Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Tailor-Made Test of Intransitive Choice

Aurélien Baillon, Han Bleichrodt, Alessandra Cillo

OPERATIONS RESEARCH
Vol. 63, No. 1, January–February 2015, pp. 198–211
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.2014.1329

METHODS

A Tailor-Made Test of Intransitive Choice

Aurélien Baillon, Han Bleichrodt
Erasmus School of Economics, PO Box 1738, 3000 DR Rotterdam, The Netherlands {baillon@ese.eur.nl, bleichrodt@ese.eur.nl}

Alessandra Cillo
Bocconi University and IGIER, 20136 Milan, Italy, alessandra.cillo@unibocconi.it

This paper reports a new test of intransitive choice using individual measurements of regret- and similarity-based intransitive models of choice under uncertainty. Our test is tailor-made and uses subject-specific stimuli. Despite these features, we observed only a few intransitivities. A possible explanation for the poor predictive performance of intransitive choice models is that they only allow for interactions between acts. They exclude within-act interactions by retaining the assumption that preferences are separable over states of nature. Prospect theory, which relaxes separability but retains transitivity, predicted choices better. Our data suggest that descriptively realistic models must allow for within-act interactions but may retain transitivity.

Subject classifications: utility/preference: estimation; decision analysis: risk.
Area of review: Decision Analysis.
History: Received August 2013; revision received May 2014; accepted September 2014. Published online in *Articles in Advance* December 22, 2014.

1. Introduction

Transitivity is a fundamental axiom of rational choice. It underlies most theories of decision making and is commonly assumed in applied decision analysis. There is broad agreement that transitivity is normative, but its empirical status is less clear. Starting with May (1954) and Tversky (1969), many studies have observed systematic and substantial violations of transitivity, suggesting that transitivity does not describe people's preferences well (Brandstätter et al. 2006, González-Vallejo 2002, Loomes et al. 1991). However, these violations are controversial. Regenwetter et al. (2011a) showed that these studies suffered from methodological limitations and that the actual rate of transitivity violations was within statistical error and, therefore, not sufficiently convincing to abandon transitive theories.[1]

All previous tests of transitivity faced the problem of choosing the right stimuli. Typically, stimuli were selected somewhat haphazardly, based on intuitive reasoning or on some hypothesized parameterization of models of intransitive choice. All subjects were then confronted with the same stimuli. An obvious drawback of this "one-size-fits-all approach" is that it is somewhat blunt. Subjects may be intransitive, but the selected parameterization may hit the critical range for only a minority of subjects. This approach does not account for the extensive heterogeneity in preferences that is usually observed in empirical studies.

An alternative, "tailor-made approach," which we adopt in this article, is to select a model of intransitive choice, measure it for each individual separately, and then use these measurements to select the individual-specific stimuli that will produce intransitive cycles according to the model. It is widely believed that real-valued utility functions representing choices require transitivity and, consequently, that the possibility of intransitive choice excludes the existence of such functions. We show that intransitive models can be measured even at the individual level and without simplifying parametric assumptions. This measurement allows us to perform the first tailor-made tests of intransitive choices.

Two important classes of intransitive choice models are models based on regret and models based on similarity judgments.[2] Examples of regret models include Bell (1982, 1983), Loomes and Sugden (1982, 1987), Fishburn's (1982) skew-symmetric bilinear (SSB) theory and, more recently, the random regret minimization model (Chorus 2012), which is used in transport modeling.

Examples of similarity models include Leland (1994, 1998), Mellers and Biagini (1994), and Rubinstein (1988). The intuition underlying these models is that subjects pay less attention to dimensions that are similar and give weight instead to dissimilar ones. A limitation of these models is that they treat similarity judgments as dichotomous: there is a threshold (not clearly specified) above which subjects take stimuli into account and below which stimuli become inconsequential. Loomes (2010) proposed the perceived relative argument model, a more general model that allows continuous rather than dichotomous similarity judgments. A related model is González-Vallejo's (2002) proportional difference model, which extends the deterministic similarity models by adding a stochastic term reflecting decision error.

We show how regret- and similarity-based intransitive models can be measured, and we apply these measurements to derive subject-specific tests of intransitivity. Despite using subject-specific tests, we found very little evidence of intransitive cycles, and we could reject the predictions of the intransitive models.

Our subjects deviated from expected utility, the model that is traditionally used in decision analysis. There are two approaches to explain deviations from expected utility. The first approach, embodied by the intransitive choice models, explains these deviations through interactions between acts. The second approach, which contains prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992), excludes such between-act interactions but permits within-act interactions through the ranking of the outcomes of an act. We also measured and tested prospect theory. Prospect theory predicted our choice data better than the intransitive theories. We conclude that descriptively realistic models of choice should allow for within-act interactions. Whether between-act interactions also play a role remains an open question, but we found no evidence that transitivity should be abandoned.

2. Theory

2.1. A General Intransitive Additive Model

Consider a decision maker who faces uncertainty, modeled through a set $\mathcal{S}$ of possible *states of nature*. Subsets of $\mathcal{S}$ are *events*. P is a probability measure defined over events. We will write $(p_1, x_1; \ldots; p_n, x_n)$ if there are events E_j that obtain with probability p_j such that the decision maker receives money amount x_j if E_j obtains, and the events E_j partition the state space. The decision maker chooses between *acts* $(p_1, x_1; \ldots; p_n, x_n)$ and $(p_1, y_1; \ldots; p_n, y_n)$, where we implicitly assume that p_j in $(p_1, x_1; \ldots; p_n, x_n)$ and p_j in $(p_1, y_1; \ldots; p_n, y_n)$ refer to the same event E_j, $j = 1, \ldots, n$.

Let $\succeq$ denote the decision maker's preference relation over acts. As usual, $\succ$ and $\sim$ denote strict preference and indifference. We assume that preferences between acts $(p_1, x_1; \ldots; p_n, x_n)$ and $(p_1, y_1; \ldots; p_n, y_n)$ are represented by an intransitive additive model (Bouyssou 1986; Fishburn 1990, 1991):

$$(p_1, x_1; \ldots; p_n, x_n) \succeq (p_1, y_1; \ldots; p_n, y_n)$$
$$\Leftrightarrow \quad \sum_{j=1}^{n} p_j \varphi(x_j, y_j) \geqslant 0. \tag{1}$$

The function φ in Equation (1) is real-valued, strictly increasing in its first argument and strictly decreasing in its second argument, and it satisfies *symmetry*: for all money amounts x_i, y_i, $-\varphi(x_i, y_i) = \varphi(y_i, x_i)$. Symmetry implies $\varphi(0, 0) = 0$. The function φ is a ratio scale, unique up to the unit of measurement.

Equation (1) captures the key property of models of intransitive choice, that there exist interactions between acts. Equation (1) can account for intransitive cycles if φ is either *convex* for all $x_i \succeq y_i \succeq z_i$, $\varphi(x_i, z_i) \geqslant \varphi(x_i, y_i) + \varphi(y_i, z_i)$, or *concave* for all $x_i \succeq y_i \succeq z_i$, $\varphi(x_i, z_i) \leqslant \varphi(x_i, y_i) + \varphi(y_i, z_i)$. To explain the common deviations from expected utility, φ should be convex (Loomes and Sugden 1987).

Equation (1) includes several models as special cases. For instance, if $\varphi(x_i, y_i) = u(x_i) - u(y_i)$, then expected utility results. Of course, in that case, intransitivities cannot occur. Another special case of Equation (1) is regret theory (Bell 1982, Loomes and Sugden 1982), where $\varphi(x_i, y_i) = Q(u(x_i) - u(y_i))$. The strictly increasing *utility function* u is an interval scale, unique up to scale and unit, and the strictly increasing function Q, which reflects the influence of regret, is unique up to unit. The convexity of φ implies convexity of Q, which is called *regret aversion*. Other special cases are Fishburn's (1982) SSB model and the general regret model of Loomes and Sugden (1987). As Fishburn (1992) showed, Tversky's (1969) additive difference model results from Equation (1) if an interstate uniformity axiom is added. Equation (1) does not account for intransitive models that result from lexicographic orders, such as the priority heuristic (Brandstätter et al. 2006). In these models, attributes are considered sequentially, and this process may end before all attributes have been considered. Equation (1) always considers all payoffs and all probabilities.

2.2. The Perceived Relative Argument Model

Equation (1) only permits between-act interactions on the payoff dimension. It takes probabilities as they stand and therefore excludes between-act interactions on the probability dimension. Loomes (2010) proposed a model, the perceived relative argument model (PRAM), that permits between-act interactions on both the probability and the payoff dimension.

Because we only use acts with at most three different states of nature in our experiment, we will explain PRAM for such acts. Let $X = (p_1, x_1; p_2, x_2; p_3, x_3)$ and $Y = (p_1, y_1; p_2, y_2; p_3, y_3)$ be any two acts. According to PRAM,

$$X \succeq Y \quad \Leftrightarrow \quad \psi(b_X, b_Y) \geqslant \xi(u_Y, u_X). \tag{2}$$

In Equation (2), ψ reflects the perceived argument for X versus Y on the probability dimension, and ξ reflects the perceived argument for Y versus X on the payoff dimension. The term b_X equals the sum of the probabilities of the states in which X gives a strictly better outcome than Y, and the term b_Y equals the sum of the probabilities of the states in which Y gives a strictly better outcome than X. For example, if $x_1 > y_1$, $x_2 < y_2$, and $x_3 = y_3$—i.e., X gives a better outcome than Y in the first state, Y gives a better outcome than X in the second state, and X and Y give the same outcome in the third state—then $b_X = p_1$ and $b_Y = p_2$.

