inv}(i/6)$ for all i.

Our choiceless method for measuring utility is based on direct strength-of-preference judgments (*SP method*). For each participant, an amount s_2 was assessed such that the strength of preference between s_2 and t_1 was judged to be the same as between t_1 and t_0, the values obtained from the TO method (for details see Appendix A). Similarly, we elicited amounts s_3, ..., s_6 such that the strength of preference between s_i and s_{i-1} was judged to be the same as that between t_1 and t_0, for all i. Following Alt (1936) and others, the SP method assumes that strength-of-preference judgments correspond with utility differences, implying

$$U(s_6) - U(s_5) = \ldots = U(s_3) - U(s_2) = U(s_2) - U(t_1) = U(t_1) - U(t_0).$$

Using the scaling convention $U(t_1)–U(t_0) = 1/6$ (as in Eq. 5.1), we have

$$U(s_i) = \frac{i}{6} \text{ for all i.} \tag{5.2}$$

Note that these strength-of-preference measurements indeed do not involve observed choices in the sense of revealed preferences (Samuelson 1938a; Varian 1993). The various attempts to relate strength-of-preference judgments to choice making in a theoretical model after all, through side payments such as hours of labor, repeated or probabilistic choices, etc., are all based on separability assumptions that beg the question of cardinal utility. Strengths of preferences have, therefore, not been part of the commonly accepted empirical domain under the ordinal view of utility.

Analysis. In each test in this paper, the null hypothesis H_0 assumes identical utility functions for the various methods. For testing group averages, we considered paired *t*-tests and Wilcoxon signed rank tests, all two-tailed, which always gave the same results. Conclusions based on accepted null hypotheses are most convincing under the most powerful tests, i.e. the *t*-tests. Hence, we usually report those. To reckon with individual differences, our main conclusions, presented in later sections, will be based on analysis of variance with repeated measures whenever possible. These analyses always give the same conclusions as paired *t*-tests.

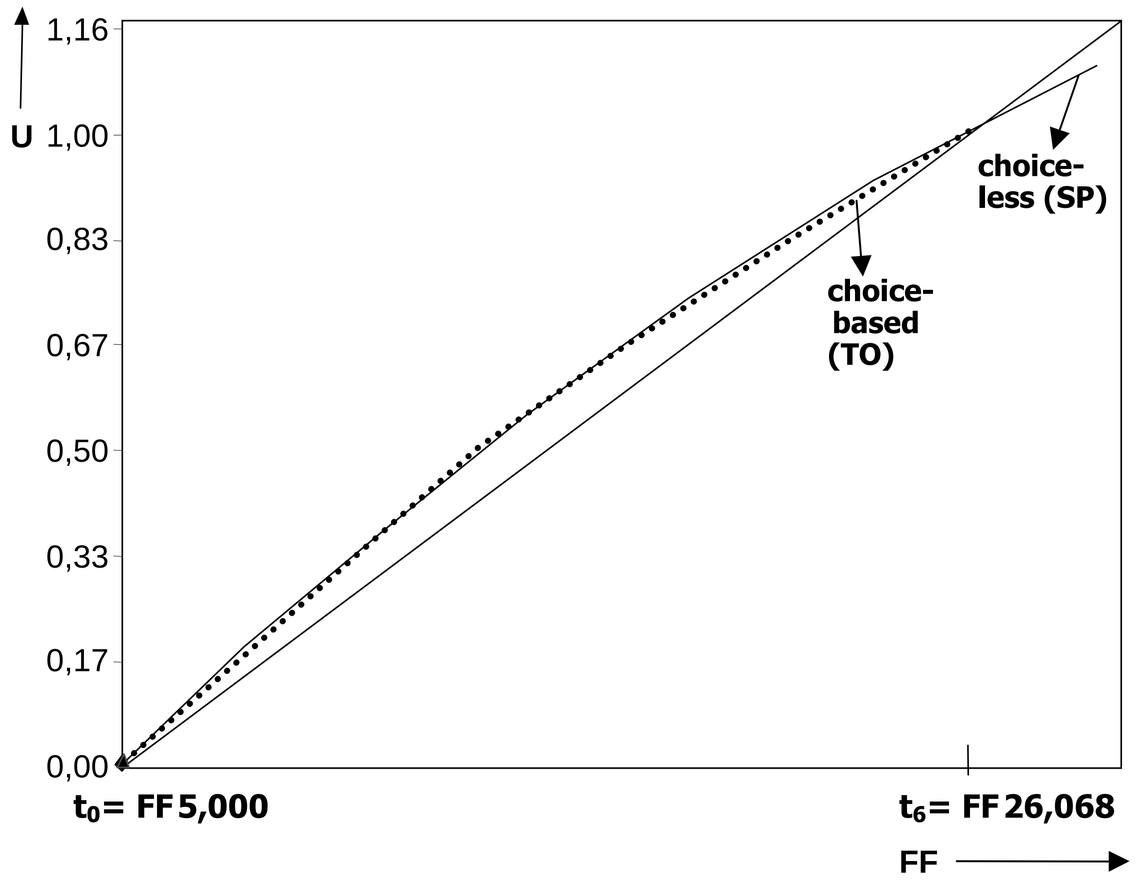

FIGURE 2. A choice-based and a choiceless utility function (for mean values).

Results. The mean values of the variables t_i and s_i are depicted in the TO and SP curves in Figure 2, which were obtained through linear interpolation. Numerical details are in Table 1 in Appendix B. The curves, based on averages, can be interpreted as the utility functions of a representative agent. The figure suggests that the choice-based and choiceless utility curves are the same. This suggestion is confirmed by statistical analyses.

For each j we have $s_j = t_j$ under H_0 because both should then have utility j/6 (Eqs. 5.1 and 5.2). H_0 is rejected for no j, with p-values ranging from .118 to .211. The equality is confirmed by parametric fittings, depicted in the upper two panels of Figure 4 and analyzed in Appendix D.

Linearity of the TO- and SP utility curves in Figure 2 was tested through Friedman tests, and was rejected for both TO and SP (H_0 for TO: $t_{j-1} - t_j$ is independent of j, $\chi^2_5 = 29.6$, $p < .001$; H_0 for SP is similar, $\chi^2_5 = 38.05$, $p < .001$). It was also rejected by the parametric analyses in Appendix D.

Psychological explanation for the equality of choiceless SP utilities and choice-based TO utilities. Under expected utility, the risky utility function was traditionally distinguished from riskless concepts because the former should comprise all aspects of risk attitudes, which obviously play no role for the latter concepts. Under prospect theory, aspects of risk attitudes beyond the utility of outcomes can be modeled through probability weighting (and loss aversion for negative outcomes). It then becomes conceivable, at least as an empirical hypothesis to be tested, that the utility function of prospect theory agrees with riskless concepts.

From a psychological perspective, it is not surprising that the choice-based and choiceless utilities measured in this paper agree, because the TO method appeals to a perception of preference in an indirect manner: In the indifference $(⅓,t_i; r) \sim (⅓,t_{i-1}; R)$, a perceived strength of preference between t_i and t_{i-1}, associated with probability ⅓, offsets the same counterargument of receiving R instead of r, associated with probability 2/3, for each i. Because the relevant probabilities are the same for each i, it is plausible that the perceived strength of preference between t_i and t_{i-1} is the same for each i (Wakker & Deneffe 1996). In this way, it is not surprising that the TO and SP methods gave similar results.

6. Verifying the Validity of Measurements

A pessimistic interpretation of the equality found in the preceding section can be devised, in agreement with the constructive view of preference (Gregory, Lichtenstein, & Slovic 1993; Loomes, Starmer, & Sugden 2003): The participants may simply have used similar heuristics in both methods used, and their TO answers may not reflect genuine preference. To investigate this possibility, we used a third, traditional, method for measuring utility, a certainty-equivalent method. For the first 13 participants, only TO and SP measurements were conducted. Then it was realized that further questions were feasible. Therefore, for the remaining 34 participants not only TO and SP measurements, but also two certainty equivalent measurements were conducted.

