Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 138 (2007) 356–378
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory

Mohammed Abdellaoui[a,*], Carolina Barrios[b], Peter P. Wakker[c]

[a]*GRID, CNRS, ENSAM & ESTP, 30 Avenue du Président Wilson 94230 Cachan, France*
[b]*Department of Industrial Engineering, Universidad de los Andes, Bogotá, Colombia*
[c]*Econometric Institute, Erasmus University, P.O. Box 1738, Rotterdam 3000 DR, The Netherlands*

Available online 7 August 2006

Abstract

In an experiment, choice-based (revealed-preference) utility of money is derived from choices under risk, and choiceless (non-revealed-preference) utility from introspective strength-of-preference judgments. The well-known inconsistencies of risky utility under expected utility are resolved under prospect theory, yielding one consistent cardinal utility index for risky choice. Remarkably, however, this cardinal index also agrees well with the choiceless utilities, suggesting a relation between a choice-based and a choiceless concept. Such a relation implies that introspective judgments can provide useful data for economics, and can reinforce the revealed-preference paradigm. This finding sheds new light on the classical debate on ordinal versus cardinal utility.

JEL Classification: B21, D81

Keywords: Cardinal utility; Ordinal revolution; Prospect theory; Risky utility; Strength of preference

1. Introduction

Utility has been a controversial concept throughout the history of economics, with interpretations shifting over time. Since the beginning of the 20th century, after what has since become known as the ordinal revolution, utility has been taken as an ordinal concept,

*Corresponding author. Tel.: +1 41 98 37 62; fax: +1 41 98 37 67.
E-mail address: abdellaoui@grid.ensam.estp.fr (M. Abdellaoui).

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.05.025

based solely on observable choice, in mainstream economics (Pareto, 1906). Ordinalism has dominated economics ever since (Hicks and Allen, 1934; Robbins, 1932).

In view of the many anomalies of observed choice that have been discovered in the 20th century, several authors have argued that a reinterpretation of utility that is 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. Under the name of happiness, introspective measurements of utility, studied extensively in the psychological literature, have also attracted interest in economics recently (Frey and Stutzer, 2002; van Praag and Ferrer-i-Carbonell, 2004).

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 on the measurement of utility and welfare in a special issue of the Journal of Econometrics:

> 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 utilities to experimental measurements of choiceless utilities, and we will 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 abandon the achievements of the ordinal revolution.

Prospect theory plays a crucial role in our analysis of choice-based utility. Most empirical measurements of utility in the economic literature today still assume expected utility (Holt and Laury, 2002). Many empirical studies have, however, revealed descriptive violations of expected utility (Starmer, 2000). Descriptive improvements have been developed, such as prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Our analysis will first show, in agreement with previous findings (Hershey and Schoemaker, 1985), that the measurement of choice-based utility leads to inconsistencies under expected utility. This may explain why there have not been many detailed empirical measurements of utility, in spite of its central importance in economics (Gollier, 2001, p. 424 ff.; Gregory et al., 2002, p. 227).

We next show that the inconsistencies of choice-based utility can be resolved by means of prospect theory. This resolution corroborates similar resolutions obtained for health outcomes by Bleichrodt et al. (2001). The present paper shows that the resulting consistent choice-based utility also agrees with choiceless utility, corroborating a similar finding for health outcomes by Stalmeier and Bezembinder (1999). We argue that the obtained agreements can restore one consistent concept of utility, which combines the advantages of cardinal utility of the 19th century in being widely applicable and of ordinal utility of the 20th century in being well linked to observable choice.

Some decision theorists assumed, contrary to the ordinalists, that cardinal utility can be meaningful in riskless contexts, and studied relations between such riskless cardinal utility and risky[1] cardinal utility (Dyer and Sarin, 1982). These studies assumed expected utility and, thus, were prone to the empirical violations thereof (Starmer, 2000). In terms of this approach from decision theory, our finding suggests that the difference between risky and riskless utility disappears when data are analyzed using the, empirically preferable, prospect theory.

Another contribution of this paper concerns the introduction of a new parametric family of utility, the expo-power family. It is a one-parameter family derived from a more general two-parameter family of Saha (1993). Our family, contrary to commonly used families, allows for the simultaneous fulfillment of three economic desiderata: concave utility, decreasing absolute risk aversion, and increasing relative risk aversion.

1.1. Outline of the paper

Section 2 briefly describes the history of utility in economics up to 1950, which has been described previously by Stigler (1950), Blaug (1962, 1997), and others. Because of further developments in utility theory during recent decades, an update of this history is called for (Section 3) and this will lead to the main research question of this paper. Section 4 provides notation and definitions, and introduces the expo-power family of utility. Section 5 presents the main empirical results. First, it measures choice-based utilities using a recently introduced method, the tradeoff method. Next, choiceless cardinal utility is measured without using any choice making or risk. No significant differences are found between these two measurements of utility. A psychological explanation is given. To verify that tradeoff utilities do reflect choice making, Section 6 compares those utilities to utilities derived from a third, traditional, measurement method ("$CE_{1/3}$"). 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 to utilities derived from a fourth measurement method ("$CE_{2/3}$"). When analyzed in terms of expected utility, the utilities derived from the fourth method deviate significantly from those derived from 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 in terms 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 is obtained, leading to one utility function consistently measured in four different ways.

Section 7 contains a discussion, and Section 8 our conclusions. Appendix A gives details of our experimental method for eliciting indifferences, developed to minimize biases. Appendices B and C describe statistical tests.

2. The history of ordinal versus cardinal utility up to 1950

Table 1 shows the various concepts of utility discussed in this paper. An important step forward for the interpretation of utility was made at the beginning of the 20th century, when the views of utility changed profoundly as a result of the ordinal revolution.

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

Table 1
Various concepts of utility

	Choice-based	Choiceless
ordinal utility	– Market equilibria	
cardinal utility	– Intertemporal – Welfare – Risk ⇔	⇔ – Strength of preferences – Experienced (Kahneman)

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 concepts 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.

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 market phenomena such as 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 are not meaningful.

Alt (1936) 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) and Robbins (1932) 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, were developed in psychology (Edwards, 1954, p. 385).

von Neumann and Morgenstern (1944) raised new hope for the existence of cardinal utility by deriving cardinal utility from risky choice. After some debates, the consensus became that this risky index is cardinal in the mathematical sense of being unique up to unit and level, but not cardinal in the sense of being the neo-classical index of goodness that emerged at the end of the 19th century (Varian, 1993).[2] Ordinalism has continued to dominate in mainstream economics ever since.