Loomes (2010) assumed that

$$\psi(b_X, b_Y) = (b_X / b_Y)^{(b_X + b_Y)^{\alpha}}. \tag{3}$$

In Equation (3), α is a person-specific variable, and its value may vary from one individual to another to reflect differences in perception. If $\alpha = 0$, then the individual takes probability ratios as they are, as in expected utility. To capture the common violations of expected utility, α should be negative. More negative values of α imply that probability ratios become less important when the absolute differences between the probabilities become smaller. For example, a negative α means that if both b_X and b_Y are scaled down, and their absolute difference decreases (they become more similar), they carry less weight. A comparable idea underlies the similarity models of Leland (1994, 1998), Mellers and Biagini (1994), and Rubinstein (1988).[3] In these models, similarity is dichotomous: above some unspecified threshold, two stimuli are considered dissimilar, but below it they become so similar that the difference between them is ignored (α goes to $-\infty$ in Equation (3)). PRAM allows more diverse applications of similarity theories by making the similar/dissimilar judgment continuous.

Loomes (2010) further assumed that there exists a real-valued utility function u defined over the set of outcomes.[4] Let u_Y denote in utility terms the advantage that Y has over X, and let u_X denote the advantage that X has over Y. For example, if $u(x_1) - u(y_1) > 0$ and $u(x_2) - u(y_2) = u(x_3) - u(y_3) < 0$, then $u_Y = u(y_2) - u(x_2) = u(y_3) - u(x_3)$ and $u_X = u(x_1) - u(y_1)$. Loomes (2010) assumed that

$$\xi(u_Y, u_X) = (u_Y/u_X)^{\delta}, \quad \text{where } \delta \geqslant 1. \tag{4}$$

Expected utility is the special case of PRAM where $\alpha = 0$ and $\delta = 1$. If $\delta > 1$, then whichever is the larger of u_Y and u_X receives disproportionate attention, and this disproportionality increases as u_Y and u_X become more and more different. In §3, we show that PRAM predicts intransitive cycles when $\delta > 1$.

3. Predicting Intransitivities

In this section, we introduce our tailor-made approach in a deterministic context, i.e., using algebraic models and assuming that subjects' choices reveal preferences that are perfectly represented by the algebraic models. In the next section, we will show how, in the experimental implementation and analysis, we incorporated probabilistic components to take into account that people's choices may randomly deviate from the deterministic models.

3.1. First Part: Measurement of a Standard Sequence of Money Amounts

Our procedure for predicting intransitivities consists of three parts. The first part uses the trade-off method of Wakker and Deneffe (1996) to elicit for each subject a *standard sequence* of money amounts $x_0, x_{i,1}, \ldots, x_{i,5}$. The subscript i expresses that the elements of a subject's standard sequence (except the first) depend on the subject's answers and that the standard sequences are different for each subject. This statement follows because in our method, φ in Equation (1) and δ in Equation (4) are allowed to be individual specific, which we also express through a subscript i below.

Two gauge outcomes R and r ($R > r > 0$), a probability p, and a starting outcome x_0 were selected, and we elicited $x_{i,1}$ such that $(p, x_{i,1}; 1-p, r) \sim (p, x_0; 1-p, R)$. The details of the elicitation procedure are provided in the next section. According to Equation (1), this indifference implies

$$\varphi_i(x_{i,1}, x_0) = \frac{1-p}{p}\varphi_i(R, r). \tag{5}$$

We then elicited $x_{i,2}$ such that $(p, x_{i,2}; 1-p, r) \sim (p, x_{i,1}; 1-p, R)$, which gives

$$\varphi_i(x_{i,2}, x_{i,1}) = \frac{1-p}{p}\varphi_i(R, r). \tag{6}$$

It follows from Equations (5) and (6) that $\varphi_i(x_{i,2}, x_{i,1}) = \varphi_i(x_{i,1}, x_0)$. Because φ_i is increasing in its first argument and decreasing in its second, this equality implies that in terms of φ_i, the distance between $x_{i,2}$ and $x_{i,1}$ is equal to the distance between $x_{i,1}$ and x_0. Continuing this procedure, we elicited indifferences $(p, x_{i,j+1}; 1-p, r) \sim (p, x_{i,j}; 1-p, R)$, $j = 0, \ldots, 4$, and thus obtained a standard sequence for which successive elements are equally spaced in terms of φ_i. Because φ_i is a ratio scale, we can set $\varphi_i(x_{i,1}, x_0) = 1$. Under regret theory and the additive difference model, we obtain that $u_i(x_{i,j}) - u_i(x_{i,j-1}) = u_i(x_{i,1}) - u_i(x_{i,0})$, and successive elements of the standard sequence are equally spaced in terms of utility. Because u_i is an interval scale, we can set $u_i(x_0) = 0$ and $u_i(x_{i,5}) = 1$. Then $u_i(x_{i,j}) = j/5$, $j = 0, \ldots, 5$.

Under PRAM, the indifferences $(p, x_{i,j+1}; 1-p, r) \sim (p, x_{i,j}; 1-p, R)$, $j = 0, \ldots, 4$, imply

$$\left(\frac{u_i(x_{i,j+1}) - u_i(x_{i,j})}{u(R) - u(r)}\right)^{\delta_i} = \frac{1-p}{p}. \tag{7}$$

Consequently, $u_i(x_{i,j+1}) - u_i(x_{i,j}) = u_i(x_{i,1}) - u_i(x_0)$, $j = 1, \ldots, 4$. If we scale u_i such that $u_i(x_0) = 0$ and $u_i(x_{i,5}) = 1$, then $u_i(x_{i,j}) = j/5$.

3.2. Second Part: Measurement of φ_i and δ_i

In the second part, we selected for each subject outcomes x_0, $x_{i,3}$, and $x_{i,4}$ from the individual's standard sequence, and we elicited $z_{i,p}$ such that $(p, x_{i,4}; 1-p, x_0) \sim (p, x_{i,3}; 1-p, z_{i,p})$. The second part provided information about the individual functions φ_i in Equation (1) (and consequently about the functions Q_i in regret theory) and the parameters δ_i in PRAM. Under Equation (1) and the chosen scaling $\varphi_i(x_{i,1}, x_0) = \varphi_i(x_{i,j}, x_{i,j-1})$, the indifference $(p, x_{i,4}; 1-p, x_0) \sim (p, x_{i,3}; 1-p, z_{i,p})$ implies

$$\varphi_i(z_{i,p}, x_0) = \frac{p}{1-p}. \tag{8}$$

Equation (8) defines φ_i as a function of $z_{i,p}$. By varying p, we can measure as many points of φ_i as desired.[5]

The second part also allows measurement of the parameter δ_i in PRAM. According to Equations (2)–(4), the indifference $(p, x_{i,4}; 1-p, x_0) \sim (p, x_{i,3}; 1-p, z_{i,p})$ implies

$$\left(\frac{u_i(z_{i,p}) - u_i(x_0)}{u_i(x_{i,4}) - u_i(x_{i,3})}\right)^{\delta_i} = (5u_i(z_{i,p}))^{\delta_i} = \frac{p}{1-p}. \quad (9)$$

From Equation (9), we can compute δ_i once we know $u_i(z_{i,p})$. This value is generally unknown but could be estimated using the utility function measured in the first part.

3.3. Third Part: Triples to Test Intransitivity

The elements of the standard sequence were used in the third part to create triples to test intransitivity. The triples were tailor-made because we used for each subject the individual's own standard sequence. For example, we presented the following three choices:

(i) $A = (1/3, x_{i,4}; 1/3, x_{i,4}; 1/3, x_{i,2})$ versus $B = (1/3, x_{i,2}; 1/3, x_{i,5}; 1/3, x_{i,3})$.

(ii) B versus $C = (1/3, x_{i,3}; 1/3, x_{i,3}; 1/3, x_{i,4})$.

(iii) C versus A.

According to Equation (1), the comparison between A and B gives

$$A \succeq B \quad \Leftrightarrow \quad \tfrac{1}{3}\varphi_i(x_{i,4}, x_{i,2}) + \tfrac{1}{3}\varphi_i(x_{i,4}, x_{i,5}) + \tfrac{1}{3}\varphi_i(x_{i,2}, x_{i,3}) \geqslant 0. \quad (10)$$

By the symmetry of φ_i, the chosen scaling, and the properties of the standard sequence, $\varphi_i(x_{i,4}, x_{i,5}) = \varphi_i(x_{i,2}, x_{i,3}) = -\varphi_i(x_{i,1}, x_0)$. Hence, Equation (10) can be written as

$$A \succeq B \quad \Leftrightarrow \quad \varphi_i(x_{i,4}, x_{i,2}) - 2 \geqslant 0. \quad (11)$$

In other words, the decision maker will prefer A over B if (and only if) φ_i is convex. A similar analysis shows that a decision maker with convex φ_i will prefer B over C and C over A. Hence, convex φ_i entails the cycle ABC (ABC stands for "A preferred to B, B preferred to C, and C preferred to A"). Concave φ_i implies the cycle BCA.