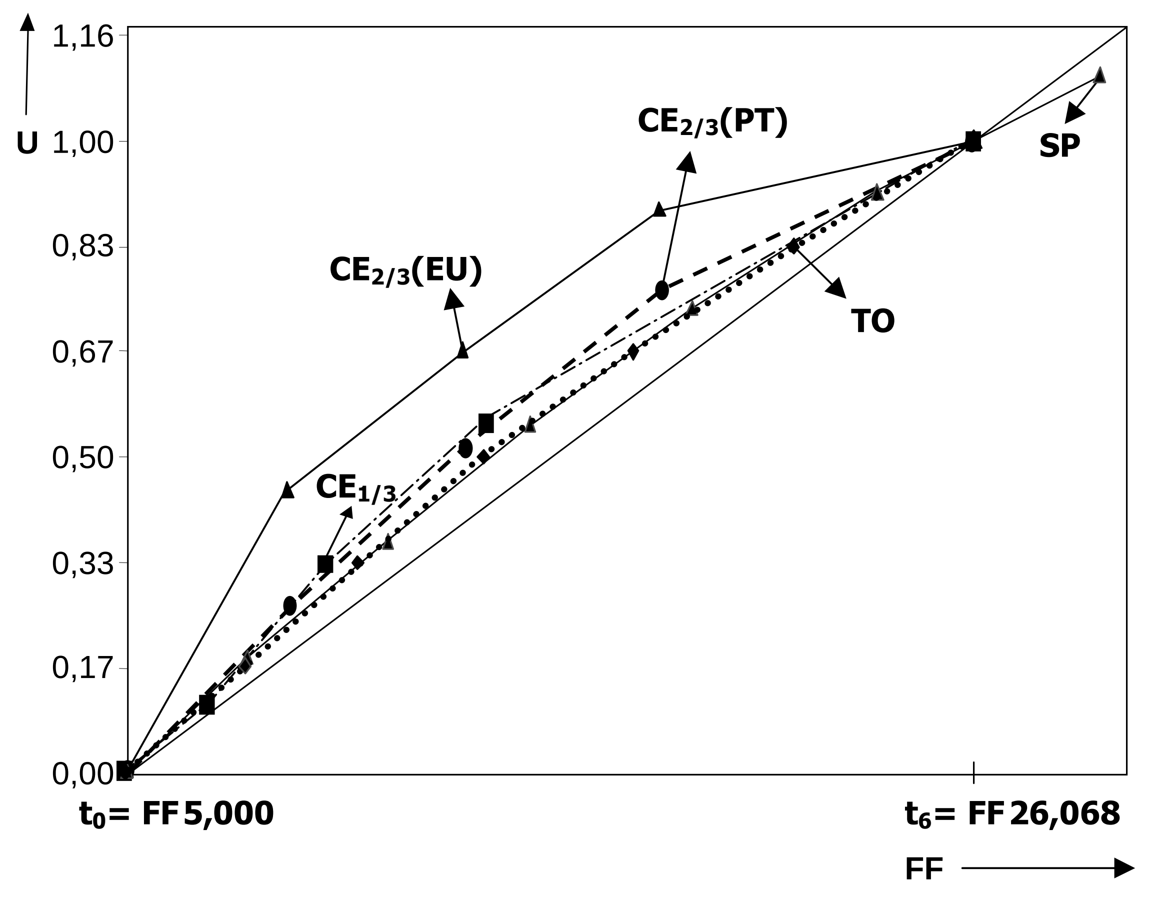

FIGURE 3. All utility functions (for mean values).

Certainty-equivalent methods compare sure amounts of money to two-outcome prospects and have been used in many studies (von Winterfeldt & Edwards 1986).

They have a format different than TO and SP methods. Therefore, if heuristics are used, it is plausible that they will be different for certainty equivalents than for the TO and SP methods, and that they will not generate the same utilities. The third method considered prospects that assign probability 1/3 to the best outcome. The reason for this particular choice of probability will be explained at the end of this section. The third method is called the *$CE_{1/3}$ method*. Amounts c_2, c_1, and c_3 were elicited such that $c_2 \sim (\frac{1}{3},t_6; t_0)$, $c_1 \sim (\frac{1}{3},c_2; t_0)$, and $c_3 \sim (\frac{1}{3},t_6; c_2)$.

We first analyze this method in the classical manner, i.e., assuming EU. We will see later that the following equalities and analysis remain valid under prospect theory. With $U(t_0) = 0$ and $U(t_6) = 1$, we get:

$$U(c_2) = \frac{1}{3},\ U(c_1) = \frac{1}{9},\ \text{and } U(c_3) = \frac{5}{9}. \tag{6.1}$$

All nonparametric utility curves measured in our experiment, based on group averages and linear interpolation, are assembled in Figure 3. Figure 4 gives the average result of parametric fittings, explained in Appendix C. The figures suggest that the average utility function resulting from the $CE_{1/3}$ observations agrees well with the TO and SP utility functions. Analysis of variance with repeated measures for the parametric fittings confirms the equality of the TO, SP, and $CE_{1/3}$ measurements while taking into account differences at the individual level, with $F(2, 66) = 0.54$, $p = 0.58$. The same conclusion follows from other statistical analyses reported in Appendices B and D.

At this point, two concerns can be raised. First, it may be argued that the assumption of EU used in the preceding analysis is not descriptively valid. Second, it may be conjectured that our design does not have the statistical power to detect differences (apart from nonlinearity of the utility curves). To investigate these concerns, we used a fourth method for measuring utility, another certainty-equivalent method. This method considered prospects that assign probability 2/3 to the best outcome and is, therefore, called the *$CE_{2/3}$ method*. The same 34 individuals participated as in the $CE_{1/3}$ method. Amounts d_2, d_1, and d_3 were elicited such that $d_2 \sim (\frac{2}{3},t_6; t_0)$, $d_1 \sim (\frac{2}{3},d_2; t_0)$, and $d_3 \sim (\frac{2}{3},t_6; d_2)$. We first analyze this method assuming EU. With $U(t_0) = 0$ and $U(t_6) = 1$, the following equalities are implied.

$$U(d_1) = \frac{4}{9},\ U(d_2) = \frac{2}{3},\ \text{and } U(d_3) = \frac{8}{9}. \tag{6.2}$$

The average utility function resulting from the $CE_{2/3}$ observations under EU is depicted as the $CE_{2/3}$(EU) curve in Figure 3 for linear interpolation, and in the middle right panel in Figure 4. The function strongly deviates from the other curves. Whereas analysis of variance with repeated measures for the parametric fittings concluded that the three measurements (TO, SP, $CE_{1/3}$) are the same, addition of $CE_{2/3}$(EU) leads to the conclusion that the four measurements (TO, SP, $CE_{1/3}$, $CE_{2/3}$(EU)) are not the same, $F(3,99) = 6.39$, $p = 0.001$. That $CE_{2/3}$(EU) is different from the other measurements, is confirmed by other statistical analyses, such as pairwise comparisons, presented in Appendices B and D. This finding falsifies EU and agrees with the EU violations documented in the literature.

We reanalyze the results of the certainty-equivalent methods by means of prospect theory, and correct the utility measurements for probability weighting. Such corrections were suggested before by Fellner (1961, p. 676), Wakker & Stiggelbout (1995), Stalmeier & Bezembinder (1999), and Bleichrodt, Pinto, & Wakker (2001). We assume the probability weighting function of Figure 1 for all individuals. This assumption obviously is an approximation because in reality the probability weighting function will depend on the individual. The descriptive performance of prospect theory could be improved if information about individual probability weighting were available. In the absence of such information, we expect that, on average, PT with the probability weighting function of Figure 1 will yield better results than EU, which also assumes that the weighting function is the same for all individuals but, furthermore, assumes that it is linear. Let us repeat that the analysis of the TO method remains valid under PT, irrespective of the individual probability weighting functions. Therefore, contrary to the CE methods, it is not affected by individual variations in probability weighting.

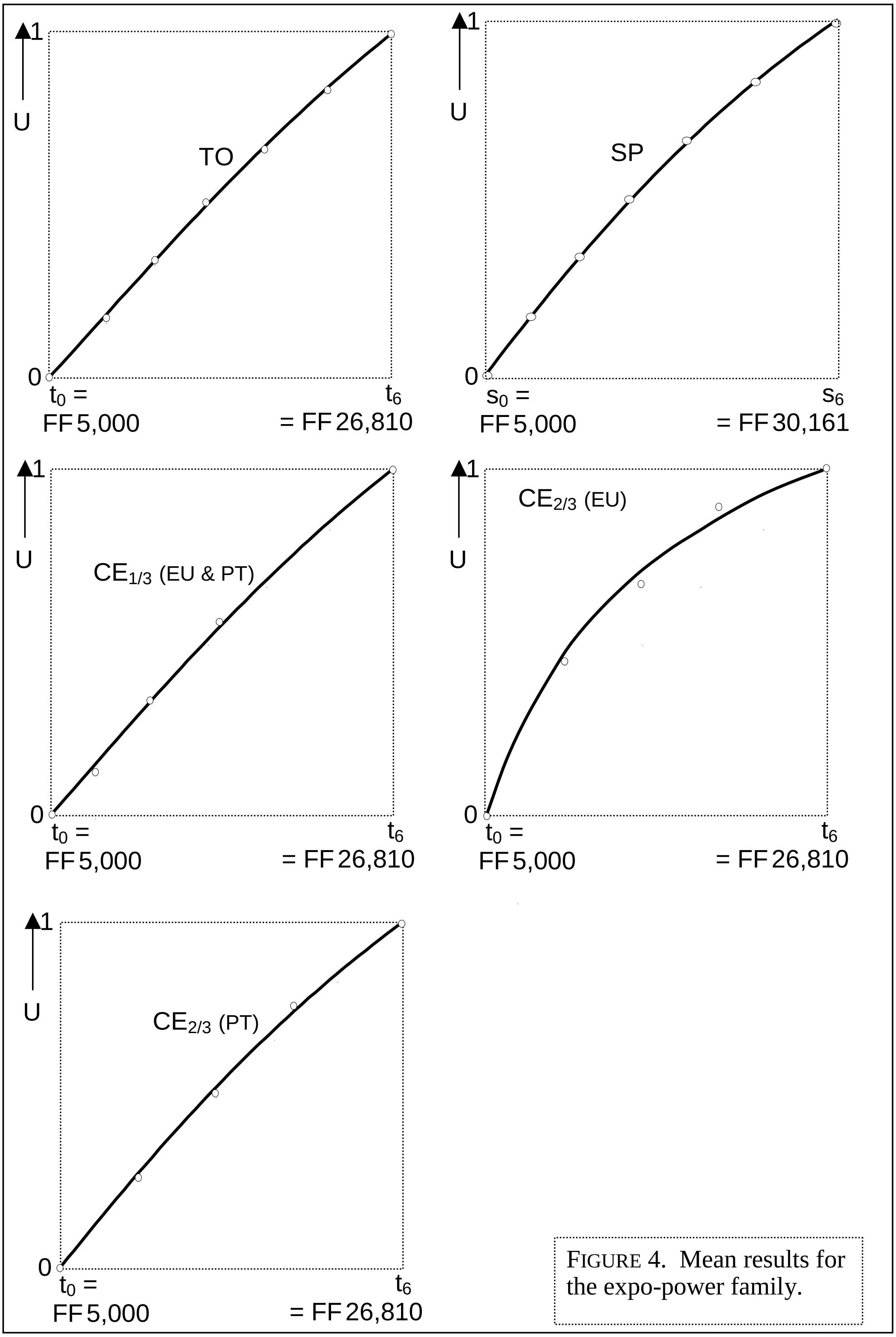

FIGURE 4. Mean results for the expo-power family.