[2]For recent deviating viewpoints, see Harsanyi (1978), Loomes and Sugden (1982), and Rabin (2000, footnote 3). It is worthy of note 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. Ordinal versus cardinal utility after 1950

This section describes the history of utility in the second half of the 20th century, which followed after von Neumann and Morgenstern's contribution. We are not aware of an account of this history in the literature. There are several accounts of the history up to and including the ordinal revolution (Stigler, 1950; Blaug, 1962, 1997). Yet, many developments have taken place since then, and an update is called for. We distinguish two independent developments after 1950. One took place in mainstream economics (Section 3.1), the other in decision theory (Section 3.2).

3.1. Ordinal utility in the economics literature after 1950

In the beginning of the ordinal period, promising results were obtained, including preference representations and derivations of equilibria (Samuelson, 1938b; Savage, 1954; Debreu, 1959). Soon, however, problems arose (Allais, 1953; Ellsberg, 1961; 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 and Morgenstern, 1944), but also for intertemporal evaluations (Samuelson, 1937), for utilitarian welfare evaluations (Harsanyi, 1955; Young, 1990), and for quality-of-life measurements in health (Gold et al., 1996). The consensus became that such cardinal indexes are relevant, but should be restricted to the specific domain to which they apply, and should not be equated with each other or with 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). 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, including Broome (1991), Gilboa and Schmeidler (2001), Kapteyn (1994), Loomes and Sugden (1982), Rabin (2000 footnote 3), Robson (2001 Section III.D), Tinbergen (1991), and van Praag (1968, 1991). A drawback of extending the inputs of utility is, obviously, that it makes the prediction of economic decisions difficult. The present paper presents an experimental investigation, based on prospect theory, into broader interpretations of utility, showing that they can contribute positively 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, utility differences are meaningful, and they commonly used strength-of-preference judgments to measure them (Dyer and Sarin, 1982). 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), and other psychophysical measurements (Camerer et al., 2004). In this study, we restrict our 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 (Samuelson, 1950, p. 121), and nonlinear empirical relations between risky and riskless utility have been studied (Bouyssou and Vansnick, 1988; 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 with this classical approach may have been the empirical deficiency of expected utility (Starmer, 2000). Different methods for measuring risky utility, that should yield the same utilities, exhibited systematic discrepancies (Karmarkar, 1978; Hershey and Schoemaker, 1985; McCord and de Neufville, 1986). These were as pronounced as the differences between risky and riskless utility (McCord and de Neufville, 1983, p. 295). This finding led some authors working on risky versus riskless utility to suggest abandoning the classical expected-utility approach (Krzysztofowicz and Koch, 1989; McCord and de Neufville, 1984).

Since the 1980s, several models that deviate from expected utility have been proposed (Starmer, 2000). The most popular one is prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), which agrees with rank-dependent utility (Gilboa, 1987; Quiggin, 1982; Schmeidler, 1989) on the domain considered in this paper. 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 (Bleichrodt et al., 2001; Krzysztofowicz and Koch, 1989). Previously, 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 for classical decision-theoretic studies to distinguish between risky and riskless utility disappears. It then becomes conceivable, at least as an empirical hypothesis to be tested, that the utility function of prospect theory agrees with riskless concepts. For the health domain, this hypothesis was investigated by Stalmeier and Bezembinder (1999), who compared riskless strength-of-preference judgments to risky certainty-equivalent judgments for health outcomes (the percentage of time suffering from migraine). They found that these measurements give similar results if analyzed in terms of prospect theory. The experiments of the present paper will, similarly, find no systematic differences between risky and riskless utility for money if the data are analyzed in terms of prospect theory.

4. Expected utility, prospect theory, and a new parametric family of utility

Throughout this paper, $U : \mathbb{R} \to \mathbb{R}$ denotes a utility function of money that is strictly increasing. We examine situations in which U is measurable or cardinal in a mathematical sense, i.e. U is determined up to unit and level. 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.

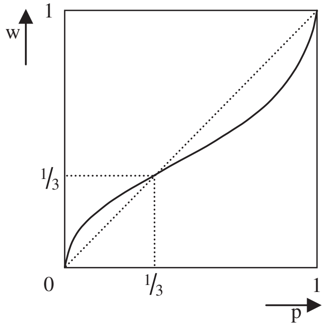

Fig. 1. The common weighting function.

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 level. Prospect theory assumes that probabilities are weighted nonlinearly, by the *probability weighting function*, denoted by 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 $x\geqslant y\geqslant 0$. EU is the special case where w is the identity. For the prospects considered in this paper, which yield only two outcomes that are both gains, original prospect theory (Kahneman and Tversky, 1979, Eq. (2)), rank-dependent utility (Quiggin, 1982), and their combination, cumulative prospect theory (Tversky and Kahneman, 1992), agree, and our conclusions apply to all these theories.

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. Fig. 1 depicts the prevailing shape (Abdellaoui, 2000; Bleichrodt and Pinto, 2000; Gonzalez and Wu, 1999; Quiggin, 1982; Tversky and Kahneman, 1992). The transformation of probability is minimal at $p=1/3$, and maximal at $p=2/3$, which is why we will use these probabilities in certainty equivalents later in the paper.

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 support Rabin's claims, and will show that prospect theory can explain these data.

For utility, two parametric families are commonly used, the power family (constant relative risk aversion, CRRA) and the exponential family (constant absolute risk aversion, CARA).[3] The prevailing empirical finding is, however, increasing relative risk aversion and decreasing absolute risk aversion (Arrow, 1971, p. 97), which is between CRRA and CARA. To accommodate this empirical pattern, we developed a new parametric family of utility. Our family was introduced in a preliminary version of this paper (Abdellaoui et al., 2000), and a variation thereof was subsequently used by Holt and Laury (2002) in their well-known study of risk attitudes.

[3]Under nonexpected utility, risk aversion cannot be equated with utility curvature, which is why we will avoid the terms CRRA and CARA.