We illustrate the advantage of our tailor-made approach over a one-size-fits-all approach using a simple example. Consider a decision maker who behaves according to regret theory with mildly concave utility (a power function with coefficient 0.80) and a mildly convex function Q_i (a power function with coefficient 1.05). Suppose that the mean utility function in the population is linear. Then the best that a one-size-fits-all approach can do is to select the stimuli in the third part such that $x_{j+1} - x_j = x_1 - x_0$, $j = 1, \ldots, 4$. However, for these stimuli, the decision maker would not exhibit cycles, whereas with our tailor-made stimuli, cycles would be exhibited.

PRAM also predicts the cycle ABC. $A = (1/3, x_{i,4}; 1/3, x_{i,4}; 1/3, x_{i,2})$ gives a better outcome than $B = (1/3, x_{i,2}; 1/3, x_{i,5}; 1/3, x_{i,3})$ in the first state (which has probability 1/3), and the utility difference is $u_i(x_{i,4}) - u_i(x_{i,2}) = 2/5$. B gives a better outcome than A in the second and in the third state (with a joint probability of 2/3), and the utility difference is $u_i(x_{i,5}) - u_i(x_{i,4}) = u_i(x_{i,3}) - u_i(x_{i,2}) = 1/5$. According to Equations (2)–(4),

$$A \succeq B \quad \Leftrightarrow \quad ((2/5)/(1/5))^{\delta_i} = 2^{\delta_i} \geqslant (2/3)/(1/3) = 2. \quad (12)$$

It follows from Equation (12) that $A \succ B$ if and only if $\delta_i > 1$. PRAM with $\delta_i > 1$ also predicts that $B \succ C = (1/3, x_{i,3}; 1/3, x_{i,3}; 1/3, x_{i,4})$ (because $u_i(x_{i,4}) - u_i(x_{i,3}) = u_i(x_{i,1}) - u_i(x_0)$) and that $C \succ A$. Hence, unless $\delta_i = 1$, in which case the decision maker is indifferent between A, B, and C, PRAM predicts the intransitive cycle ABC. This prediction does not depend on the value of α_i. The parameter α_i drops out of Equation (12) because the probabilities of the states in which the outcomes differ between the two acts sum to 1, and $1^{\alpha_i} = 1$.[6]

4. Experiment

The previous section showed that violations of transitivity are closely connected with nonlinearity of the individual functions φ_i and with the individual parameters δ_i in PRAM. The aim of our experiment was to explore whether these relations could indeed be observed empirically.

4.1. Subjects

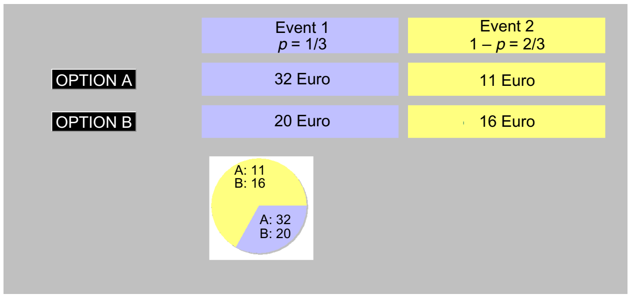
The subjects were 54 students (22 male) from Erasmus University Rotterdam, aged between 18 and 33 years (median age 21), with various academic backgrounds. They were paid a €10 participation fee. In addition, each subject had a 10% chance to play out one of the choices for real. After the experiment, subjects drew a ticket from a nontransparent bag containing 10 tickets, one of which was a winning ticket. If a subject drew the winning ticket, the computer randomly selected the choice to be played for real. The subject then played the preferred option in the selected choice, with payoffs determined by another randomly drawn number.

4.2. Procedures

The experiment was computer-run[7] and administered in sessions of two subjects with one experimenter present. Sessions lasted 55 minutes on average. The subjects were asked to make choices between pairs of acts. The indifferences in the first two parts of our method were elicited through a series of choices that "zoomed in" on subjects' indifference values. This iteration procedure is explained in the online companion (available as supplemental material at http://dx.doi.org/10.1287/opre.2014.1329).

Figure 1 shows how choices were presented. Subjects were asked to choose between two acts, A and B, by clicking on their preferred option. They were then asked

Figure 1. (Color online) Example of the computer interface used in the experiment.



to confirm their choice. If they confirmed, the next question was displayed. If not, the choice was displayed again. We recorded the choice that was confirmed.

Acts were presented both in a matrix format and as pie charts, with the sizes of the pies corresponding to the sizes of the probabilities of the events. We counterbalanced across questions what was option *A* and what was option *B*. We also varied across subjects in which event column (right or left) the stimuli changed during the iteration process. Hence, for half the subjects, the change occurred in the left column; for the other half, it occurred in the right column. Table 1 summarizes the questions asked in the first two parts of our method.

Table 2 shows the triples used to test intransitivity in the third part. Because the elements of the standard sequence elicited in the first part differed between subjects, the choices differed between subjects and were tailor-made to produce intransitivities. For each triple, a convex φ_i predicts the intransitive cycle ABC. Obviously, all special cases of Equation (1) (regret, SSB, additive difference model) that we discussed in §2 make the same prediction.

We used two sets of triples. In triples 1 to 7, the probabilities of the different outcomes were all equal to 1/3. In triples 8 to 14, the probabilities differed and were equal to 1/5, 2/5, and 2/5. Figure 2 shows the presentation of the choices in the third part of the experiment. Triples 1, 2, 3, 6, 8, and 9 also tested PRAM. The other triples had three unequal utility differences, and Loomes (2010) did not explain how to analyze such cases under PRAM.

Previous studies have observed event-splitting effects (also called coalescing), which occur when the same outcome is received under two different states of the world (Humphrey 1995, Starmer and Sugden 1993). An event with a given probability typically receives more weight when it is split into two subevents than when it is presented as a single event. To prevent event-splitting effects from confounding the results, acts always had the same number of states, and subevents were not combined or split. Humphrey (2001) found that event-splitting effects were mainly caused by a preference for more positive outcomes and an aversion to the outcome zero. In our triples, the number of positive outcomes was always the same, and we avoided the outcome zero and used $x_0 = €20$ instead.

Prior to the actual experiment, subjects answered two training questions. After these questions, we elicited the outcome t that led to indifference between $(1/3, t; 2/3, 11)$ and $(1/3, 40; 2/3, 16)$. These answers were not used in the final analyses and only served to monitor for confusion about the experimental instructions.[8]

Each experimental session began by eliciting the elements of the standard sequence, which were used as inputs in the other two parts. The order of the second and the third part was counterbalanced. The measurement of the standard sequence had to be performed in a fixed order, but the order of the choices in the other parts was random.

Thus far, to ease presentation, we have presented our method in a deterministic context. We now explain how we assessed the effect of response errors, i.e., the possible deviation of choices from preferences. We combined three approaches to assess the effect of response errors. First, we repeated certain choices during the experiment. Second, we incorporated the possibility of various types of response errors into our statistical analysis. Third, we ran a simulation study with several error specifications to study the robustness of our results and the power of our tests.

4.3. Repeated Choices

To test for response error, 13 choices were repeated in total. After the first part, we repeated the third choice of the iteration procedure for two randomly selected questions. After the second part, we repeated the third choice of the iteration procedure for each of $z_{i,2/5}$, $z_{i,3/5}$, and $z_{i,3/4}$. Subjects were generally close to indifference in the third choice, and response errors were therefore more likely. We also repeated eight randomly selected choices from the third part.

We finally repeated the entire elicitations of $x_{i,1}$, the first element of the standard sequence, and $z_{i,1/4}$. These repetitions gave insight into the imprecision of the elicited indifference values. A difference between the two elicited values of $x_{i,1}$ might also signal strategic responding, a potential limitation of the trade-off method. In the trade-off method, answers are used as inputs in later questions. By overstating their answers, subjects could increase the attractiveness of later questions. Because we used a choice-based elicitation procedure and forced choices, subjects did not actually see their elicited indifference values and were less likely to detect the chained nature of the experiment. Even if they did, subjects could not be aware of chaining in the first question, the original elicitation of $x_{i,1}$, because this question did not use previous responses. If the subjects answered strategically in the remaining questions and overstated their indifference values, the repeated measurement of $x_{i,1}$, which could be affected by strategic responding, should exceed the original measurement, which could not be affected by strategic responding.