It has been found that, on average, w(⅓) is approximately ⅓ (see Figure 1). Therefore, our analysis of $CE_{1/3}$ needs no modification and Eq. 6.1 and the utility function depicted in Figure 1 remain valid under PT. Accordingly, the agreement between the $CE_{1/3}$ utilities and the TO utilities also remains valid. It has been found that w(⅔) is approximately .51 (see the references given at Figure 1). Hence, the analysis of $CE_{2/3}$ that was based on EU needs modification. We now find

$$U(d_1) = 0.26,\ U(d_2) = 0.51,\ \text{and}\ U(d_3) = 0.76 \quad (6.3)$$

instead of Eq. 6.2. The resulting corrected utility curves, denoted $CE_{2/3}$(PT), are depicted in Figures 3 and 4. They agree well with the TO, SP, and $CE_{1/3}$ curves. Analysis of variance with repeated measures for the parametric fittings confirms the equality of the TO, SP, $CE_{1/3}$, and $CE_{2/3}$(PT) measurements, with $F(3,99) = 0.63$, $p = 0.6$. In other words, replacing $CE_{2/3}$(EU) by $CE_{2/3}$(PT) restores the equality of utility. The equality is confirmed by other statistical analyses, reported in Appendices B and D.

7. Discussion

The statistical analyses suggested that the TO, (SP), $CE_{1/3}$, and $CE_{2/3}$(PT) utilities are the same, but that $CE_{2/3}$(EU) gives different values. According to PT, the discrepancy between the $CE_{2/3}$ utilities, derived under EU, and the other utilities found, is caused by violations of EU. After correction for these violations, a reconciliation of the different risky utility measurements, TO, $CE_{1/3}$, and $CE_{2/3}$, results. The reconciliation suggests one consistent cardinal index of utility for risk, supporting the results of the TO measurements indeed. It entails a positive result within the revealed-preference paradigm. The further agreement of this index with the SP index extends beyond the domain of revealed preference, and is the main message of this paper.

The role of real incentives has often been debated, and their importance is now generally acknowledged (Binmore 1999; Smith 1982). Real incentives are commonly implemented for moderate amounts of money. Utility measurement is, however, of interest only for significant amounts of money, for two reasons. First, important decisions typically involve large amounts of money and, second, utility is close to linear for moderate amounts so that no measurement is needed there (Marshall 1890; Rabin 2000; Savage 1954 p. 60). For these reasons, we had to use significant amounts and could not implement real incentives.

Camerer & Hogarth (1999) and Hertwig & Ortmann (2001) surveyed the role of real incentives. Real incentives improve performance in cognitively demanding tasks such as predicting company bond ratings (Camerer & Hogarth 1999, Table 1). Real incentives reduce variance and increase general risk aversion but do not affect results otherwise for simple tasks such as choices between simple prospects, the topic of this paper. Kachelmeier & Shehata (1992) confirmed this claim for high stakes. Some studies have reported negative effects of real incentives upon intrinsic motivations (Frey & Oberholzer-Gee 1997; Gneezy & Rustichini 2000; Loewenstein 1999 Section 5). In summary, because real incentives do not have much impact on choices in the domain of our study, we have carried out this investigation even though real incentives could not be implemented. The utility function for money is central in economics and its experimental measurement deserves investigation (Stigler 1950 Section IV.c), even if a resort to hypothetical choices cannot be avoided (Shafir, Diamond, & Tversky 1997, p. 350).

We used two other utility measurement methods not reported here, an unchained certainty equivalent method where we elicited values x_j equivalent to $(j/6,t_6; t_0)$, and a lottery equivalent method (McCord & de Neufville 1986) where we elicited probabilities q_j to give equivalences $(q_j,t_j; t_0) \sim (0.75,t_6; t_0)$, $j = 1,\dots,5$. The former method gave the same results as the methods reported in this paper, with utilities diverging significantly from TO, $CE_{1/3}$, … etc. under expected utility, but convergence re-established under prospect theory. Under the lottery equivalent method, there was partial divergence from TO etc. under expected utility, but prospect theory did not improve the case and even enlarged the divergence. The results of the lottery equivalent method may be explained by a bias upward due to scale compatibility that has been found to bias probability matching questions (Bleichrodt 2002). The data of the two methods discussed here were noisier than those of the other methods, and these two methods have not been widely used in the literature. Therefore, we do not report their details. They are available in Barrios (2003).

We next compare our findings to existing empirical findings in the literature, beginning with studies of choice-based utilities. For this context, there have been several studies that found results similar to ours. Karmarkar (1978) and McCord & de

Neufville (1986) found that utilities, measured through certainty equivalents with different probabilities, are inconsistent when analyzed by means of EU. Abdellaoui (2000) found that the utilities measured by the TO method are not affected by probability weighting. Bleichrodt, Pinto, & Wakker (2001) found that corrections by means of the probability weighting function of Figure 1 reconcile discrepancies between choice-based utilities. Tversky & Fox (1995 p. 276) pointed out the appealing feature of the ⅓ probability that it does not lead to systematic probability transformations in CE questions. Finally, Rabin (2000) argued on theoretical grounds that utility is more linear than commonly thought, and that most of the commonly observed risk aversion is due to factors other than utility curvature.

Rabin's argument is based on a paradox entailing that, if risk attitude is based solely on utility curvature as in expected utility, then a moderate and realistic degree of risk aversion for moderate stakes necessarily implies an extreme and unrealistic degree of risk aversion for high stakes. We used prospect theory, where risk attitude consists of other factors besides utility curvature, to estimate the utility function. Our empirical findings of moderate utility curvature confirm Rabin's predictions. Our contribution to Rabin’s paradox is to demonstrate that not only does it refute expected utility, but also it can be accommodated by prospect theory.

For the economic literature, the novelty of our study lies in the comparison of choice-based utilities, derived from prospect theory, with choiceless utilities derived from strength-of-preference judgments. The direct agreement between these measurements (alluded to by Camerer 1995, p. 625) is remarkable. Our measurements satisfy Birnbaum & Sutton's (1992) principle of scale convergence, according to which different ways to measure utility should give the same result. It would, indeed, be desirable if one concept of utility could emerge that is relevant for many contexts, such as decision under risk, welfare evaluations, intertemporal discounted utility, case-based reasonings (Gilboa & Schmeidler 1995), etc. (cf. Broome's 1991 index of goodness, or Robson 2001).

In applied domains, e.g. in health economics, it is common practice to use utilities measured in one context for applications in other contexts (Gold et al. 1996; Torrance, Boyle, & Horwood 1982). For example, risky utilities measured through the "standard gamble method" have been used in policy decisions about interpersonal tradeoffs (treating elderly versus young people) or intertemporal decisions (current prevention measures against future health impairments). Also choiceless utilities, measured through direct scaling questions or otherwise, have been used for decision making. The pragmatic justification is that no better data are available, and decisions have to be taken as good as possible with whatever is available. Empirical relations between various utilities have been studied extensively. See Frey & Stutzer (2000 p. 920), Pennings & Smidts (2000), Revicki & Kaplan (1993), Robinson, Loomes, & Jones-Lee (2001). Our improved procedures based on prospect theory provide a new way of studying such relationships.

We only compared risky choice-based utilities to riskless choiceless utilities derived from strengths of preferences, and we did not consider utilities derived from other tradeoffs such as interpersonal or intertemporal. We hope that future empirical studies will consider such other tradeoffs, and that Birnbaum and Broome's scale convergence can be established with one unified concept of utility relevant to many domains in social sciences. Then the use of choiceless data in applications, such as health economics, can become more acceptable to mainstream economists and ordinalists, not only for pragmatic reasons (Manski 2004), but also conceptually.

8. Conclusion

In the classical economic debate between cardinalists and ordinalists, the latter assumed that direct judgments, having no preference basis, are not meaningful. In the light of today's advances in experimental methods in economics, the question whether relations exist between direct judgments and preferences can be investigated empirically. The first investigations of such relations were conducted in decision theory. These investigations assumed expected utility theory, so that their results were distorted by the descriptive deficiencies of this theory. Prospect theory provided descriptive improvements. Using this theory, our experiment suggests a simple relation between direct strength-of-preference judgments and risky-decision utilities.

If an empirical relationship between direct judgments and preferences can be firmly established, then direct judgments will provide useful data for economic analyses in contexts where preferences are hard to measure because of choice anomalies (Kahneman 1994). Conversely, such links provide a consistency basis for direct judgments. The result will be that direct judgments reinforce the revealed preference approach and vice versa. We, therefore, hope for further empirical investigations of the relations between direct judgments and revealed preferences.

Appendix A. A Two-Step Procedure for Eliciting Indifferences

This appendix describes the new two-step procedure that was developed so as to obtain reliable indifferences. We first consider the measurement of t_1 for the TO method. That is, a value x (= t_1) was to be found to yield an indifference A = (1/3,5000; 2000) ~ (1/3,x; 1000) = B. Figure 5 displays these prospects (called propositions there) for x = 11000. The first step of our procedure established an interval containing t_1. It started with x = 5000, which clearly is a lowerbound for t_1 because the right prospect B then is dominated by the left prospect A. By means of a scrollbar, the experimenter next increased x to 25000, and here all participants preferred the right prospect B, so that 25000 is an upper bound for t_1 for all participants. These questions, yielding a preliminary interval [5000, 25000] containing t_1, served only to familiarize the subjects with the choices. The interval containing t_1 that we searched for was to be a narrower subinterval of [5000, 25000], and was obtained as follows.

The scrollbar was again placed at its initial value x = 5000, where B is dominated by A. The experimenter increased x until the participant was no longer sure that she prefers A. Next a smaller outcome x was found for which the participant was still sure to prefer A to B, say x = α > 5000. Similarly, an outcome x of B was found for which the participant was sure to prefer B to A, say x = β < 25000. Obviously, $\beta > \alpha$; if not, the participant did not understand the procedure and it was repeated. Thus, an interval $[\alpha, \beta]$ was obtained that contained the indifference value t_1. We wanted this interval to be of the same length for all participants. Hence, we asked participants to be more precise if their interval $[\alpha, \beta]$ was too long. Commonly it was shorter, in which case the computer automatically enlarged it. In this manner, an interval of a fixed length was obtained for the second step. Figure 5 displays the final result of Step 1 for a participant with $[\alpha, \beta]$ = [7000, 11000] as the interval of fixed length containing t_1.