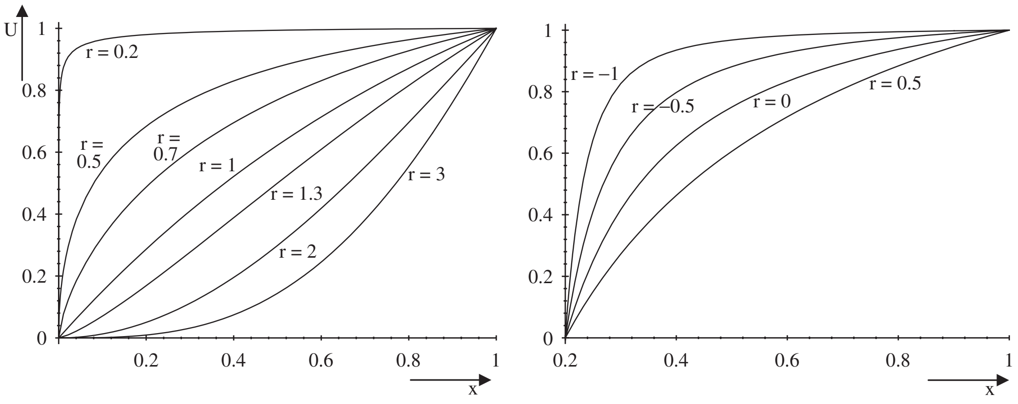

Fig. 2. Some normalized curves of the expo-power family. For $r = 1.3$, the curve is convex up to 0.40 and concave after. The right panel, with several examples of $r \leqslant 0$, excludes $x = 0$ from the domain, because there the functions are not defined for $r \leqslant 0$.

We first define the two common families $U(x)$. For each family hereafter, proper rescalings $x \mapsto a(x-b)$ $(a>0)$ of the arguments have to be considered, depending on the context, for instance if the following formulas are to be applied only to the interval [0,1].

The *power family* is defined by

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

The *exponential family* is defined by

- $1-e^{-rx}$ if $r>0$;
- x if $r = 0$;
- $e^{-rx}-1$ if $r<0$.

Our new family is the *expo-power family*, defined by

- $-\exp(-z^r/r)$ for $r \neq 0$;[4]
- $-1/z$ for $r = 0$.

Fig. 2 depicts some examples of the expo-power family, normalized at $x = 0$ and $x = 1$ for $r>0$ (left panel), and normalized at $x = 0.2$ and $x = 1$ for $r \leqslant 0$ (right panel). For $r \leqslant 0$, the expo-power function is $-\infty$ at $x = 0$, similar to the power function. The expo-power family is a variation of a two-parameter family introduced by Saha (1993). The functional form of our submember fits its name well, which is why we maintain this name. On the

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

interval [0,1], resulting from normalizations of x, 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).
- The family allows both concave ($r \leqslant 1$) and convex ($r \geqslant 2$) functions.
- There is a subclass ($0 \leqslant r \leqslant 1$) that combines some desirable features.
  (i) The functions are concave;
  (ii) The Arrow-Pratt measure of absolute risk aversion, $-u''(x)/u'(x) = (1-r)/x + x^{r-1}$, is decreasing in x;
  (iii) The measure of relative risk aversion, $-xu''(x)/u'(x) = 1 - r + x^r$, is increasing in x.

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 of utility 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, which means that it does not fit individual data well, and it does not satisfy the above conditions (i) and (iii). Bell (1988) and Farquhar and Nakamura (1987) characterized the family of all polynomial combinations of exponential functions, containing the general sumex family (linear combinations of exponential functions; Nakamura, 1996). These families have many parameters, and useful subfamilies remain to be identified.

5. An experimental comparison of choice-based and choiceless utilities

This section presents the main empirical finding of this paper, relating the, choice-based, tradeoff method of measuring utility to the, choiceless, strength-of-preference method.

5.1. 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), and 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 taken any course 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. Prospects were displayed as pie charts on a computer screen (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.

We developed software for carefully observing indifferences while avoiding biases; see Appendix A for details. 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.

5.2. Measurement methods

For the tradeoff method (TO *method*), we used "gauge outcomes" R and r with $R = \text{FF } 2000 > r = \text{FF } 1000$ (FF $1 \approx \$0.17$). An outcome t_0 was set at FF 5000. For each participant, the outcome $t_1 > t_0$ was assessed such that $(\frac{1}{3}, t_1; r) \sim (\frac{1}{3}, t_0; R)$. Next, $t_2 > t_1$ was assessed such that $(\frac{1}{3}, t_2; r) \sim (\frac{1}{3}, t_1; R)$, …, and, finally, $t_6 > t_5$ was assessed such that $(\frac{1}{3}, t_6; r) \sim (\frac{1}{3}, t_5; R)$. Under prospect theory, the indifferences imply the five equalities $U(t_6) - U(t_5) = \cdots = U(t_1) - U(t_0)$, independent of how the participant transforms probabilities (Wakker and 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:

$$U(t_i) = \frac{i}{6} \text{ for all } i. \tag{1}$$

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 of s_2 over t_1 was judged to be the same as that of t_1 over t_0, the values obtained from the TO method (see Appendix A for details). Similarly, we elicited amounts s_3, …, s_6 such that the strength of preference of s_i over s_{i-1} was judged to be the same as that of t_1 over 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. (1)), we have

$$U(s_i) = \frac{i}{6} \text{ for all } i. \tag{2}$$

5.3. 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 used two-tailed paired t-tests. To correct for individual differences, we also carried out analyses of variance with repeated measures. These analyses always gave the same conclusions as paired t-tests.

5.4. Results

The mean values of the variables t_i and s_i are depicted as the TO and SP curves in Fig. 3, which were obtained through linear interpolation. The other curves in the figure will be explained later. Table 2 in Appendix B provides numerical details. The figure suggests that the choice-based TO curve and the choiceless SP curve 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