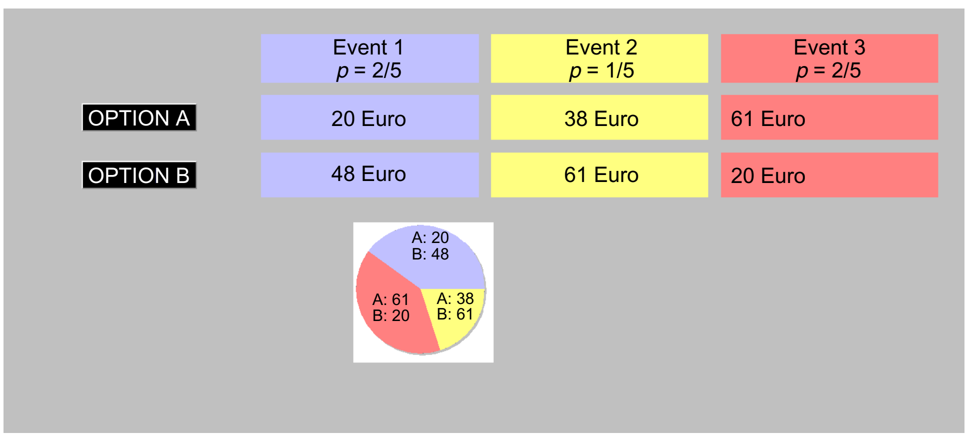
Table 1. Summary of the first two parts of our measurement procedure.

	Elicited value	Indifference
First part: Measurement of the standard sequence	$x_{i,j},\ j=1,\ldots,5$	$(1/3, x_{i,j}; 2/3, 11) \sim (1/3, x_{i,j-1}; 2/3, 16)$
Second part: Measurement of φ_i and δ_i	$z_{i,p},\ p=1/4, 2/5, 3/5, 3/4$	$(p, x_{i,4}; 1-p, 20) \sim (p, x_{i,3}; 1-p, z_{i,p})$

Table 2. The fourteen choice triples used to test intransitivity.

Triple	Act A	Act B	Act C
1	$(1/3, x_{i,2}; 1/3, x_{i,2}; 1/3, x_{i,2})$	$(1/3, x_0; 1/3, x_{i,3}; 1/3, x_{i,3})$	$(1/3, x_{i,1}; 1/3, x_{i,1}; 1/3, x_{i,4})$
2	$(1/3, x_{i,3}; 1/3, x_{i,3}; 1/3, x_{i,2})$	$(1/3, x_{i,1}; 1/3, x_{i,4}; 1/3, x_{i,3})$	$(1/3, x_{i,2}; 1/3, x_{i,2}; 1/3, x_{i,4})$
3	$(1/3, x_{i,4}; 1/3, x_{i,4}; 1/3, x_{i,2})$	$(1/3, x_{i,2}; 1/3, x_{i,5}; 1/3, x_{i,3})$	$(1/3, x_{i,3}; 1/3, x_{i,3}; 1/3, x_{i,4})$
4	$(1/3, x_{i,4}; 1/3, x_{i,3}; 1/3, x_{i,2})$	$(1/3, x_{i,1}; 1/3, x_{i,5}; 1/3, x_{i,3})$	$(1/3, x_{i,2}; 1/3, x_{i,2}; 1/3, x_{i,5})$
5	$(1/3, x_{i,5}; 1/3, x_{i,1}; 1/3, x_{i,2})$	$(1/3, x_{i,2}; 1/3, x_{i,3}; 1/3, x_{i,3})$	$(1/3, x_{i,3}; 1/3, x_0; 1/3, x_{i,5})$
6	$(1/3, x_{i,4}; 1/3, x_{i,2}; 1/3, x_{i,1})$	$(1/3, x_0; 1/3, x_{i,4}; 1/3, x_{i,3})$	$(1/3, x_{i,2}; 1/3, x_0; 1/3, x_{i,5})$
7	$(1/3, x_{i,4}; 1/3, x_{i,2}; 1/3, x_{i,1})$	$(1/3, x_0; 1/3, x_{i,5}; 1/3, x_{i,2})$	$(1/3, x_{i,1}; 1/3, x_{i,1}; 1/3, x_{i,5})$
8	$(1/5, x_0; 2/5, x_{i,1}; 2/5, x_{i,5})$	$(1/5, x_{i,2}; 2/5, x_{i,3}; 2/5, x_{i,2})$	$(1/5, x_{i,4}; 2/5, x_0; 2/5, x_{i,4})$
9	$(1/5, x_0; 2/5, x_{i,3}; 2/5, x_{i,3})$	$(1/5, x_{i,2}; 2/5, x_{i,5}; 2/5, x_0)$	$(1/5, x_{i,4}; 2/5, x_{i,2}; 2/5, x_{i,2})$
10	$(1/5, x_{i,1}; 2/5, x_{i,2}; 2/5, x_{i,4})$	$(1/5, x_{i,3}; 2/5, x_{i,5}; 2/5, x_0)$	$(1/5, x_{i,5}; 2/5, x_{i,1}; 2/5, x_{i,3})$
11	$(1/5, x_{i,1}; 2/5, x_{i,2}; 2/5, x_{i,4})$	$(1/5, x_{i,3}; 2/5, x_{i,5}; 2/5, x_0)$	$(1/5, x_{i,4}; 2/5, x_{i,1}; 2/5, x_{i,3})$
12	$(1/5, x_{i,1}; 2/5, x_{i,2}; 2/5, x_{i,4})$	$(1/5, x_{i,3}; 2/5, x_{i,5}; 2/5, x_0)$	$(1/5, x_{i,5}; 2/5, x_{i,2}; 2/5, x_{i,2})$
13	$(1/5, x_0; 2/5, x_{i,2}; 2/5, x_{i,4})$	$(1/5, x_{i,2}; 2/5, x_{i,5}; 2/5, x_0)$	$(1/5, x_{i,4}; 2/5, x_{i,2}; 2/5, x_{i,2})$
14	$(1/5, x_0; 2/5, x_0; 2/5, x_{i,5})$	$(1/5, x_{i,2}; 2/5, x_{i,4}; 2/5, x_0)$	$(1/5, x_{i,4}; 2/5, x_0; 2/5, x_{i,3})$

Figure 2. (Color online) Example of a choice question in the third part of the experiment.



4.4. Analysis of Errors

Testing axioms of measurement theory, such as transitivity, requires accounting for the inherently unreliable nature of choice behavior. As has been noted by Duncan Luce (Luce 1995, 1997), this process involves two steps: (1) to recast a deterministic model, such as Equation (1) and PRAM, as a probabilistic model and (2) to use the appropriate statistical methodology for testing that probabilistic model.

We used two types of probabilistic models. Most of our analyses were based on random utility models, in which the decision maker has a fixed deterministic strength of preference, and the probability of selecting one act over another is a function of this strength of preference. This class is also referred to as Fechnerian error models. We also used a tremble model, in which the decision maker has a fixed preference relation but makes occasional errors (or trembles) when choosing between acts. Empirical evidence suggests that the random utility model describes preferences better (Loomes et al. 2002, Loomes and Sugden 1995), but we included the tremble model to test the robustness of our findings.

To make the third part probabilistic, we first analyzed the aggregate data assuming that all individuals have the same fixed preferences but make errors. Regenwetter et al. (2010) have shown that statistical testing of hypotheses under this model is complex because we may face an order-constrained inference problem where the likelihood ratio statistic does not have an asymptotic χ^2 distribution. Davis-Stober (2009) derived the appropriate tests for these cases (which use weighted χ^2 distributions). We used the QTEST package (Regenwetter et al. 2014) to implement these tests.[9] QTEST requires at least 20 observations per choice pair, and we therefore pooled the data for this analysis. This pooling causes no problems if a theory makes a precise prediction as in the case of Equation (1) with convex φ_i and PRAM. We allowed error rates up to 50%, i.e., subjects could mistakenly choose the wrong prospect up to 50% of the time. That is, the preferred prospect should be the modal choice up to sampling variability. Both the random utility and the tremble model make this prediction.

To capture individual heterogeneity, we computed for each individual the curvature of the φ_i function and the PRAM parameter δ_i using the data collected in parts 1 and 2 of the experiment. We used two measures of the curvature of a subject's φ_i function. First, we computed for each subject the area under the (normalized) φ_i-function minus the area under the diagonal. If φ_i is convex (concave), then this area is negative (positive). We also estimated each subject's φ_i-function by a power function using nonlinear least squares. Convex (concave) utility corresponds with a power coefficient greater (less) than 1. We measured δ_i using nonlinear least squares. We then used these measures of φ_i and δ_i to predict the number of cycles (using Poisson regressions) and the individual choices (using random probit models) in the third part.

Poisson regressions are particularly suitable for analyzing count data, as in our case where we count the number of cycles for each subject (Kleiber and Zeileis 2008). Poisson regressions are Fechnerian models where more convex φ_i and higher δ_i should lead to more ABC cycles. Because there were more subjects who displayed no cycles than the Poisson distribution permits, we used the zero-inflated Poisson regression, which is a mixture of a Poisson count component and an additional point mass at zero.[10]

We finally analyzed the individual choices. Equation (1) predicts that the likelihood that a subject chooses A over B, B over C, and C over A in the third part increases with the convexity of φ_i. The same prediction holds for PRAM if δ_i increases. We tested these predictions using random-effect probit analysis, which is also a Fechnerian type of analysis, where the probability of choosing in line with Equation (1) or PRAM is a function of the intensity of preference. The error terms in the probit analyses were subject-dependent to reflect that some subjects make larger errors than others.[11]

4.5. Simulation Study

The above analyses assume that the standard sequences and the $z_{i,p}$ were measured without error. To account for error in all three parts, we used a simulation study. We assumed the subjects to be regret maximizers with u_i and Q_i power functions.[12] The power coefficients were drawn from beta distributions with means m_u and m_Q, standard deviations s_U and s_Q, minimum values 0 and maximum values 5.[13] We implemented both a random utility and a tremble model. In the random utility model, each subject had an individual distribution of Fechner errors. This distribution was normal with mean zero and subject-specific standard deviation. The subject-specific standard deviation was drawn from a uniform distribution on [0, sdmax]. In the tremble model, each subject had an individual tremble p_i drawn from a uniform distribution on [0, $p$max].