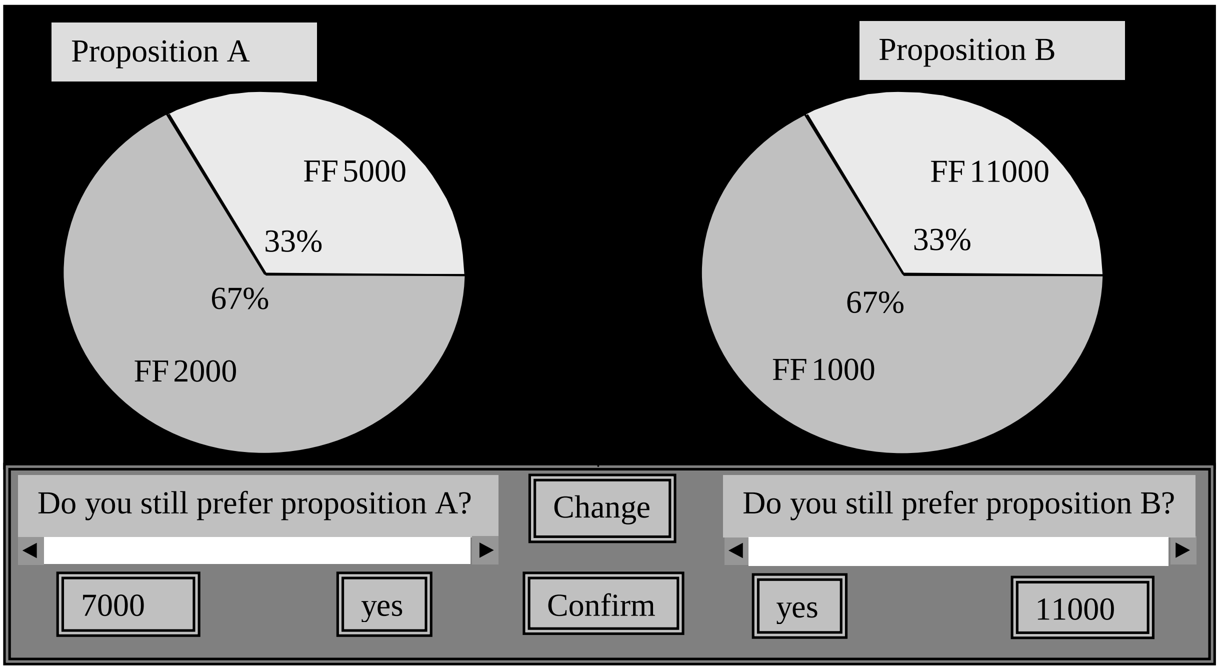

FIGURE 5. A screen used in the first step.

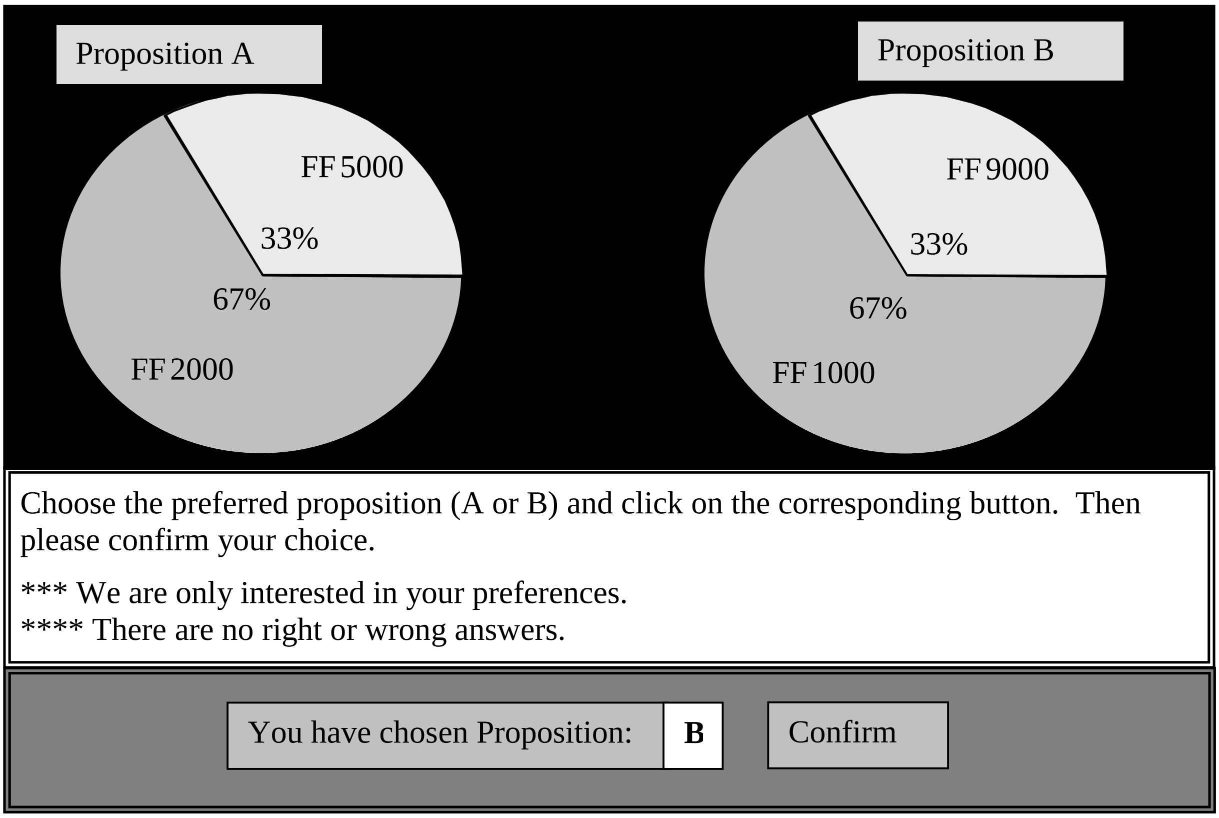

FIGURE 6. Presentation of the prospects in the second step.

In Step 2 of our procedure to elicit t_1, a choice-based bisection procedure was used to find the indifference value $x = t_1 \in [\alpha,\beta]$; see Figure 6. The midpoint $(\alpha+\beta)/2$ of the interval of Step 1 was substituted for x, and the participant was asked to choose between the prospects—indifference was not permitted. The midpoint was subsequently combined with the left or right endpoint of the preceding interval, depending on the preference expressed. In this manner, a new interval resulted that contained t_1 and that was half as large as the preceding interval. After five similar iterations, the interval was sufficiently narrow and its midpoint was taken as t_1. To test for consistency, we repeated the choice of the third iteration; it was virtually always (≥ 92% for each measurement) consistent in our experiment.

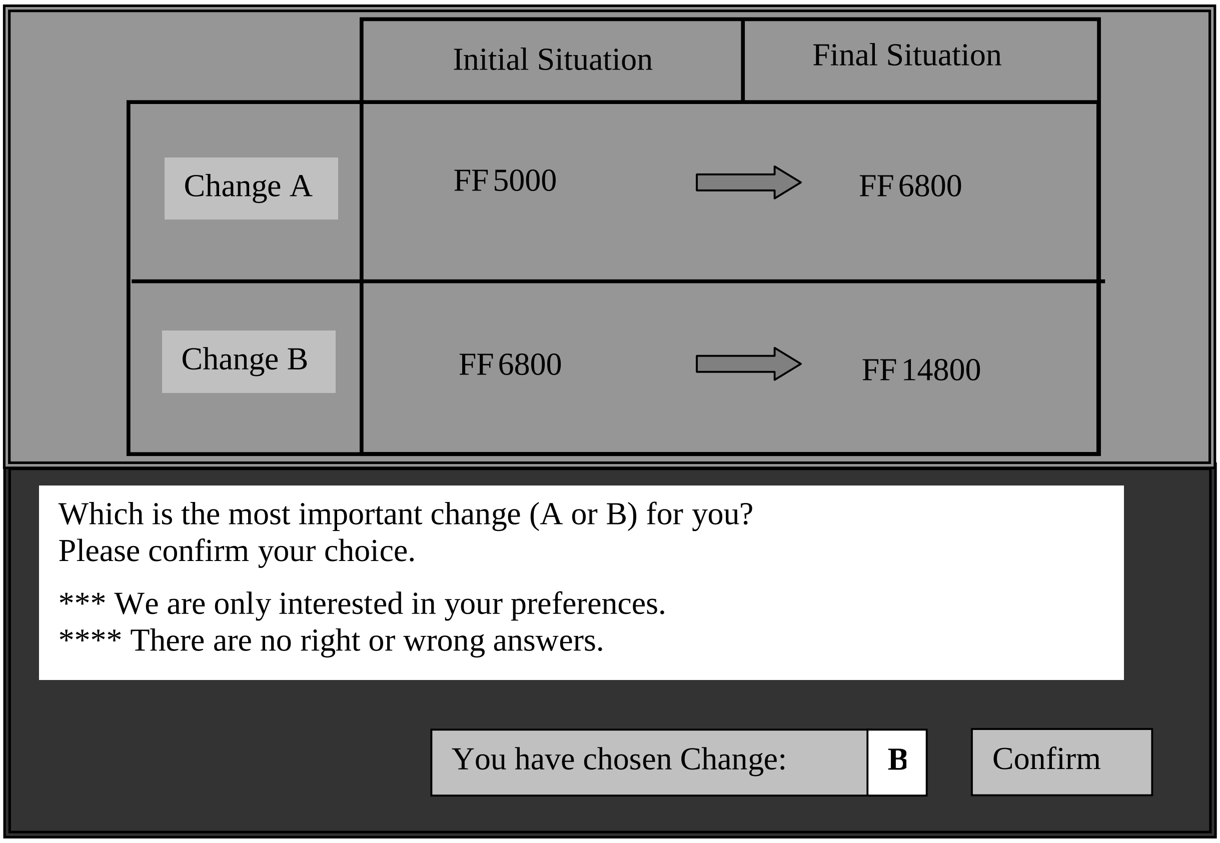

FIGURE 7. Presentation of strength of preference questions.