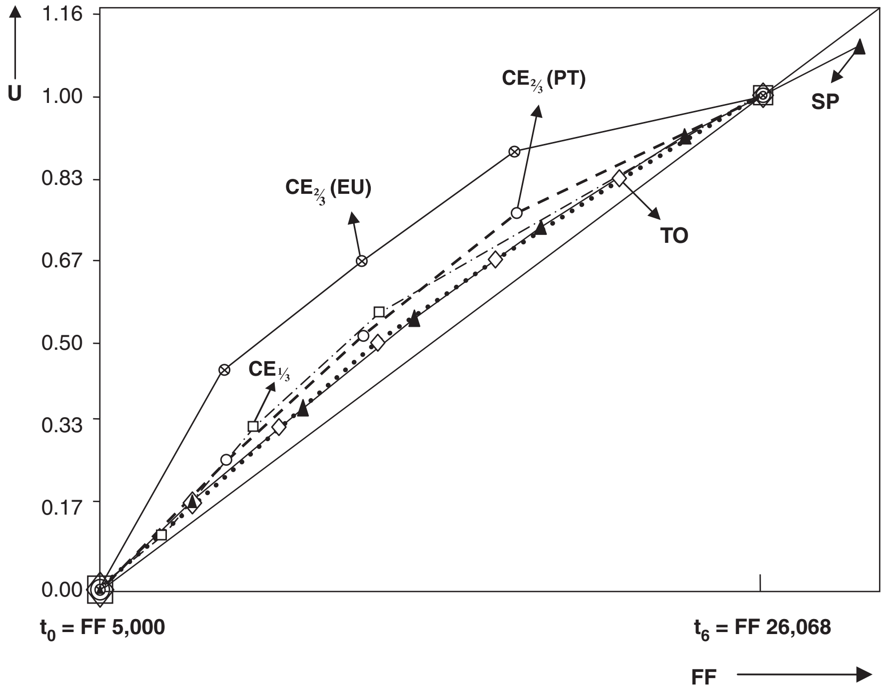

Fig. 3. All utility functions (for group averages and with linear interpolation).

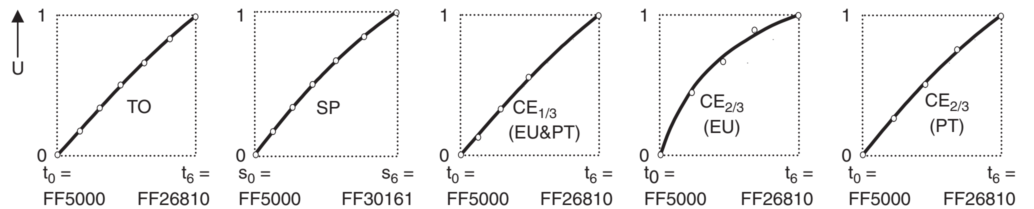

Fig. 4. Mean results for the expo-power family.

then have utility $j/6$ (Eqs. (1 and 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 left two panels of Fig. 4 and analyzed in Appendix C. The other panels in Fig. 4 will be explained later.

Linearity of the TO- and SP utility curves in Fig. 3 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, $\chi^2_5 = 38.05$, $p < .001$). Linearity was also rejected by the parametric analyses in Appendix C.

5.5. Psychological explanation for the equality of choiceless SP utilities and choice-based TO utilities

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 strength of preference in an indirect manner: In the indifference $(\frac{1}{3}, t_1; r) \sim (\frac{1}{3}, t_0; R)$, a perceived strength of preference of t_1 over t_0 (associated with probability $\frac{1}{3}$), offsets the counterargument of receiving R instead of r (associated with probability $\frac{2}{3}$). Similarly, for all i, in the indifference $(\frac{1}{3}, t_i; r) \sim (\frac{1}{3}, t_{i-1}; R)$, the perceived strength of preference of t_i over t_{i-1} (always associated with probability $\frac{1}{3}$) offsets the same counterargument of receiving R instead of r (always associated with probability $\frac{2}{3}$) as above. Hence, it is plausible that all strength of preferences of t_i over t_{i-1} are perceived as equally strong (Köbberling and Wakker, 2004). From this perspective, 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 a pessimistic interpretation of the constructive view of preference (Slovic, 1995): The participants may simply have used similar heuristics in both methods, 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. At that point we realized that further questions were feasible. Therefore, for the remaining 34 participants, in addition to TO and SP measurements, also two certainty-equivalent measurements were conducted.

Certainty-equivalent methods have a format different from 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. Our third method, the $CE_{1/3}$ *method*, considered prospects that assign probability 1/3 to the best outcome. 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 $c_3 \sim (\frac{1}{3}, t_6; c_2)$, with the t's derived from the TO measurements as described before.

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 obtain:

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

Fig. 3 in the preceding section depicts the nonparametric utility curve for $CE_{1/3}$, and the third panel in Fig. 4 gives the average result of parametric fittings. 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 C.

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, the $CE_{2/3}$ *method*, using prospects that assign probability 2/3 to the best outcome. 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 $d_3 \sim (\frac{2}{3}, t_6; d_2)$. With $U(t_0) = 0$ and $U(t_6) = 1$, the following equalities are implied:

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

The resulting average utility function is depicted as the $CE_{2/3}$(EU) curve in Fig. 3, and in the fourth panel in Fig. 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, the 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 C. 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. We assume the probability weighting function of Fig. 1 for all individuals. This assumption is obviously 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 Fig. 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 recall that the analysis of the TO method remains valid under PT, irrespective of the individual probability weighting functions. Therefore, contrary to the CE methods, the TO method is not affected by individual variations in probability weighting.

It has been found that, on average, $w(\frac{1}{3})$ is approximately $\frac{1}{3}$ (see Fig. 1 and the references there). Therefore, our analysis of $CE_{1/3}$ needs no modification and Eq. (3) and the utility function depicted in Fig. 1 remain valid under PT (Tversky and Fox, 1995, p. 276). Accordingly, the agreement between the $CE_{1/3}$ utilities and the TO utilities also remains valid. It has been found that $w(\frac{2}{3})$ is approximately .51 (Fig. 1). Hence, the analysis of $CE_{2/3}$ that was based on EU needs modification. We now find

$$\text{under PT,} \quad U(d_2) = 0.51, \quad U(d_1) = 0.26, \quad \text{and} \quad U(d_3) = 0.76 \tag{5}$$

instead of Eq. (4). The resulting corrected utility curves, denoted by $CE_{2/3}$(PT), are depicted in Figs. 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. Our finding satisfies Birnbaum and Sutton's (1992) principle of scale convergence, according to which it is desirable if different ways to measure utility give the same result.

7. Discussion

For the choice-based utilities, the statistical analyses suggested that the TO, $CE_{1/3}$, 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 $CE_{2/3}$(EU) 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, thereby supporting the results of the TO measurements. It entails a positive result within the revealed-preference paradigm. The further agreement of this index with the SP index extends beyond revealed preference, and is the new contribution of this paper.

It would have been desirable to test hypotheses at the individual level and not just at the level of group averages, as we did. In particular, it would then be desirable to adopt an error theory for individual choice. Although several such theories have been considered recently, there is no agreement on which is the best theory (Loomes et al., 2002). Data about individual choice under risk is usually very noisy. In our sample, analyses at the individual level did not give significant results and have not been reported. Considerably larger datasets are required to test hypotheses at the individual level.