We started the simulation by selecting a set of values for m_u, m_Q, s_u, s_Q, sdmax, $p$max, and a specific error model. We then used exactly the same elicitation method as in the experiment. From the simulated elements of the standard sequence, we could estimate utility and compare it to the true (selected) value. From the simulated values of $z_{i,p}$, we could estimate Q_i and compare it to the true value. We then computed the number of cycles in the third part and performed random effects probit analyses to test whether the choices in the third part were significantly predicted by the convexity of φ_i, which gave us an impression of the power of our method and its robustness to errors.

We selected 24 sets of parameter estimates with errors ranging from small to very large (for details, see the appendix). For each set of parameters, we simulated 100 experiments with 50 subjects each. Hence, we simulated 120,000 subjects in total. For the mean power utility coefficient, we selected 0.80 and 0.90, corresponding to mild concavity, which is usually observed in studies using the trade-off method to measure utility. For the standard deviation of the power coefficient, we selected 0.10 and 0.50. Most empirical studies found a standard deviation close to 0.10, but we also wanted to explore the impact of using a larger value. For Q_i, we selected power coefficients 1.20 and 2, corresponding to mild and stronger convexity, and standard deviations of 0.20 and 1.00. We selected larger standard deviations than for utility because Bleichrodt et al. (2010) found more volatility for Q than for utility.

5. Results

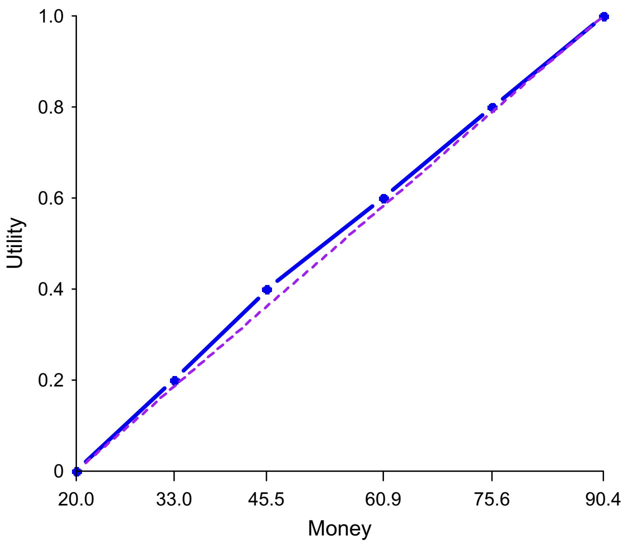
5.1. Repeated Choices

We repeated 13 choice questions and used them to compute replication rates, i.e., the rate of identical choices between the original and the repeated choices. The replication rates were 79.6% in the first part, 72.7% in the second part, and 73% in the third part. These rates are comparable to the values observed in previous research (Stott 2006). There was no difference between the two measurements of $x_{i,1}$ and $z_{i,1/4}$ (paired t-test, $p = 0.42$ for $x_{i,1}$ and $p = 0.52$ for $z_{i,1/4}$). The absence of a difference between the two measurements of $x_{i,1}$ indicates that the measurements using the trade-off method were not affected by strategic responding.

5.2. First Part: Measurement of the Standard Sequence

For two subjects, the repeated measurement of $x_{i,1}$ was lower than the original measurement by more than three times the standard deviation (22 instead of 76 for one subject and 27 instead of 58.5 for the other). Their responses most likely reflected confusion, and therefore we excluded them from the remaining analyses.[14]

Figure 3. (Color online) Utility based on the mean data.



The horizontal axis of Figure 3 shows the mean values of the elements of the standard sequence. The medians were similar. The difference between successive elements increases slightly, but the null hypothesis that it is constant could only be rejected at the 10% level (repeated measures ANOVA, $p = 0.08$).

Under regret theory, Tversky's additive difference model, and PRAM, the elements of the standard sequence determine the utility function. Figure 3 shows the utility function using the mean data. The dotted line indicates linear utility. Utility was close to linear, which agrees, of course, with the finding that the differences between successive elements of the standard sequence were small. We also estimated utility assuming that it belongs to the power family. The estimated power coefficient using the pooled data was 0.87 (se = 0.04).

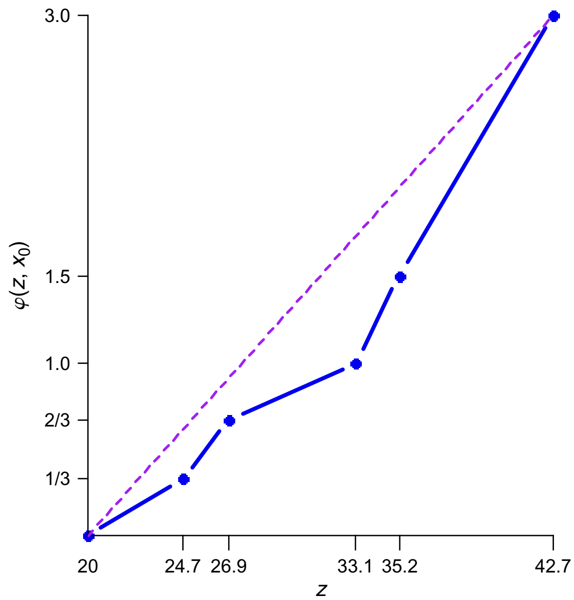
At the individual level, utility was also close to linear. The mean of the individual power coefficients was 0.97, which did not differ from 1 (t-test, $p = 0.70$). There were 16 subjects whose fitted power coefficient was significantly less than 1 (at the 5% level), corresponding to concave utility, and 14 subjects for whom it significantly exceeded 1, corresponding to convex utility (binomial test, $p = 0.86$). For the remaining 22 subjects, the estimated utility function did not differ from linearity.

5.3. Second Part: Measurement of φ_i and δ_i

For two subjects, the original measurement of $z_{i,1/4}$ exceeded the repeated measurement by more than three times the standard deviation. Their responses most likely reflected confusion, and therefore we excluded these subjects.

Figure 4 shows that the estimated φ function based on the mean data was convex. The estimated power coefficient using the pooled data was 1.67 (standard error = 0.23), which differed significantly from 1 (t-test, $p < 0.01$).

Figure 4. (Color online) The function φ based on the mean data.



Most individual functions φ_i were also convex. Based on the area under the individual φ_i-functions, 35 subjects had a convex φ_i, and 15 subjects had a concave φ_i. The mean of the individual power coefficients in the estimation of φ_i at the subject level was 1.37. The mean of the individual estimates of δ_i in the PRAM model was 1.57. Under regret theory, our findings are similar to the findings of Bleichrodt et al. (2010).

5.4. Third Part: Triples to Test Intransitivity

Table 3 shows the response patterns for the 14 triples used to test intransitivity. The two intransitive patterns are shaded, and the final row shows the total proportion of cycles for each of the 14 triples. Intransitive cycles were rare. The proportion of cycles is comparable to Birnbaum and Schmidt (2008) and Loomes (2010) and lower than in Loomes et al. (1991) and Starmer and Sugden (1998), who observed intransitivity rates of approximately 20%. The latter two studies also found that the cycle ABC, which is predicted by Equation (1) with convex φ and by PRAM, was more common than the opposite cycle BCA. We did not observe this asymmetry.

The QTEST analysis indicated that the dearth of intransitive cycles was not due to response errors. We could reject the null hypothesis that ABC cycles, which are predicted by Equation (1) with convex φ_i, were the modal choice ($p < 0.001$). We could also reject the null hypotheses that BCA cycles were the modal choice ($p < 0.001$) and that PRAM was the modal choice ($p < 0.001$).

According to Equation (1), ABC cycles should increase with the convexity of φ_i and with δ_i because they imply that the strength of preference of A over B, of B over C, and

Table 3. The proportions of subjects displaying each of the eight possible response patterns in the triples to test intransitivity.

	Triple													
Pattern	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)	10 (%)	11 (%)	12 (%)	13 (%)	14 (%)
ABC	4	4	6	2	4	0	0	6	4	2	6	6	2	10
ABA	26	24	14	10	14	26	22	8	14	20	10	2	6	12
ACC	8	14	12	18	4	6	12	4	36	14	26	30	36	4
ACA	24	30	22	24	8	20	30	8	20	24	30	24	18	6
BBC	10	4	12	12	18	12	4	18	10	12	2	10	6	26
BBA	18	10	18	18	32	14	10	28	6	12	8	14	8	10
BCC	8	8	12	10	8	14	12	20	8	10	4	12	16	30
BCA	2	6	4	6	12	8	10	8	2	6	14	2	8	2
ABC + BCA	6	10	10	8	16	8	10	14	6	8	20	8	10	12

Notes. The light shaded patterns are the intransitive patterns. The pattern ABC is predicted by Equation (1) with φ convex and by PRAM. The final row shows for each triple the total proportion of subjects who cycled.

of C over A increases. We found no effect of the convexity of φ_i. In the Poisson regressions, the coefficient of the area measure of convexity on the number of ABC cycles even had the wrong (negative) sign (t-test, $p = 0.04$). There was also no evidence that BCA cycles increased with the concavity of φ_i (t-test, $p = 0.79$). However, we did find evidence that the number of ABC cycles increased with the value of δ_i, as predicted by PRAM (t-test, $p = 0.01$).