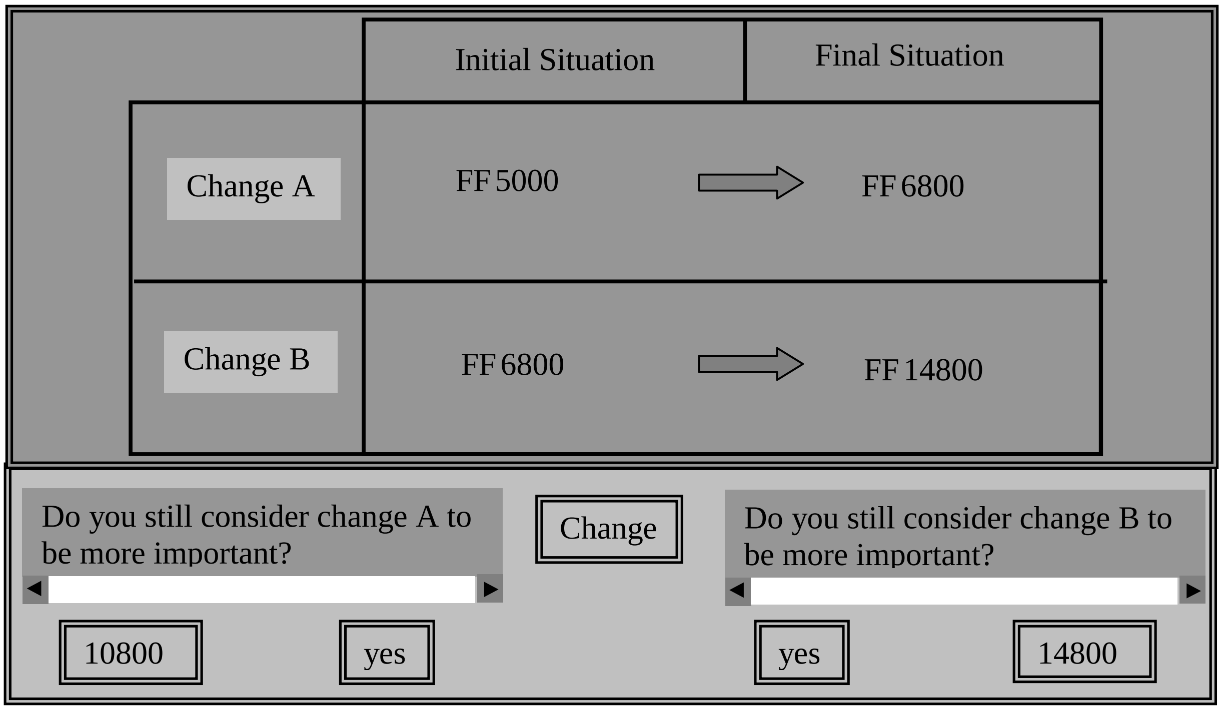

FIGURE 8. Presentation of strength of preference questions.

We adopted the above elaborate method of eliciting indifference values so as to obtain high-quality data, avoiding many biases that have been known to arise from direct matching questions (Bostic, Herrnstein, & Luce 1990). The indifference values $t_2,\ldots,t_6$ were elicited similarly. For the CE measurements we used the same way to elicit indifferences as for the TO measurements, now with one option being riskless. For the strength-of-preference measurements, a similar two-stage procedure was used but the stimuli were different because no prospects were involved. Figures 7 and 8 show the screens presented to the participants in the two stages.

Appendix B. Statistical Analysis of Raw Data

Table 1 gives descriptive statistics of our measurements. Paired t-tests of the equality of TO and SP measurements were described in the main text. We next consider paired *t*-test comparisons of the other two measurements with TO.

TABLE 1. Mean values in French francs. Standard deviations are in parentheses.

i	t_i	s_i	c_i	d_i
0	5000 *(0)*	5000 *(0)*		
1	8048 *(1318)*	8048 *(1318)*	7047 *(1055)*	8976 *(1964)*
2	11002 *(3022)*	11482 *(3067)*	10011 *(2201)*	13329 *(3754)*
3	14244 *(5332)*	15076 *(4932)*	13979 *(4214)*	18205 *(7338)*
4	18023 *(7864)*	19268 *(7275)*		
5	22165 *(11076)*	24210 *(10285)*		
6	26810 *(14777)*	30161 *(14644)*		

To compare TO with $CE_{1/3}$, note that $c_2 = t_2$ under H_0 because both should have utility ⅓ (Eqs. 5.1 and 6.1). Further comparisons between c- and t-values cannot be made directly because the observations concern different points in the domain. To nevertheless obtain comparisons, we use linear interpolations. (Other parametric fittings will be the topic of Appendix B.) Linear interpolation is best done on the scale with most observations, i.e. the TO scale. For example, if $U(t_0) = 0$ and $U(t_1) = 1/6$ (Eq. 5.1) then, by linear interpolation, $U(⅔t_1 + ⅓t_0) \approx 1/9$ and $⅔t_1 + ⅓t_0$ can be compared to c_1 (Eq. 5.2). As indicated in the fourth row of Table 2, $⅔t_4 + ⅓t_3$ can similarly be compared to c_3. The t_{33}- and p-values in the table indicate that no equality of the c-values and the corresponding (interpolations of) t-values is rejected statistically.

For the comparison of the $CE_{2/3}$ measurements with TO, note that $d_2 = t_4$ under H_0, because both should have utility ⅔ (Eqs. 5.1 and 6.2). Further comparisons require linear interpolations, indicated in Table 2. All equalities between TO- and $CE_{⅔}$-values, predicted by EU, are strongly rejected. If we reanalyze the data through PT, and adapt the linear interpolations correspondingly as indicated in the table, then the equality of utility is re-established.

TABLE 2. Direct tests of the consistency of choice-based methods.

theory	CEs	Utility	TOs*	t_{33}	p-value
EU & PT	c_1	1/9	$\frac{2}{3}t_1+\frac{1}{3}t_0$	0.09	.928
EU& PT	c_2	1/3	t_2	–1.49	.146
EU& PT	c_3	5/9	$\frac{1}{3}t_4+\frac{2}{3}t_3$	–1.52	.138
EU	d_1	4/9	$\frac{2}{3}t_3+\frac{1}{3}t_2$	–5.41	**.000**
EU	d_2	2/3	t_4	–4.30	**.000**
EU	d_3	8/9	$\frac{1}{3}t_6+\frac{2}{3}t_5$	–3.96	**.000**
PT	d_1	0.26	$.58\,t_2 + .42\,t_1$	–1.45	.158
PT	d_2	0.51	$.08\,t_4 + .92\,t_3$	–1.19	.244
PT	d_3	0.76	$.58\,t_5 + .42\,t_4$	–1.78	.084

*: interpolated t_i's

All tests in this appendix confirm the conclusions based on analysis of variance with repeated measures, reported in the main text. Nevertheless, a number of objections can be raised against the analyses of this appendix. For the scale that is interpolated, a bias downward is generated because utility is usually concave and not linear. For scales with few observations such as the CE scales, the bias can be big and, therefore, a direct comparison of $CE_{1/3}$ and $CE_{2/3}$ is not well possible. The latter problem is aggrevated because the different CE measurements focus on different parts of the domain.

The pairwise comparisons of the different points in Table 2 are not independent because the measurements are chained. Biases in measurements may propagate. This may explain why all five s_j values in Table 1 exceed the corresponding t_j values, although the difference is never significant. The differences can be explained by an overweighting of t_0 and t_1, due to their role as anchor outcomes in the SP measurements. While distorting the s_j's upwards, this bias hardly distorts the elicited utility curvature. For the latter, not the values of s_j or t_j per se, but their equal spacedness in utility units, is essential. This equal spacedness is affected only for the interval $[U(t_0),U(t_1)]$ under the SP method, which then is somewhat underestimated. For these reasons, it is preferable to investigate the curvature of utility, as opposed to the directly observed inverse utility values (this is what our observations t_i, s_i, c_i, d_i, in fact are). We investigate the curvature of utility through parametric fittings in the following appendices.

Appendix C. Fitting Parametric Utility Families

We fitted a number of parametric families to our data, and used the resulting parameters in the statistical analyses. All families hereafter were normalized so as to be on a same scale, and in this manner their numerical fits were compared. Because normalizations do not affect the empirical meaning of cardinal utility, we give non-normalized formulas hereafter as their notation is simpler. First, we considered the two families that have been used most frequently in the literature. Parametric fittings directly concern the curvature of utility, and smoothen out irregularities in the data. A drawback is that the results may depend on the particular parametric families chosen.

The *power family* is defined by

- x^r if $r > 0$
- $\ln(x)$ if $r = 0$
- $-x^r$ if $r < 0$.

A rescaling $z = x/t_6$ or $x/(t_6 - t_0)$ does not affect the preferences and, hence, need not be applied here. The translation $z = x - t_0$ leads to another family that will be discussed later. {#This family is most commonly used in the literature, and is also knows as the family of constant relative risk aversion (CRRA). Our results in Table 1 agree with those commonly found for individual choices with moderate stakes (Tversky & Kahneman 1992). In macro-economics and finance, market data are considered that concern bigger stakes, and then usually lower (more negative) powers are required to achieve required levels of concavity (Aït-Sahalia & Lo 2000; Bliss & Panigirtzoglou 2004; Gregory, Lamarche, & Smith 2002; Perraudin & Sorensen 2000; van Soest, Das, & Gong 2005). An additional reason why such studies find negative r is that they assume expected utility so that risk aversion generated by probability weighting is (mis)modeled through concave utility.#}

The *exponential family*, also knows as the family of constant absolute risk aversion (CARA), is defined by

- e^{rz} if $r > 0$
- z if $r = 0$
- $-e^{rz}$ if $r < 0$

where the domain $[t_0,t_6]$ is mapped into the unit interval through the transformation $z = (x - t_0)/(t_6 - t_0)$.