It is common to implement real incentives in experiments for moderate amounts of money. Utility measurement is, however, of interest only for significant amounts of money, because utility is close to linear for moderate amounts so that no measurement is needed there anyway (Rabin, 2000; Savage, 1954, p. 60). Hence, we had to use significant amounts and could not implement real incentives. Camerer and Hogarth (1999) and Hertwig and Ortmann (2001) contain surveys of hypothetical choice versus real incentives. For simple choices that are not cognitively demanding, as in our experiment, and where there are no direct contrast effects between real and hypothetical choice as in Holt and Laury (2002), no big differences are found. The utility function for money is central in economics and, hence, its experimental measurement deserves investigation (Stigler 1950, Section IV.c) even if a resort to hypothetical choice cannot be avoided.

Rabin (2000) argued on theoretical grounds that utility is more linear than commonly thought, and that much of the commonly observed risk aversion is due to factors other than utility curvature. He suggested factors put forward by prospect theory. Our study confirms his suggestion empirically.

In applied domains, e.g. in health economics, it is common practice, based on pragmatic grounds, to use utilities measured in one context, possibly choiceless, for applications in other contexts (Gold et al., 1996). Empirical relations between various utilities, including choiceless, have been studied extensively (Pennings and Smidts, 2000; Revicki and Kaplan, 1993; Stalmeier and Bezembinder, 1999; Young, 1990). Our contribution is to establish relations between choiceless concepts and the revealed-preference paradigm of economics, reinforcing the usefulness of both.

8. Conclusion

In the classical economic debate between cardinalists and ordinalists, the latter defined economics as the study of revealed choice and assumed that direct judgments, having no preference basis, are not meaningful for economics. In light of today's advances in experimental methods in economics, the relations between direct judgments and preferences can be investigated empirically. Using prospect 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 the data resulting from direct judgments are useful and meaningful for economics.

The use of direct judgments will be particularly called for in contexts where preferences are hard to measure due to choice anomalies. More and more contexts of this kind were discovered in the 20th century (Kahneman, 1994). With relations between choice-based and choiceless data firmly established, 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. 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.

If relations between utility can be established across different decision contexts (intertemporal, welfare, etc.), then one consistent cardinal index of utility may result that is applicable to many domains, in the spirit of Broome's (1991) index of goodness. We, therefore, hope to see further empirical investigations of the relations between direct judgments and revealed preferences.

Acknowledgement

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

Appendix A. A two-step procedure for eliciting indifferences

This appendix describes our new two-step procedure for eliciting reliable indifferences. For the TO method, a value $x = t_1$ was to be found such that $A = (1/3,5000; 2000) \sim (1/3, x; 1000) = B$ (see Fig. 5, with $x_1 = 11,000$, and prospects called propositions). The first step established an interval containing t_1. We started with $x = 5000$ (instead of 11,000 in

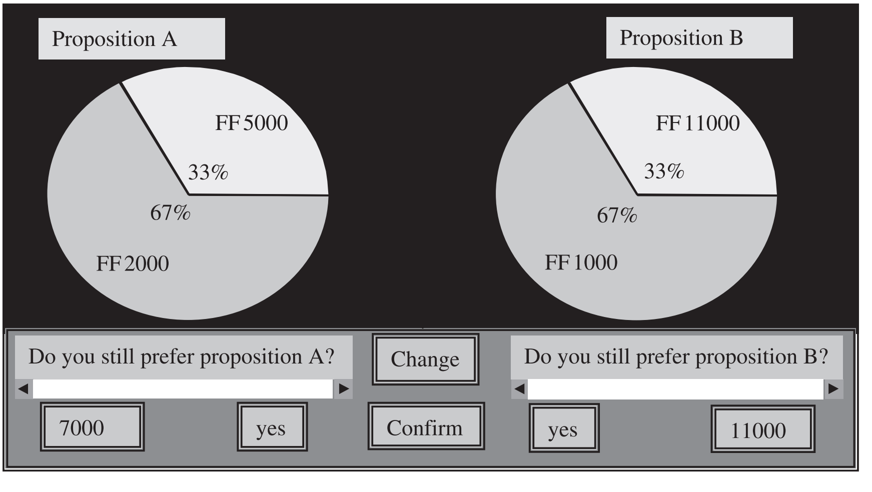

Fig. 5. A screen used in the first step.

Fig. 5), clearly a lowerbound for t_1 because of dominance of A over B. Using a scrollbar, the experimenter next increased x to 25,000, where all participants preferred the right prospect B so that $x = 25,000$ is an upper bound for t_1 for all participants. These questions, yielding a preliminary interval [5000, 25,000] containing t_1, served to familiarize the subjects with the stimuli. We next narrowed the interval down, as follows.

The scrollbar was again placed at $x = 5000$, where A dominates B, and was increased until the participant was no longer sure to prefer A. A smaller x was subsequently found for which the participant was still sure to prefer A, say $x = \alpha > 5000$. Similarly, an x was found with sure preference for B, say $x = \beta < 25,000$. Obviously, $\beta > \alpha$ should hold; if not, the participant did not understand the procedure and it was repeated. Thus, an interval $[\alpha, \beta]$ containing t_1 was obtained. We wanted this interval to be of length 4000 for all participants. Hence, we asked participants to be more precise if $[\alpha, \beta]$ was too long. Commonly it was shorter, in which case the computer automatically enlarged it. Thus, an interval of length 4000 was obtained. Fig. 5 displays the final result of Step 1 for a participant with $[\alpha, \beta] = [7000, 11000]$ as the interval that contains t_1.

In Step 2, the indifference value $x = t_1 \in [\alpha,\beta]$ was found using a choice-based bisection (see Fig. 6). With the midpoint $(\alpha + \beta)/2$ (9000 in Fig. 6) substituted for x, the participant chose 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. A new interval containing t_1 resulted in this manner, being half as large as the preceding one. After five similar iterations, the interval was sufficiently narrow and its midpoint was taken as t_1. We repeated the choice of the third iteration so as to test for consistency. It was virtually always consistent ($\geqslant 92\%$ for each measurement).

The indifference values $t_2, \ldots, t_6$ were elicited similarly. Similar procedures were also used for the CE and strength-of-preference measurements (Figs. 7 and 8).