Table 3 presents the results of all subjects, including subjects for whom no intransitivities were predicted, because φ_i did not deviate from linearity or δ_i did not differ from 1. Tversky (1969) already observed that only a minority of subjects displayed intransitive choice behavior. Intransitive cycles remained rare when we restricted attention to the subjects who were most likely to display intransitive choice behavior (subjects for whom φ_i was significantly different from linearity based on their estimated power coefficients and subjects for whom δ_i was significantly larger than 1). None of these subjects displayed more than three intransitive cycles. However, the probability of ABC cycles now increased with the convexity of φ_i (t-test, $p = 0.05$) and the probability of BCA cycles with the concavity of φ_i (t-test, $p < 0.01$), as predicted by Equation (1). As before, and in agreement with PRAM, the number of intransitive cycles increased with the value of δ_i (t-test, $p < 0.01$).

In the random effects probit analyses, the convexity of φ_i had no effect on the probability of choosing according to the predictions of Equation (1) with convex φ. The probability of choosing according to PRAM also did not increase with the value of δ_i. These conclusions remained true when we restricted attention to those subjects who had significantly convex φ_i and significantly positive δ_i.

5.5. Simulation Study

The appendix shows the results from the simulation analysis. The picture that emerges is as follows. First, regardless of the error specification (random utility or tremble) and the parameters chosen, the number of predicted cycles is much higher than what we observed in Table 3. The proportion of cycles always exceeds 25% and for the more realistic scenarios is approximately 50%. Second, in nearly all specifications, the proportion of ABC cycles is much higher than the proportion of BCA cycles, contrary to what we observed in Table 3. Third, the power of detecting a significant effect of the convexity of φ in the random effects probit analyses is very high. In the random utility model, it is close to 100%, and in the tremble model, it is typically above 80%. The simulations showed that a random utility model reflected the repeated choices and measurements better than a tremble model did (see the appendix). We conclude that the lack of support for Equation (1) and PRAM in our experiment is not due to low power.

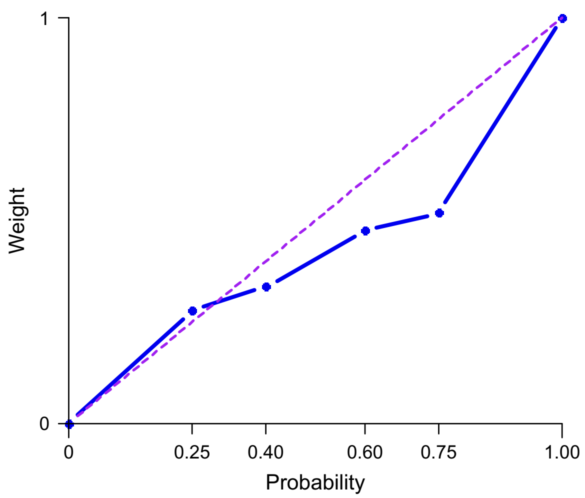
6. Prospect Theory

Expected utility rules out both interactions between acts (through transitivity) and interactions within acts (through Savage's sure thing principle). A key property of intransitive models is that they allow interactions between acts. Our data indicated that expected utility did not hold: the observed convexity of φ is inconsistent with expected utility. However, we observed little evidence of intransitive cycles and hence of interactions between acts. This result raises the question of whether deviations from expected utility are better explained through interactions within acts. In this section, we briefly study this question.

We assume prospect theory, the main descriptive theory of decision under uncertainty that allows for interactions within acts (Tversky and Kahneman 1992). In prospect theory, preferences are defined over gains and losses relative to a reference point. As our experiment used only gains, we will describe prospect theory for gains, in which case it coincides with rank-dependent utility (Quiggin 1981, 1982). We assume without loss of generality that all acts $(p_1, x_1; \ldots; p_n, x_n)$ are *rank-ordered*; i.e., $x_1 \geqslant \cdots \geqslant x_n$. Prospect theory evaluates acts $(p_1, x_1; \ldots; p_n, x_n)$ as

$$\sum_{j=1}^{n} \pi_j u(x_j). \tag{13}$$

Figure 5. (Color online) The probability weighting function based on the mean data.



The *decision weights* π_j are defined as $w(\sum_{i=1}^{j} p_i) - w(\sum_{i=1}^{j-1} p_i)$ with w a nondecreasing *probability weighting function* that satisfies $w(0) = 0$ and $w(1) = 1$. Equation (13) shows that under prospect theory, the evaluation of an act does not depend on the other acts in the choice set. Hence, prospect theory excludes between-act interactions. Because the weight given to the utility of an outcome depends on its rank, prospect theory includes within-act interactions.

Wakker and Deneffe (1996) showed that the trade-off method can measure the utility function in rank-dependent utility. Consequently, the first part of our measurement procedure measures u in Equation (13), and Figure 3 illustrates what the mean utility function looks like. In §5, we found that if we assume power utility, the pooled estimate of the power coefficient was equal to 0.87, which is approximately the same as the estimate obtained by Tversky and Kahneman (1992).

From the indifference $(p, x_{i,4}; 1-p, x_0) \sim (p, x_{i,3}; 1-p, z_{i,p})$, elicited in the second part, we obtain under prospect theory

$$w_i(p) = \frac{u_i(z_{i,p})}{u_i(x_{i,4}) - u_i(x_{i,3}) + u_i(z_{i,p})}. \qquad (14)$$

The utility of $z_{i,p}$ was generally unknown but could be estimated through interpolation from the individual utility function measured in the first part. This estimation gave the probability weights for $p = 1/4$, $p = 2/5$, $p = 3/5$, and $p = 3/4$. Figure 5 illustrates the probability weighting function based on the mean data. The diagonal shows expected utility, in which probability weighting is linear. The elicited probability weighting function was inverse S-shaped, overweighting small probabilities and underweighting larger probabilities, which is consistent with earlier findings (Abdellaoui 2000, Bleichrodt and Pinto 2000, Gonzalez and Wu 1999). The pooled estimate of the probability weighting parameter using the family proposed by Tversky and Kahneman (1992)[15] was 0.57, which is close to the value of 0.61 obtained by Tversky and Kahneman (1992). The mean of the individual probability weighting parameters was 0.72 (st. error = 0.05).

To assess the predictive power of prospect theory, we used for each subject the utilities and probability weights obtained in the first two parts to predict the individual's choices in the third part. We estimated the weights of probabilities 1/5, 1/3, 2/3, and 4/5 used in the triples of the third part by fitting for each subject the one-parameter probability weighting function proposed by Tversky and Kahneman (1992). We could then estimate for each subject the probability weights based on the subject's obtained probability weighting parameter estimate.[16]

All but five of our subjects had an inverse S-shaped probability weighting function. Prospect theory with an inverse S-shaped probability weighting function made a clear prediction in nine of the triples in the third part. We analyzed these unambiguous triples using QTEST and could reject the null hypothesis that the choices predicted by prospect theory were the modal choices ($p < 0.01$). In most triples, prospect theory predicted the choices well, but in some it was clearly off. This result may be because the aggregate analysis does not account for strength of preference. Indeed, the random effects probit analyses showed that the difference in prospect theory value had a strong and significant effect in predicting the choices in the third part (t-test, $p < 0.01$). By contrast, in §5 we concluded that neither the convexity of φ_i nor the PRAM parameter δ_i predicted choices.

7. Concluding Remarks

Intransitive choices were thin on the ground. This conclusion is consistent with Birnbaum and Schmidt (2008) and Regenwetter et al. (2011a) even though we used tests that were specifically designed to uncover violations of transitivity. The intransitive theories that we tested did not predict choices. These theories include regret theory and PRAM, a rich model of intransitive choice that was recently proposed by Loomes (2010) and that extends the similarity models of Rubinstein (1988), Leland (1994, 1998), and Mellers and Biagini (1994) by permitting continuous similarity judgments. The lack of support for PRAM is consistent with a recent paper by Guo and Regenwetter (2014). We explored several strategies for modeling the stochastic nature of choice, including a simulation exercise, but the evidence against the intransitive models was robust.

One explanation for the lack of support for the intransitive models might be that we used two-outcome acts in the measurement of the intransitive models but three-outcome acts in the tests of intransitivity. Theoretically this difference should not matter, but increasing the number of outcomes complicates the experimental tasks. The more complex a task, the more likely subjects are to resort to simple heuristics (Payne 1976, Swait and Adamowicz 2001).