Several authors have suggested that utility is logarithmic (Bernoulli 1738; Savage 1954 p. 94) but this family did not fit our data well. It allows for concave utility only, whereas several participants exhibited convexities. Let us recall here that utility functions, when corrected for probability weighting, are less concave than traditional measurements have suggested. We also considered the translated power family where x is replaced by $x - t_0$. This family supported the empirical hypotheses of this paper equally well. We do not report its results because this family seems to be of limited empirical interest: Its derivatives at t_0 are extreme and the domain is not easily extended below t_0.

We introduce a new, third, parametric family, which we call the *expo-power family*, and which is defined by

$-\exp(-\frac{z^r}{r})$ for $r \neq 0$;[4]

$-\frac{1}{z}$ for $r = 0$.

We rescaled $z = \frac{x}{t_6}$. Figure 9 depicts some examples.

The expo-power family is a variation of a two-parameter family introduced by Saha (1993). The rescaling $z = \frac{x}{t_6}$ maps our domain $[t_0,t_6]$ to $[t_0/t_6,1] \subset [0,1]$. On $[0,1]$, the family exhibits some desirable features.

- r has a clear interpretation, being an anti-index of concavity (the smaller r the more concave the function is).
- The family allows for both concave ($r \leq 1$) and convex ($r \geq 2$) functions.
- There exists a subclass of this family ($0 \leq r \leq 1$) that combines a number of desirable features.
  - (i) The functions are concave;
  - (ii) The measure of absolute risk aversion, the Arrow-Pratt measure $-u''(x)/u'(x) = (1-r)/x + x^{r-1}$, is decreasing in x.

[4] For r close to zero, the strategically equivalent function $-\exp(-\frac{z^r}{r}+1/r)$ is more tractable for numerical purposes.

(iii) Finally, the measure of proportional risk aversion, $-xu''(x)/u'(x) = 1 - r + x^r$, is increasing in x.[5]

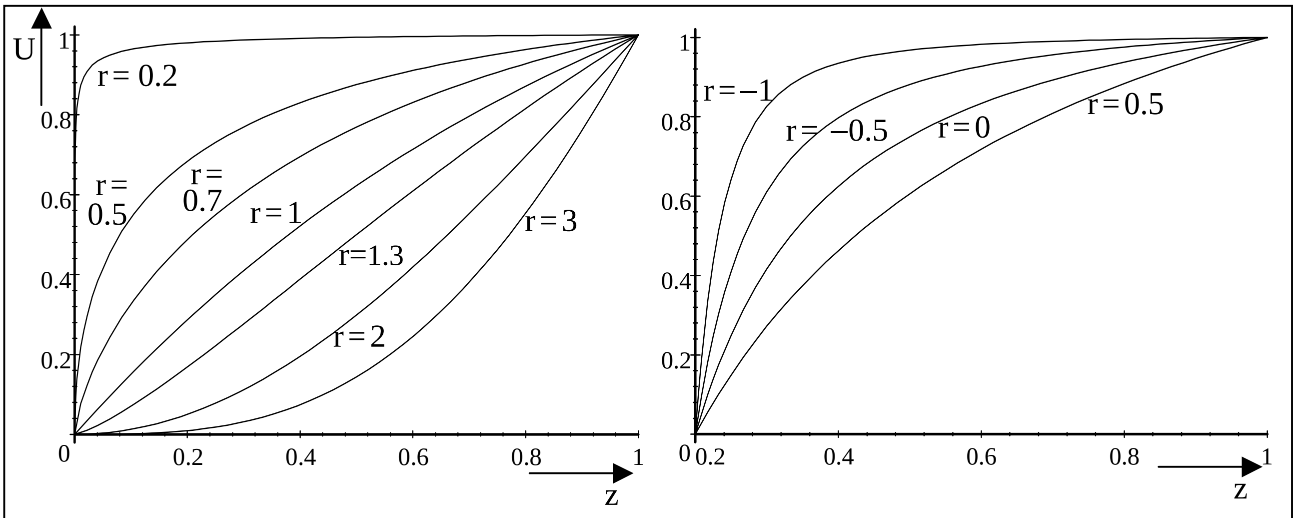

FIGURE 9. Some normalized curves of the expo-power family. For r=1.3, the curve is convex up to 0.40 and concave after.

Conditions (i), (ii), and (iii) are considered most relevant in the economics literature (Arrow 1971 p. 97; Binswanger 1981), but cannot be combined by any of the traditional parametric families. To allow for simultaneous satisfaction of all of these conditions, we developed the above variation of Saha's family. Necessarily, a one-parametric family with decreasing absolute risk aversion cannot contain linear functions and this is a drawback of our family. For r = 1.3, the curves are close to linear.

Other families have been considered in the literature. Merton (1971) introduced the HARA family with hyperbolic absolute risk aversion. This family does not allow for convex functions, because of which it does not fit our data well. In addition, it does not allow for a combination of the three desirable features (i), (ii), and (iii) listed above.

Bell (1988) and Farquhar & Nakamura (1987) characterized the family of all polynomial combinations of exponential functions. A subclass thereof is the general sumex family, consisting of all linear combinations of exponential functions and characterized by Nakamura (1996). In general, these families have many parameters

[5] So as to preserve this feature, we changed only the scale and not the location in the substitution $x \mapsto z(x)$.

and useful subfamilies remain to be identified. We did consider one two-parameter subfamily, being the sum of two exponential functions. The CE methods have only three data points, which is insufficient to determine the parameters in any reliable manner. The TO and SP methods have more data points and estimations of the two parameters were obtained. The null hypothesis of identity of the parameters was not rejected. Unfortunately, the parameter estimations were still unreliable and the test had little power. Therefore, it is not reported here.

Appendix D. Further Statistical Analyses of Parametric Estimations

Table 3 gives descriptive statistics for individual parametric estimates. Figure 4 in the main text depicts the optimal parametric fittings of the expo-power family for a representative agent. The parameters used there are: r = 1.242 for TO, r = 1.128 for SP, r = 1.206 for $CE_{1/3}$, r = 0.393 for $CE_{2/3}$(EU), r = 1.136 for $CE_{2/3}$(PT). These curves are based on averages of t_6 and t_1/t_6,..., t_5/t_6 for TO, s_1/t_6,..., and s_6/t_6 for SP, t_6 and c_1/t_6, c_2/t_6, c_3/t_6 for $CE_{1/3}$, and, finally, t_6 and d_1/t_6, d_2/t_6, d_3/t_6 for $CE_{2/3}$(EU) and $CE_{2/3}$(PT). The curves for power and exponential fittings are very similar.

TABLE 3

	Parametric Families								
	Power			Exponential			Expo-power		
	Median	Mean	St. Dev.	Median	Mean	St. Dev.	Median	Mean	St. Dev.
TO	0.77	0.91	0.70	0.28	0.29	0.90	1.29	1.33	0.75
SP	0.64	1.10	2.04	0.42	–0.14[a]	2.51	1.12	1.46	2.08
$CE_{1/3}$	0.88	1.03	1.23	0.10	0.39	1.73	1.31	1.44	1.21
$CE_{2/3}$(EU)	–0.33	–0.32	0.97	1.82	2.21	1.86	0.17	0.39	0.56
$CE_{2/3}$(PT)	0.77	0.83	1.01	0.23	0.25	1.95	1.30	1.27	0.94

[a] If one outlier, participant 28, is excluded then the mean parameter is 0.18 and the standard deviation is 1.35.

Wilcoxon tests rejected linear utility for the power family (H_0: r = 1), both for TO (z = –2.24, p < 0.05) and for SP (z = –2.32, p < 0.05), and likewise rejected linearity for the exponential family (H_0: r = 0; TO: z = –2.72, p < 0.05; SP: z = –2.42, p < 0.05). Because the expo-power family does not contain linear functions, no test of linearity was carried out for this family.

Table 4 presents the results of tests of equalities of utility parameters. The answers of participant 44 for the $CE_{2/3}$ questions could not be accommodated by the exponential family and the parametric fitting did not converge. This participant is excluded from the analysis of this family.

TABLE 4. Results of paired *t*-tests

	Parametric Families					
	Power		Exponential		Expo-power	
	t	p	*t*	p	*t*	p
TO – SP	t_{46}=–0.17	.867	t_{45}=–0.63	.532	t_{46}=–0.42	.677
TO – $CE_{1/3}$	t_{33}=–0.54	.590	t_{32}=–0.41	.682	t_{33}=–0.67	.511
TO – $CE_{2/3}$(EU)	t_{33}=6.76	**.000**	t_{32}=–6.27	**.000**	t_{33}=6.25	**.000**
TO – $CE_{2/3}$(PT)	t_{33}=0.002	.999	t_{32}=0.070	.945	t_{33}=0.23	.820
SP – $CE_{1/3}$	t_{33}=0.35	.730	t_{32}=–1.67	.105	t_{33}=0.61	.546
SP – $CE_{2/3}$(EU)	t_{33}=4.05	**.000**	t_{32}=–4.76	**.000**	t_{33}=2.98	**.005**
SP – $CE_{2/3}$(PT)	t_{33}=0.69	.493	t_{32}=–1.16	.255	t_{33}=0.91	.368
$CE_{1/3}$ – $CE_{2/3}$(EU)	t_{33}=5.23	**.000**	t_{32}=–7.19	**.000**	t_{33}=5.27	**.000**
$CE_{1/3}$ – $CE_{2/3}$(PT)	t_{33}=0.57	.572	t_{32}=0.43	.672	t_{33}=0.91	.370
$CE_{2/3}$(EU)– $CE_{2/3}$(PT)	t_{33}=–13.34	**.000**	t_{32}=10.09	**.000**	t_{33}=–8.13	**.000**

The conclusions are the same for all families and agree with the conclusions in the main text. The $CE_{2/3}$ measurements, when analyzed through EU, differ significantly from all the other measurements. The other measurements, including the $CE_{2/3}$ measurements when analyzed through PT, agree.