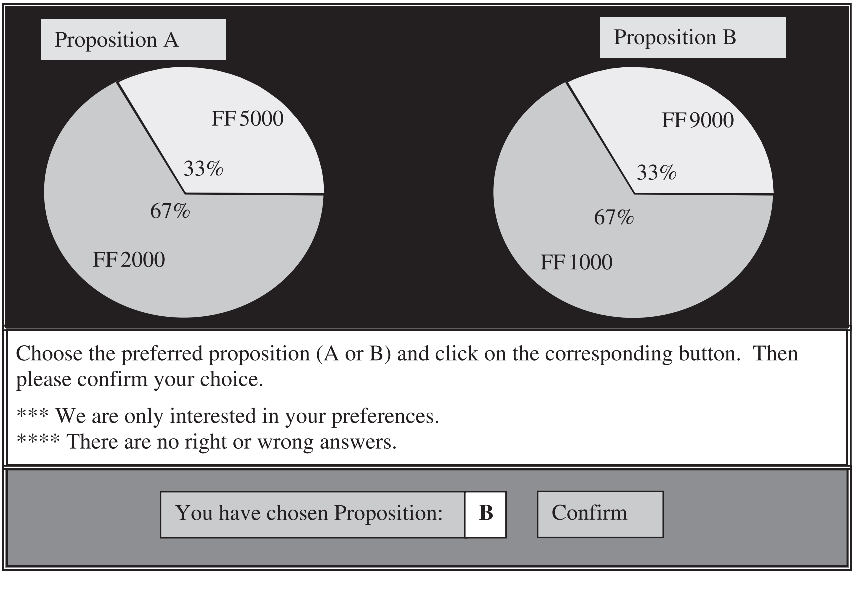

Fig. 6. Presentation of the prospects in the second step.

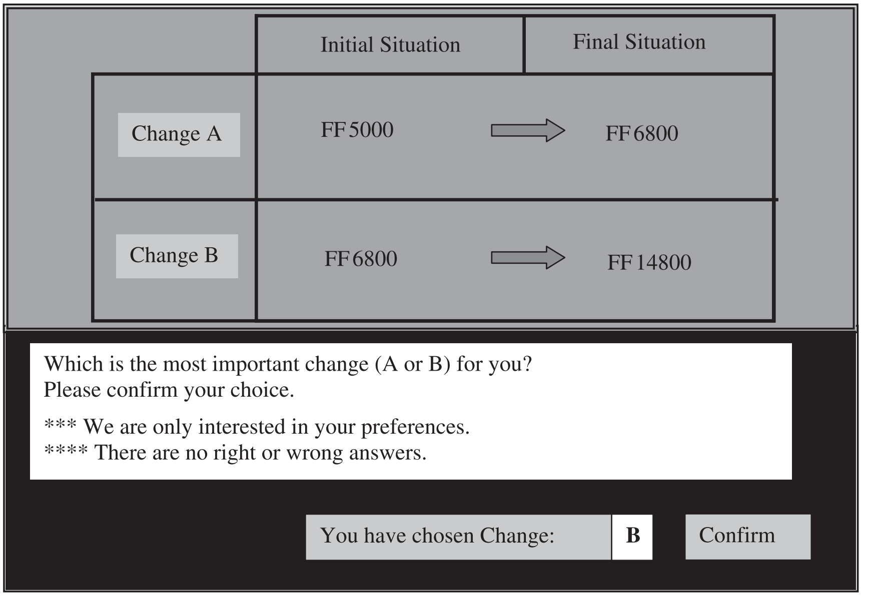

Fig. 7. First presentation of strength of preference questions.

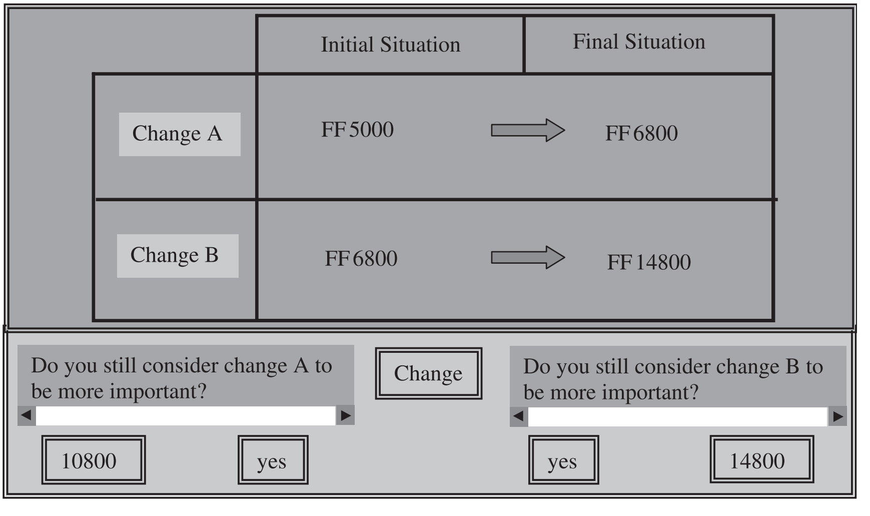

Fig. 8. Second presentation of strength of preference questions.

Appendix B. Statistical analysis of raw data

Table 2 gives descriptive statistics of our measurements. Paired t-tests of TO versus SP are in the main text.

Table 2
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)		

Table 3
Direct tests of the consistency of choice-based methods

Theory	CEs	Utility	TOs[a]	t_{33}	p-value
EU & PT	c_1	1/9	$\frac{2}{3}t_1 + \frac{1}{3}t_0$	0.09	.928
EU and PT	c_2	1/3	t_2	−1.49	.146
EU and 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	$.58t_2 + .42t_1$	−1.45	.158
PT	d_2	0.51	$.08t_4 + .92t_3$	−1.19	.244
PT	d_3	0.76	$.58t_5 + .42t_4$	−1.78	.084

[a]Interpolated t_i's.

To compare TO to $CE_{1/3}$, note that $c_2 = t_2$ under H_0 because then $U(c_2) = U(t_2) = \frac{1}{3}$ (Eqs. (1) and (3)). Because other c- and t-values concern different points in the domain, we use linear interpolation on the scale with most observations, i.e. the TO scale, to make comparisons. For example, if $U(t_0) = 0$ and $U(t_1) = 1/6$ (Eq. (1)) then, by linear interpolation, $U(\frac{2}{3}t_1 + \frac{1}{3}t_0) \approx 1/9$ and $\frac{2}{3}t_1 + \frac{1}{3}t_0$ can be compared to c_1 (Eq. (3)). Similarly, $\frac{1}{3}t_4 + \frac{2}{3}t_3$ can be compared to c_3 (fourth row of Table 3). The table indicates that no equality of c-values and corresponding (interpolations of) t-values is rejected statistically.