Prospect theory explained our data better than the intransitive models, even though the QTEST analysis showed that its fit was not perfect. Perhaps this less than perfect fit could be expected given that our study was not specifically designed to measure prospect theory. The better performance of prospect theory suggests another explanation for the lack of support for the intransitive models, namely that they do not describe preferences well. Concepts such as regret and similarity are intuitive, and evidence from neuroscience suggests that they play a fundamental role in regulating choice behavior (Camille et al. 2004). However, the general intransitive model and PRAM may not be the appropriate way to model this intuition. In particular, by assuming separability across events, they rule out all within-act interactions. Our data suggest that accounting for the violations of expected utility requires abandoning event separability. Allowing for between-act interactions while retaining event separability is not a viable modeling strategy.

This conclusion does not imply that between-act interactions play no role in decision making under uncertainty. We found little support for the between-act interactions modeled by the intransitive models, but there may be other ways in which between-act interactions affect choices. For example, they may shape reference points, a key ingredient of prospect theory. Prospect theory does not predict how reference points are formed. One possibility is that they are determined by a comparison between the acts under consideration. Evidence for such between-act interaction was obtained by Bleichrodt et al. (2001), Hershey and Schoemaker (1985), and van Osch et al. (2006). To explore this possibility further and to develop formal models capturing such between-act interactions is an important topic for future research.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2014.1329.

Acknowledgments

Michael Birnbaum, Enrico Diecidue, Graham Loomes, Tony Marley, Ulrich Schmidt, Peter Wakker, an anonymous reviewer, and in particular Michel Regenwetter provided helpful comments. The research of Aurélien Baillon and Han Bleichrodt was made possible through grants from the Netherlands Organization for Scientific Research (NWO) [453-06-001 and 451-10-011]. The research of Alessandra Cillo was made possible through the Bocconi Research Division and the Spanish ministry of science and innovation, Project ECO2008-05155/ECON.

Endnotes

1. Other studies that make this point include Iverson and Falmagne (1985), Sopher and Gigliotti (1993), Luce (2000), Birnbaum and Gutierrez (2007), Birnbaum and Schmidt (2008), Birnbaum (2010), and Birnbaum and Schmidt (2010). In contrast, Myung et al. (2005), who reanalyzed Tversky's (1969) data using a sophisticated Bayesian approach, concluded that the violations of transitivity were real. However, the set of models they considered was restricted, and their main purpose was to show the ability of the Bayesian approach to select among nonnested models.

2. There are intransitive models that do not belong to these two classes, e.g., models based on a lexicographic ordering (Brandstätter et al. 2006, Tversky 1969). These models are not the topic of our research. Birnbaum (2010) and Regenwetter et al. (2011b) tested these models and found them to be inconsistent with their data.

3. González-Vallejo's (2002) proportional difference (PD) model is also based on the concept of similarity. It embeds a deterministic similarity core in a stochastic framework. In the online supplement, we show that our data are inconsistent with the PD model.

4. Loomes (2010) used the letter c to denote this function. For consistency with the rest of the paper, we use the letter u.

5. Strictly speaking, we measured $\varphi_i(\cdot, x_0)$. For notational convenience, we will write φ_i.

6. The parameter α_i could be measured by adding one question. We could ask, for instance, for the value of z_i that led to indifference between $(p_1, x_{i,4}; p_2, x_0; p_3, x_0)$ and $(p_1, x_{i,3}; p_2, z_i; p_3, x_0)$, which implies $(1/(5u_i(z_i)))^{\delta_i} = (p_2/p_1)^{(p_1+p_2)^{\alpha_i}}$. The parameter α_i can immediately be solved from this equation.

7. The experiment can be found at http://regret.unibocconi.it/.

8. The mean response was €53.50 (median €50), and all subjects reported a value of t exceeding €40, suggesting that they understood the task.

9. We also analyzed the data under Birnbaum's true and error model, which makes different independence assumptions (Birnbaum 2011, Regenwetter et al. 2011b). The conclusions were similar.

10. To test for robustness, we also estimated a hurdle model, which is a two-part model where the first part is a binary model that answers the question "does a subject display any cycles at all?" and the second part is a count part that answers the question "if a subject displays cycles, how many?" The hurdle model led to the same conclusions.

11. Therefore, the number of parameters was four, leaving 2,096 degrees of freedom. We also estimated a model with both subject- and choice-dependent errors to reflect that errors were more likely in some choices than in others (leaving 2,054 degrees of freedom). This model led to the same conclusions.

12. We could not use φ_i in the simulations. In part two of the experiment, we measured $\varphi_i(\cdot, x_0)$ and could assess the convexity of φ_i in its first variable. However, for the simulation exercise, we needed to know the entire function φ_i.

13. Permitting larger maximum values would be even more favorable for our approach because it would increase the number of predicted ABC cycles.

14. Retaining them did not affect the conclusions.

15. $w(p) = p^\gamma/(p^\gamma + (1-p)^\gamma)^{1/\gamma}$. For $0.27 < \gamma < 1$, w is inverse S-shaped.

16. We also estimated the weights based on linear interpolation. The results were similar.

Appendix. Simulation Results

						u			Q			φ		% cycles		Random probit significance of φ				Consistency		
m_U	s_U	m_Q	s_Q	Error type	*p*max/ *sd*max	Mean	Std	Median of abs diff	Mean	Std	Median of abs diff	Mean	Std	ABC	BCA	N.S.	10%	5%	1%	Repeated choices	x_1	$z_{1/4}$
0.9	0.1	1.2	0.2	Tremble	0.2	0.92	0.43	0.12	1.07	0.43	0.16	1.05	0.41	44.9	6.4	4	2	15	79	90.8	1	1
0.8	0.1	2	0.2	Tremble	0.2	0.83	0.48	0.14	1.42	0.8	0.44	1.37	0.77	73.3	0.1	7	2	13	78	90.8	1	1
0.9	0.5	1.2	1	Tremble	0.2	0.98	0.83	0.14	0.96	0.82	0.15	0.9	0.78	34.1	35.9	1	0	2	97	90.4	1	1
0.8	0.5	2	1	Tremble	0.2	0.84	0.84	0.12	1.33	0.96	0.36	1.24	0.9	57.6	11.7	0	1	2	97	90.7	1	1
0.9	0.1	1.2	0.2	Tremble	0.5	0.96	0.72	0.28	0.79	0.53	0.35	0.77	0.51	29.4	5.9	12	6	18	64	79.4	2	2
0.8	0.1	2	0.2	Tremble	0.5	0.87	0.78	0.32	0.91	0.84	1.22	0.87	0.81	50.5	1.2	0	0	3	97	78.9	2	2
0.9	0.5	1.2	1	Tremble	0.5	1	1.06	0.29	0.68	0.72	0.26	0.64	0.67	23.3	28.4	8	4	8	80	79.1	4	2
0.8	0.5	2	1	Tremble	0.5	0.88	1.27	0.28	0.89	0.92	0.87	0.83	0.87	40.8	9	2	3	13	82	79.3	2	2
0.9	0.1	1.2	0.2	Tremble	1	1.09	1.39	0.49	0.51	0.53	0.75	0.49	0.52	20.6	8.2	14	7	14	65	66.3	6	5
0.8	0.1	2	0.2	Tremble	1	0.99	1.3	0.52	0.55	0.78	1.65	0.53	0.75	33.2	4.8	0	0	5	95	66.4	6	6
0.9	0.5	1.2	1	Tremble	1	1.08	1.6	0.48	0.46	0.61	0.43	0.44	0.57	16.9	22.2	26	3	13	58	66.5	10	6
0.8	0.5	2	1	Tremble	1	1.05	1.69	0.5	0.54	0.77	1.38	0.51	0.72	27.5	9.9	3	4	6	87	66.6	7	5
0.9	0.1	1.2	0.2	Fechner	5	0.96	0.73	0.29	0.88	0.49	0.33	0.86	0.48	31.5	3.5	0	0	3	97	77.1	3	2
0.8	0.1	2	0.2	Fechner	5	0.84	0.59	0.2	1.35	0.77	0.6	1.31	0.75	78.6	0.1	0	0	1	99	94.5	2	1
0.9	0.5	1.2	1	Fechner	5	1.17	1.44	0.39	0.79	0.85	0.32	0.79	0.86	31.7	15.2	0	0	0	100	72	6	2
0.8	0.5	2	1	Fechner	5	1.12	1.35	0.33	1.06	1.02	0.69	1.06	1.04	49.3	6.9	0	0	0	100	78.5	3	2
0.9	0.1	1.2	0.2	Fechner	10	0.98	0.98	0.41	0.7	0.5	0.52	0.68	0.5	23.2	4.6	1	3	4	92	71.2	5	3
0.8	0.1	2	0.2	Fechner	10	0.85	0.79	0.29	1.06	0.79	1.03	1.03	0.78	65.6	0.4	0	0	0	100	90.1	3	2
0.9	0.5	1.2	1	Fechner	10	1.22	1.71	0.48	0.7	0.85	0.35	0.71	0.87	29.6	12.9	0	0	0	100	68.4	8	3
0.8	0.5	2	1	Fechner	10	1.14	1.64	0.4	0.97	1.04	0.85	0.97	1.06	45.4	6.5	0	0	0	100	75.4	4	2
0.9	0.1	1.2	0.2	Fechner	20	1.04	1.33	0.51	0.55	0.49	0.7	0.54	0.48	19.3	5.5	6	6	13	75	66.4	8	4
0.8	0.1	2	0.2	Fechner	20	0.88	1.06	0.4	0.85	0.77	1.3	0.83	0.76	53.1	0.8	0	1	3	96	85.2	4	2
0.9	0.5	1.2	1	Fechner	20	1.22	1.64	0.54	0.63	0.82	0.43	0.64	0.84	28.7	11	0	0	0	100	65.5	9	5
0.8	0.5	2	1	Fechner	20	1.2	1.67	0.44	0.87	1.02	0.96	0.88	1.04	42.8	6.7	0	0	0	100	73.5	5	3