The statistics for analyses of variance with repeated measures described in the main text concerned the expo-power family. The other families give very similar statistics and the same conclusions.

References

Abdellaoui, Mohammed (2000), "Parameter-Free Elicitation of Utilities and Probability Weighting Functions," *Management Science* 46, 1497–1512.

Aït-Sahalia, Yacine & Andrew W. Lo (2000), "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics* 94, 9–51.

Allais, Maurice (1953), "Fondements d'une Théorie Positive des Choix Comportant un Risque et Critique des Postulats et Axiomes de l'Ecole Américaine," *Colloques Internationaux du Centre National de la Recherche Scientifique (Econométrie)* 40, 257–332. Paris: Centre National de la Recherche Scientifique. Translated into English, with additions, as "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School," *in* Maurice Allais & Ole Hagen (1979, Eds.), *Expected Utility Hypotheses and the Allais Paradox*, 27–145, Reidel, Dordrecht, The Netherlands.

Alt, Franz (1936), "Über die Messbarkeit des Nutzens." *Zeitschrift für Nationalökonomie* 7, 161–169. Translated into English by Siegfried Schach (1971), "On the Measurability of Utility." *In* John S. Chipman, Leonid Hurwicz, Marcel K. Richter, & Hugo F. Sonnenschein (Eds.), *Preferences, Utility, and Demand*, Hartcourt Brace Jovanovich, New York, Chapter 20.

Arrow, Kenneth J. (1971), "*Essays in the Theory of Risk Bearing*." North-Holland, Amsterdam.

Barrios, Carolina (2003), "Une Réconciliation de la Mesure de l'Utilité à l'Aide de la "Prospect Theory": Une Approche Experimentale," Ph.D. dissertation, ENSAM, Paris, France.

Baumol, William J. (1958), "The Cardinal Utility Which is Ordinal," *Economic Journal* 68, 665–672.

Bayoumi, Ahmed & Donald A. Redelmeier (2000), "Decision Analysis with Cumulative Prospect Theory," *Medical Decision Making* 20, 404–412.

Bell, David E. (1988), "One-Switch Utility Functions and a Measure of Risk," *Management Science* 34, 1416–1424.

Bentham, Jeremy (1789), "*The Principles of Morals and Legislation*." At the Clarendom Press, Oxford.

Bernoulli, Daniel (1738), "Specimen Theoriae Novae de Mensura Sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5, 175–192. Translated into English by Louise Sommer (1954), "Exposition of a New Theory on the Measurement of Risk," *Econometrica* 22, 23–36. Reprinted in Alfred N. Page (1968, Ed.), *Utility Theory: A Book of Readings*, Chapter 11, Wiley, New York. Revised translation in William J. Baumol &

Stephen M. Goldfeld (Eds., 1968), *Precursors in Mathematical Economics: An Anthology*. Clowes and Sons, London, Selection 2, 15–26.

Binmore, Kenneth G. (1999), "Why Experiments in Economics?," *Economic Journal* 109, F16–F24.

Binswanger, Hans P. (1981), "Attitudes towards Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal* 91, 867–890.

Birnbaum, Michael H. & Juan B. Navarrete (1998), "Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence," *Journal of Risk and Uncertainty* 17, 49–78.

Birnbaum, Michael H. & Sara E. Sutton (1992), "Scale Convergence and Utility Measurement," *Organizational Behavior and Human Decision Processes* 52, 183–215.

Blaug, Mark (1962), "*Economic Theory in Retrospect*." Cambridge University Press, Cambridge. (5th ed. 1997)

Bleichrodt, Han (2002), "A New Explanation for the Difference between SG and TTO Utilities," *Health Economics* 11, 447–456.

Bleichrodt, Han & José Luis Pinto (2000), "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis," *Management Science* 46, 1485–1496.

Bleichrodt, Han, José Luis Pinto, & Peter P. Wakker (2001), "Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility," *Management Science* 47, 1498–1514.

Bliss, Robert R. & Nikolaos Panigirtzoglou (2004), "Option-Implied Risk Aversion Estimates," *Journal of Finance* 59, 407–446.

Bostic, Raphael, Richard J. Herrnstein, & R. Duncan Luce (1990), "The Effect on the Preference-Reversal Phenomenon of Using Choice Indifferences," *Journal of Economic Behavior and Organization* 13, 193–212.

Bouyssou, Denis & Jean-Claude Vansnick (1988), "A Note on the Relationships between Utility and Value Functions." In Bertrand R. Munier (Ed.), Risk, Decision and Rationality, 103–114, Reidel, Dordrecht, The Netherlands.

Breiter, Hans C., Ithak Aharon, Daniel Kahneman, Anders Dale, & Peter Shizgal (2001), "Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses," *Neuron* 30, 619–639.

Broome, John R. (1991), "*Weighing Goods*." Basil Blackwell, Oxford, UK.

Camerer, Colin F. (1995), "Individual Decision Making." *In* John H. Kagel & Alvin E. Roth (Eds.), *Handbook of Experimental Economics*, 587–703, Princeton University Press, Princeton, NJ.

Camerer, Colin F. & Teck-Hua Ho (1994), "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty* 8, 167–196.

Camerer, Colin F. & Robin M. Hogarth (1999), "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty* 19, 7–42.

Cooter, Robert D. & Peter Rappoport (1984), "Were the Ordinalists Wrong about Welfare Economics?," *Journal of Economic Literature* 22, 507–530.

Cramer, Gabriel (1728), Letter from Cramer to Nicholas Bernoulli. Translated into English by Louise Sommer in Bernoulli, Daniel (1954), "Exposition of a New Theory on the Measurement of Risk," *Econometrica* 22, 23–36.

Debreu, Gérard (1959), "Cardinal Utility for Even-Chance Mixtures of Pairs of Sure-Prospects," *Review of Economic Studies* 26, 174–177.

Debreu, Gérard (1976), "Least Concave Utility Functions," *Journal of Mathematical Economics* 3, 121–129.

Edgeworth, F. Ysidro (1881), "Mathematical Physics, An Essay on the Application of Mathematics to the Moral Sciences." Reprinted 1967, M. Kelley, New York.

Ellingsen, Tore (1994), "Cardinal Utility: A History of Hedinometry." *In* Maurice Allais & Ole Hagen (Eds.), *Cardinalism; A Fundamental Approach*, 105–165, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Ellsberg, Daniel (1954), "Classic and Current Notions of 'Measurable Utility'," *Economic Journal* 62, 528–556.

Ellsberg, Daniel (1961), "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics* 75, 643–669.

Farquhar, Peter H. & Yutaka Nakamura (1987), "Constant Exchange Risk Properties," *Operations Research* 35, 206–214.

Fellner, Wiliam (1961), "Distortion of Subjective Probabilities as a Reaction to Uncertainty," *Quarterly Journal of Economics* 75, 670–689.

Fishburn, Peter C. (1978), "On Handa's "New Theory of Cardinal Utility" and the Maximization of Expected Return," *Journal of Political Economy* 86, 321–324.

Frey, Bruno S. & Felix Oberholzer-Gee (1997), "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review* 87, 746–755.

Frey, Bruno S. & Alois Stutzer (2000), "Happiness, Economy and Institutions," *Economic Journal* 110, 918–938.

Friedman, Milton & Leonard J. Savage (1948), "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy* 56, 279–304.

Frisch, Ragnar (1926), "Sur un Problème d'Economie Pure," *Norsk Matematisk Forenings Skrifter Serie 1* 16, 1–40. Translated into English by John S. Chipman, "On a Problem in Pure Economics." *In* John S. Chipman, Leonid Hurwicz, Marcel K. Richter, & Hugo F. Sonnenschein (1971, Eds.), *Preferences, Utility, and Demand*, Chapter 19, Hartcourt, New York.

Gilboa, Itzhak (1987), "Expected Utility with Purely Subjective Non-Additive Probabilities," *Journal of Mathematical Economics* 16, 65–88.

Gilboa, Itzhak & David Schmeidler (1995), "Case-Based Decision Theory," *Quarterly Journal of Economics* 110, 605–639.

Gilboa, Itzhak & David Schmeidler (2001), "A Cognitive Model of Individual Well-Being," *Social Choice and Welfare* 18, 269–288.

Gneezy, Uri, & Aldo Rustichini (2000), "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics* 115, 791–810.

Gold, Marthe R., Joanna E. Siegel, Louise B. Russell, & Milton C. Weinstein (1996), "*Cost-Effectiveness in Health and Medicine*." Oxford University Press, New York.

Gonzalez, Richard & George Wu (1999), "On the Shape of the Probability Weighting Function," *Cognitive Psychology* 38, 129–166.

Gregory, Allen W., Jean-François Lamarche, & Gregor W. Smith (2002), "Information-Theoretic Estimation of Preference Parameters: Macroeconomic Applications and Simulation Evidence," *Journal of Econometrics* 107, 213–233.

Gregory, Robin, Sarah Lichtenstein, & Paul Slovic (1993), "Valuing Environmental Resources: A Constructive Approach," *Journal of Risk and Uncertainty* 7, 177–197.

Grether, David M. & Charles R. Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69, 623–638.

Gul, Faruk (1991), "A Theory of Disappointment Aversion," *Econometrica* 59, 667–686.

Handa, Jagdish (1977), "Risk, Probabilities, and a New Theory of Cardinal Utility," *Journal of Political Economy* 85, 97–122.

Harbaugh, William T., Kate Krause, & Lise Vesterlund (2001), "Prospect Theory in Choice and Pricing Tasks,"

Harsanyi, John C. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63, 309–321.