To compare TO to $CE_{2/3}$, note that $d_2 = t_4$ under H_0, because then $U(d_2) = U(t_4) = \frac{2}{3}$ (Eqs. (1) and (4)). Further comparisons require linear interpolations, indicated in Table 3. All equalities between TO- and $CE_{2/3}$-values, predicted by EU, are strongly rejected. A reanalysis through PT, with adapted linear interpolations indicated in the table, re-establishes the equality of utility.

All tests in this appendix confirm the conclusions based on the analyses of variance with repeated measures that were 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 mostly 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 aggravated because the different CE measurements focus on different parts of the domain.

The pairwise comparisons of the different points in Table 3 are not independent because the measurements are chained. Biases in measurements may propagate. This may explain why all five s_j values in Table 2 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 in the SP measurements, due to their role as anchor outcomes. While distorting the s_j's upwards, this bias hardly distorts the elicited utility curvature. For the latter, it is not the values of s_j or t_j per se that are essential, 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 appendix.

Appendix C. Further statistical analyses of parametric estimations

We fitted power utility, exponential utility, expo-power utility, and the sumex family (sum of two exponential functions), and used the resulting parameters in the statistical analyses. 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.

Logarithmic utility $\ln(x+b)$ is not suited to fit our data because it does not allow convex utility, whereas several participants exhibited convexities. Bear in mind that utility, when corrected for probability weighting, is less concave than traditionally thought. The CE methods have too few data points to obtain any reliable estimation for sumex utility. Equality of utility was confirmed for the TO and SP methods, but the estimations were unreliable and we do not report them.

All remaining families were normalized so as to be on the same scale, and in this way their numerical fits were compared. The arguments were transformed as x/t_6 for each subject. For power utility, any change of scale of x is immaterial because it does not affect preference. The translation $z = x - t_0$ supported our data well, but has the drawback that utility is not defined at and below $x = t_0$ for negative powers, which is why we do not report its results. For exponential utility, a translation $z = x - t_0$ is immaterial because it does not affect preference, and rescalings of x only amount to rescalings of the parameter r. For expo-power utility the rescaling x/x_6 leads to the desirable properties described in the main text. We did not consider the translation $x - x_0$ for the same reason as with power utility, but also so as to maintain increasing relative risk aversion.

Table 4 gives descriptive statistics for individual parametric estimates. Fig. 4 in the main text depicts the optimal parametric fittings of the expo-power family for group averages. 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 $t_1/t_6, \ldots, t_5/t_6$ for TO, $s_1/t_6, \ldots,$ 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.

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 linear utility 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

Table 4
Descriptive statistics of individual parametric estimates

	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.

Table 5
Results of paired *t*-tests

	Parametric families					
	Power		Exponential[a]		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**

[a]Participant 44 was excluded because the parameters of the exponential family did not converge for $CE_{2/3}$.

carried out for this family. Table 5 presents the results of tests of equalities of utility parameters.

The conclusions are the same for all families and agree with the conclusions in the main text. The $CE_{2/3}$ measurements, when analyzed using EU, differ significantly from all the other measurements. Those other measurements, including the $CE_{2/3}$ measurements when analyzed using PT, agree mutually. 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, M., 2000. Parameter-free elicitation of utilities and probability weighting functions. Management Science 46, 1497–1512.

Abdellaoui, M., Barrios, C., Wakker, P.P., 2000. Reconciling introspective utility with revealed preference: experimental arguments based on prospect theory. Presented at ESA Conference, Department of Economics, University of Amsterdam, The Netherlands, 20 October.

Allais, M., 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.

Alt, F., 1936. Über die Messbarkeit des Nutzens. Zeitschrift für Nationalökonomie 7, 161–169.

Arrow, K.J., 1971. Essays in the theory of risk bearing. North-Holland, Amsterdam.

Bell, D.E., 1988. One-switch utility functions and a measure of risk. Management Science 34, 1416–1424.

Birnbaum, M.H., Sutton, S.E., 1992. Scale convergence and utility measurement. Organizational Behavior and Human Decision Processes 52, 183–215.

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

Bleichrodt, H., Pinto, J.L., 2000. A parameter-free elicitation of the probability weighting function in medical decision analysis. Management Science 46, 1485–1496.

Bleichrodt, H., Pinto, J.L., Wakker, P.P., 2001. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. Management Science 47, 1498–1514.

Bouyssou, D., Vansnick, J.-C., 1988. A note on the relationships between utility and value functions. In: Munier, B.R. (Ed.), Risk, Decision and Rationality. Reidel, Dordrecht, The Netherlands, pp. 103–114.

Broome, J.R., 1991. Weighing Goods. Basil Blackwell, Oxford, UK.

Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: a review and capital–labor–production framework. Journal of Risk and Uncertainty 19, 7–42.

Camerer, C.F., Loewenstein, G.F., Prelec, D., 2004. Neuroeconomics: why economics needs brains. Scandinavian Journal of Economics 106, 555–579.

Debreu, G., 1959. Theory of Value. An Axiomatic Analysis of Economic Equilibrium. Wiley, New York.

Dyer, J.S., Sarin, R.K., 1982. Relative risk aversion. Management Science 28, 875–886.

Edwards, W., 1954. The theory of decision making. Psychological Bulletin 51, 380–417.

Ellsberg, D., 1961. Risk, ambiguity and the Savage axioms. Quarterly Journal of Economics 75, 643–669.

Farquhar, P.H., Nakamura, Y., 1987. Constant exchange risk properties. Operations Research 35, 206–214.

Fellner, W., 1961. Distortion of subjective probabilities as a reaction to uncertainty. Quarterly Journal of Economics 75, 670–689.

Frey, B.S., Stutzer, A., 2002. What can economics learn from happiness research? Journal of Economic Literature 40, 403–435.

Gilboa, I., 1987. Expected utility with purely subjective non-additive probabilities. Journal of Mathematical Economics 16, 65–88.

Gilboa, I., Schmeidler, D., 2001. A cognitive model of individual well-being. Social Choice and Welfare 18, 269–288.

Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C., 1996. Cost-effectiveness in health and medicine. Oxford University Press, New York.

Gollier, C., 2001. The economics of risk and time. MIT Press, Cambridge, MA.

Gonzalez, R., Wu, G., 1999. On the shape of the probability weighting function. Cognitive Psychology 38, 129–166.

Gregory, A.W., Lamarche, J.-F., Smith, G.W., 2002. Information-theoretic estimation of preference parameters: macroeconomic applications and simulation evidence. Journal of Econometrics 107, 213–233.

Grether, D.M., Plott, C.R., 1979. Economic theory of choice and the preference reversal phenomenon. American Economic Review 69, 623–638.

Harsanyi, J.C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy 63, 309–321.

Harsanyi, J.C., 1978. Bayesian decision theory and utilitarian ethics. American Economic Review, Papers and Proceedings 68, 223–228.

Hershey, J.C., Schoemaker, P.J.H., 1985. Probability versus certainty equivalence methods in utility measurement: Are they equivalent? Management Science 31, 1213–1231.

Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: a challenge for psychologists? Behavioral and Brain Sciences 24, 383–403.

Hicks, J.R., Allen, R.G.D., 1934. A reconsideration of the theory of value: I; II. Economica 1, 52–75, 196–219.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–1655.

Kahneman, D., 1994. New challenges to the rationality assumption. Journal of Institutional and Theoretical Economics 150, 18–36.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.
Kahneman, D., Tversky, A. (Eds.), 2000. Choices, Values, and Frames. Cambridge University Press, New York.
Kapteyn, A., 1994. The measurement of household cost functions: revealed preference versus subjective measures. Journal of Population Economics 7, 333–350.
Karmarkar, U.S., 1978. Subjectively weighted utility: a descriptive extension of the expected utility model. Organizational Behavior and Human Performance 21, 61–72.
Köbberling, V., Wakker, P.P., 2004. A simple tool for qualitatively testing, quantitatively measuring, and normatively justifying Savage's subjective expected utility. Journal of Risk and Uncertainty 28, 135–145.
Krzysztofowicz, R., Koch, J.B., 1989. Estimation of cardinal utility based on a nonlinear theory. Annals of Operations Research 19, 181–204.
Lichtenstein, S., Slovic, P., 1971. Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology 89, 46–55.
Loomes, G., Sugden, R., 1982. Regret theory: an alternative theory of rational choice under uncertainty. Economic Journal 92, 805–824.
Loomes, G., Moffat, P.G., Sugden, R., 2002. A microeconometric test of alternative stochastic theories of risky choice. Journal of Risk and Uncertainty 24, 103–130.
Manski, C.F., 2004. Measuring expectations. Econometrica 72, 1329–1376.
McCord, M.R., de Neufville, R., 1983. Fundamental deficiency of expected utility decision analysis. In: French, S., Hartley, R., Thomas, L.C., White, D.J. (Eds.), Multi-objective decision making. Academic Press, New York, pp. 279–305.
McCord, M.R., de Neufville, R., 1984. Utility dependence on probability: an empirical demonstration. Large Scale Systems 6, 91–103.
McCord, M.R., de Neufville, R., 1986. "Lottery equivalents": reduction of the certainty effect problem in utility assessment. Management Science 32, 56–60.
Merton, R.C., 1971. Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3, 373–413.
Nakamura, Y., 1996. Sumex utility functions. Mathematical Social Sciences 31, 39–47.
Pareto, V., 1906. Manuele di Economia politica. Piccolo Biblioteca Scientifica, Milan.
Pennings, J.M.E., Smidts, A., 2000. Assessing the construct validity of risk attitude. Management Science 46, 1337–1348.
Quiggin, J., 1982. A theory of anticipated utility. Journal of Economic Behaviour and Organization 3, 323–343.
Rabin, M., 2000. Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68, 1281–1292.
Revicki, D.A., Kaplan, R.M., 1993. Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. Quality Life Research 2, 477–487.
Robbins, L., 1932. An essay on the nature and significance of economic science. MacMillan, London.
Robson, A.J., 2001. Why would nature give individuals utility functions? The Journal of Political Economy 109, 900–914.
Saha, A., 1993. Expo-power utility: a flexible form for absolute and relative aversion. American Journal of Agricultural Economics 75, 905–913.
Samuelson, P.A., 1937. A note on measurement of utility. Review of Economic Studies 4 (2), 155–161.
Samuelson, P.A., 1938a. The numerical representation of ordered classifications and the concept of utility. Review of Economic Studies 6 (1), 65–70.
Samuelson, P.A., 1938b. A note on the pure theory of consumer's behaviour. Economica, N.S 5, 61–71, 353–354.
Samuelson, P.A., 1950. The problem of integrability in utility theory. Economica, N.S 17, 355–385.
Savage, L.J., 1954. The Foundations of Statistics. Wiley, New York (Second edition 1972, Dover Publications, New York).
Schmeidler, D., 1989. Subjective probability and expected utility without additivity. Econometrica 57, 571–587.
Simon, H.A., 1955. A behavioral model of rational choice. Quarterly Journal of Economics 69, 99–118.
Slovic, P., 1995. The construction of preference. American Psychologist 50, 364–371.
Stalmeier, P.F.M., Bezembinder, T.G.G., 1999. The discrepancy between risky and riskless utilities: a matter of framing? Medical Decision Making 19, 435–447.
Starmer, C., 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, G.J., 1950. The development of utility theory: I; II. Journal of Political Economy 58, 307–327, 373–396.

Tinbergen, J., 1991. On the measurement of welfare. Journal of Econometrics 50, 7–13.

Tversky, A., Fox, C.R., 1995. Weighing risk and uncertainty. Psychological Review 102, 269–283.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

van Praag, B.M.S., 1968. Individual Welfare Functions and Consumer Behavior. North-Holland, Amsterdam.

Van Praag, B.M.S., 1991. Ordinal and cardinal utility: an integration of the two dimensions of the welfare concept. Journal of Econometrics 50, 69–89.

van Praag, B.M.S., Ferrer-i-Carbonell, A., 2004. Happiness Quantified. Oxford University Press, Oxford, UK.

Varian, H.R., 1993. Intermediate Microeconomics. Norton, New York.

von Neumann, J., Morgenstern, O., 1944 (1947, 1953). Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ.

Wakker, P.P., Deneffe, D., 1996. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. Management Science 42, 1131–1150.

Young, H.P., 1990. Progressive taxation and equal sacrifice. American Economic Review 80, 253–266.