Hershey, John C. & Paul J.H. Schoemaker (1985), "Probability versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent?," *Management Science* 31, 1213–1231.

Hertwig, Ralf & Andreas Ortmann (2001), "Experimental Practices In Economics: A Challenge for Psychologists?," *Behavioral and Brain Sciences* 24, 383–403.

Hicks, John R. & Roy G.D. Allen (1934), "A Reconsideration of the Theory of Value: I; II," *Economica* 1, 52–75; 196–219.

Houthakker, Hendrik S. (1950), "Revealed Preference and the Utility Function," *Economica*, N.S. 17, 159–174.

Kachelmeier, Steven J. & Mohamed Shehata (1992), "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review* 82, 1120–1141; for comment see Steven J. Kachelmeier & Mohammed Shehata (1994), *American Economic Review* 84, 1104–1106.

Kahneman, Daniel (1994), "New Challenges to the Rationality Assumption," *Journal of Institutional and Theoretical Economics* 150, 18–36.

Kahneman, Daniel & Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263–291.

Kahneman, Daniel & Amos Tversky (2000, Eds.), "*Choices, Values, and Frames*." Cambridge University Press, New York.

Kapteyn, Arie (1994), "The Measurement of Household Cost Functions: Revealed Preference versus Subjective Measures," *Journal of Population Economics* 7, 333–350.

Karmarkar, Uday S. (1978), "Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model," *Organizational Behavior and Human Performance* 21, 61–72.

Karni, Edi & Zvi Safra (1990), "Rank-Dependent Probabilities," *Economic Journal* 100, 487–495.

Krzysztofowicz, Roman & John B. Koch (1989), "Estimation of Cardinal Utility Based on a Nonlinear Theory," *Annals of Operations Research* 19, 181–204.

Lichtenstein, Sarah & Paul Slovic (1971), "Reversals of Preference between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology* 89, 46–55.

Loewenstein, George F. (1999), "Experimental Economics from the Vantage-Point of Behavioural Economics," *Economic Journal* 109, F25–F34.

Loomes, Graham & Robert Sugden (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty," *Economic Journal* 92, 805–824.

Loomes, Graham, Chris Starmer, & Robert Sugden (2003), "Do Anomalies Disappear in Repeated Markets?," *The Economic Journal* 113, C153–C166.

Machina, Mark J. (1982), " 'Expected Utility' Analysis without the Independence Axiom," *Econometrica* 50, 277–323.

Manski, Charles F. (2004), "Measuring Expectations," *Econometrica* 72, 1329–1376.

Marshall, Alfred (1890), "*Principles of Economics*." 8^{th} edition 1920 (9^{th} edition 1961), MacMillan, New York.

McCord, Mark R. & Richard de Neufville (1983), "Fundamental Deficiency of Expected Utility Decision Analysis." *In* Simon French, R. Hartley, L.C. Thomas, & Douglas J. White (Eds.), *Multi-Objective Decision Making*, 279–305, Academic Press, New York.

McCord, Mark R. & Richard de Neufville (1984), "Utility Dependence on Probability: An Empirical Demonstration," *Large Scale Systems* 6, 91–103.

McCord, Mark R. & Richard de Neufville (1986), " "Lottery Equivalents ": Reduction of the Certainty Effect Problem in Utility Assessment," *Management Science* 32, 56–60.

Merton, Robert C. (1971), "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory* 3, 373–413.

Nakamura, Yutaka (1996), "Sumex Utility Functions," *Mathematical Social Sciences* 31, 39–47.

Ng, Yew-Kwang (1997), "A Case for Happiness, Cardinalism, and Interpersonal Comparability," *Economic Journal* 107, 1848–1858.

Pareto, Vilfredo (1906), "*Manuele di Economia Politica*." Piccolo Biblioteca Scientifica, Milan. Translated into English by Ann S. Schwier (1971), "*Manuel of Political Economy*," MacMillan, London.

Translated into French in 1927 as "*Manuel d'Economie Politique*." Giard, Paris, second edition.

Pennings, Joost M.E. & Ale Smidts (2000), "Assessing the Construct Validity of Risk Attitude," *Management Science* 46, 1337–1348.

Perraudin, William R.M. & Bent E. Sorensen (2000), "The Demand for Risky Assets: Sample Selection and Household Portfolios," *Journal of Econometrics* 97, 117–144.

Prelec, Drazen (1998), "The Probability Weighting Function," *Econometrica* 66, 497–527.

Quiggin, John (1982), "A Theory of Anticipated Utility," *Journal of Economic Behaviour and Organization* 3, 323–343.

Rabin, Matthew (2000), "Risk Aversion and Expected-utility Theory: A Calibration Theorem," *Econometrica* 68, 1281–1292.

Ramsey, Frank P. (1931), "Truth and Probability." *In* "*The Foundations of Mathematics and other Logical Essays*," 156–198, Routledge and Kegan Paul, London.

Reprinted in Henry E. Kyburg Jr. & Howard E. Smokler (1964, Eds.), *Studies in Subjective Probability*, 61–92, Wiley, New York. (2nd edition 1980, Krieger, New York.)

Revicki, Dennis A. & Robert M. Kaplan (1993), "Relationship between Psychometric and Utility-Based Approaches to the Measurement of Health-Related Quality of Life," *Quality Life Research* 2, 477–487.

Robinson, Angela, Graham Loomes, & Michael Jones-Lee (2001), "Visual Analog Scales, Standard Gambles, and Relative Risk Aversion," *Medical Decision Making* 21, 17–21.

Robson, Arthur J. (2001), "Why Would Nature Give Individuals Utility Functions?," *The Journal of Political Economy* 109, 900–914.

Saha, Atanu (1993), "Expo-power Utility: A Flexible Form for Absolute and Relative Aversion," *American Journal of Agricultural Economics* 75, 905–913.

Samuelson, Paul A. (1937), "A Note on Measurement of Utility," *Review of Economic Studies* 4 (issue 2, February 1937), 155–161.

Samuelson, Paul A. (1938a), "The Numerical Representation of Ordered Classifications and the Concept of Utility," *Review of Economic Studies* 6 (Issue 1, October 1938), 65–70.

Samuelson, Paul A. (1938b), "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, N.S. 5, 61–71, 353–354.

Samuelson, Paul A. (1950), "The Problem of Integrability in Utility Theory," *Economica*, N.S. 17, 355–385.

Savage, Leonard J. (1954), "*The Foundations of Statistics*." Wiley, New York. (Second edition 1972, Dover Publications, New York.)

Schmeidler, David (1989), "Subjective Probability and Expected Utility without Additivity," *Econometrica* 57, 571–587.

Sen, Amartya K. (1974), "Informational Bases of Alternative Welfare Approaches. Aggregation and Income Distribution," *Journal of Public Economics* 3, 387–403.

Shafir, Eldar, Peter A. Diamond, & Amos Tversky (1997), "Money Illusion," *Quarterly Journal of Economics* 112, 341–374.

Simon, Herbert A. (1955), "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69, 99–118.

Skinner, Burrhus F. (1971), "*Beyond Freedom and Dignity*." Knopf, N.Y.

Smith, Vernon L. (1982), "Microeconomic Systems as an Experimental Science," *American Economic Review* 72, 923–955.

Starmer, Chris (2000), "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature* 38, 332–382.

Stigler, George J. (1950), "The Development of Utility Theory: I; II," *Journal of Political Economy* 58, 307–327; 373–396.
Reprinted in Alfred N. Page (1968), *Utility Theory: A Book of Readings*, Wiley, New York, 55–119.

Tinbergen, Jan (1991), "On the Measurement of Welfare," *Journal of Econometrics* 50, 7–13.

Torrance, George W., Michael H. Boyle, & Sargent P. Horwood (1982), "Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States," *Operations Research* 30, 1043–1069.

Tversky, Amos & Craig R. Fox (1995), "Weighing Risk and Uncertainty," *Psychological Review* 102, 269–283.

Tversky, Amos & Daniel Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.

van Praag, Bernard M.S. (1968), "*Individual Welfare Functions and Consumer Behavior*." North-Holland, Amsterdam, 1968.

Van Praag, Bernard M.S. (1991), "Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept," *Journal of Econometrics* 50, 69–89.

Van Soest, Arthur, Marcel Das, & Xiaodong Gong (2005), "A Structural Labour Supply Model with Flexible Preferences," *Journal of Econometrics* 107, 345–374.

Varian, Hal R. (1993), "*Intermediate Microeconomics*." Norton, New York.

von Neumann, John & Oskar Morgenstern (1944, 1947, 1953), "*Theory of Games and Economic Behavior*." Princeton University Press, Princeton NJ.

von Winterfeldt, Detlof & Ward Edwards (1986), "*Decision Analysis and Behavioral Research*." Cambridge University Press, Cambridge, UK.

Wakker, Peter P. & Daniel Deneffe (1996), "Eliciting von Neumann-Morgenstern Utilities when Probabilities Are Distorted or Unknown," *Management Science* 42, 1131–1150.

Wakker, Peter P. & Anne M. Stiggelbout (1995), "Explaining Distortions in Utility Elicitation through the Rank-Dependent Model for Risky Choices," *Medical Decision Making* 15, 180–186.

Watson, John B. (1913), "Psychology as the Behaviorist Views It," *Psychological Review* 20, 158–177.

Weber, Elke U. (1994), "From Subjective Probabilities to Decision Weights: The Effects of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events," *Psychological Bulletin* 115, 228–242.

Yaari, Menahem E. (1965), "Convexity in the Theory of Choice under Risk," *Quarterly Journal of Economics* 79, 278–290.

Yaari, Menahem E. (1987), "The Dual Theory of Choice under Risk," *Econometrica* 55, 95–115.

Zellner, Arnold (1971), *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

Zeuthen, Frederik (1937), "On the Determinateness of the Utility Function," *Review of Economic Studies* 4, 236–239.

Köbberling & Wakker waar?