Notes. Each row corresponds to 100 experiments with 50 subjects. The first 4 parameters give the population mean and standard deviation for the two functions of regret theory, u and Q. The fifth and six columns indicate the error type and the maximum error size (the maximum tremble or the maximum standard deviation of the Fechner parameter). From the simulations of parts 1 and 2, we estimated the power parameters of u and Q. We report the mean and standard deviation of these estimates and the median absolute difference between the estimates and the original, randomly selected parameters. We then report the estimates of φ and the proportion of cycles we would observe. For each of the 100 experiments, we ran a random probit regression. We report how many times φ is significant at three different significance levels (and N.S. for not significant). Finally, we report the proportion of repeated choices that were consistent and the median absolute difference between the two measurements of x_1 and of $z_{1/4}$ (which were 4.8 and 1.9 in our experiment).

In the tremble model, the median absolute difference between the two measurements of x_1 is similar to the difference between the two measurements of $z_{1/4}$. This result contradicts what we observed in our experiment, where the median absolute difference was higher for x_1 than for $z_{1/4}$. A maximum tremble of 0.5 is realistic to reproduce the observed differences for $z_{1/4}$ but too low to reproduce the observed differences for x_1. Symmetrically, a higher maximum tremble is realistic for x_1 but unrealistic for $z_{1/4}$. The simulations using the random utility model do replicate that the median absolute difference is lower for $z_{1/4}$ than for x_1. If we also consider the 73% consistency rate that we observed in the repeated choices, then we see that Fechner errors between 5 and 10 give plausible consistency results (a maximum of 20 mostly gives unrealistic inconsistencies). The table also shows that for all plausible combinations, the power of detecting a significant result in the random probit analysis is close to 100%.

References

Abdellaoui M (2000) Parameter-free elicitation of utility and probability weighting functions. *Management Sci.* 46(11):1497–1512.

Bell DE (1982) Regret in decision making under uncertainty. *Oper. Res.* 30(5):961–981.

Bell DE (1983) Risk premiums for decision regret. *Management Sci.* 29(10):1156–1166.

Birnbaum MH (2010) Testing lexicographic semiorders as models of decision making: Priority dominance, integration, interaction, and transitivity. *J. Math. Psych.* 54(4):363–386.

Birnbaum MH (2011) Testing mixture models of transitive preference: Comment on Regenwetter et al. (2011). *Psych. Rev.* 118(4):675–683.

Birnbaum MH, Gutierrez RJ (2007) Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organ. Behav. Human Decision Processes* 104(1):96–112.

Birnbaum MH, Schmidt U (2008) An experimental investigation of violations of transitivity in choice under uncertainty. *J. Risk Uncertainty* 37(1):77–91.

Birnbaum MH, Schmidt U (2010) Testing transitivity in choice under risk. *Theory Decision* 69(4):599–614.

Bleichrodt H, Pinto JL (2000) A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Sci.* 46(11):1485–1496.

Bleichrodt H, Cillo A, Diecidue E (2010) A quantitative measurement of regret theory. *Management Sci.* 56(1):161–175.

Bleichrodt H, Pinto JL, Wakker PP (2001) Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Sci.* 47(11):1498–1514.

Bouyssou D (1986) Some remarks on the notion of compensation in MCDM. *Eur. J. Oper. Res.* 26(1):150–160.

Brandstätter E, Gigerenzer G, Hertwig R (2006) The priority heuristic: Making choices without trade-offs. *Psych. Rev.* 113(2):409–432.

Camille N, Coricelli G, Sallet J, Pradat-Diehl P, Duhamel J, Sirigu A (2004) The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304(5674):1167–1170.

Chorus C (2012) Random regret minimization: An overview of model properties and empirical evidence. *Transport Rev.* 32(1):75–92.

Davis-Stober CP (2009) Analysis of multinomial models under inequality constraints: Applications to measurement theory. *J. Math. Psych.* 53(1):1–13.

Fishburn PC (1982) Non-transitive measurable utility. *J. Math. Psych.* 26(1):31–67.

Fishburn PC (1990) Skew symmetric additive utility with finite states. *Math. Soc. Sci.* 19(2):103–115.

Fishburn PC (1991) Nontransitive additive conjoint measurement. *J. Math. Psych.* 35(1):1–40.

Fishburn PC (1992) Additive differences and simple preference comparisons. *J. Math. Psych.* 36(1):21–31.

Gonzalez R, Wu G (1999) On the shape of the probability weighting function. *Cognitive Psych.* 38(1):129–166.

González-Vallejo C (2002) Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psych. Rev.* 109(1):137–155.

Guo Y, Regenwetter M (2014) Quantitative tests of the perceived relative argument model: Commentary on Loomes (2010). *Psych. Rev.* 121(4):696–705.

Hershey JC, Schoemaker PJH (1985) Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Sci.* 31(10):1213–1231.

Humphrey SJ (1995) Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *J. Risk Uncertainty* 11(3):263–274.

Humphrey SJ (2001) Are event-splitting effects actually boundary effects? *J. Risk Uncertainty* 22(1):79–93.

Iverson G, Falmagne J (1985) Statistical issues in measurement. *Math. Soc. Sci.* 10(2):131–153.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

Kleiber C, Zeileis A (2008) *Applied Econometrics with R* (Springer, New York).

Leland JW (1994) Generalized similarity judgments: An alternative explanation for choice anomalies. *J. Risk Uncertainty* 9(2):151–172.

Leland JW (1998) Similarity judgments in choice under uncertainty: A reinterpretation of the predictions of regret theory. *Management Sci.* 44(5):659–672.

Loomes G (2010) Modeling choice and valuation in decision experiments. *Psych. Rev.* 117(3):902–924.

Loomes G, Sugden R (1982) Regret theory: An alternative theory of rational choice. *Econom. J.* 92(368):805–824.

Loomes G, Sugden R (1987) Some implications of a more general form of regret theory. *J. Econom. Theory* 41(2):270–287.

Loomes G, Sugden R (1995) Incorporating a stochastic element into decision theories. *Eur. Econom. Rev.* 39(3–4):641–648.

Loomes G, Moffatt PG, Sugden R (2002) A microeconometric test of alternative stochastic theories of risky choice. *J. Risk Uncertainty* 24(2):103–130.

Loomes G, Starmer C, Sugden R (1991) Observing violations of transitivity by experimental methods. *Econometrica* 59(2):425–439.

Luce RD (1995) Four tensions concerning mathematical modeling in psychology. *Annual Rev. Psych.* 46:1–27.

Luce RD (1997) Several unresolved conceptual problems of mathematical psychology. *J. Math. Psych.* 41(1):79–87.

Luce RD (2000) *Utility of Gains and Losses*: *Measurement-Theoretical and Experimental Approaches* (Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey).

May KO (1954) Intransitivity, utility, and the aggregation of preference patterns. *Econometrica* 22(1):1–13.

Mellers BA, Biagini K (1994) Similarity and choice. *Psych. Rev.* 101(3): 505–518.

Myung JI, Karabatsos G, Iverson GI (2005) A Bayesian approach to testing decision making axioms. *J. Math. Psych.* 49(3):205–225.

Payne JW (1976) Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organ. Behav. Human Performance* 16(2):366–387.

Quiggin J (1981) Risk perception and risk aversion among Australian farmers. *Australian J. Agricultural Econom.* 25(2):160–169.

Quiggin J (1982) A theory of anticipated utility. *J. Econom. Behav. Organ.* 3(4):323–343.

Regenwetter M, Dana J, Davis-Stober CP (2010) Testing transitivity of preferences on two-alternative forced choice data. *Frontiers Psych.*, http://dx.doi.org/10.3389/fpsyg.2010.00148.

Regenwetter M, Dana J, Davis-Stober CP (2011a) Transitivity of preferences. *Psych. Rev.* 118(4):42–56.

Regenwetter M, Dana J, Davis-Stober CP, Guo Y (2011b) Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). *Psych. Rev.* 118(1):684–688.

Regenwetter M, Davis-Stober CP, Lim SH, Guo Y, Popova A, Zwilling C, et al. (2014) QTEST: Quantitative testing of theories of binary choice. *Decision* 1(1):2–34.

Rubinstein A (1988) Similarity and decision-making under risk (Is there a utility theory resolution to the allais paradox?). *J. Econom. Theory* 46(1):145–153.

Sopher B, Gigliotti G (1993) Intransitive cycles: Rational choice or random error? *Theory Decision* 35(3):311–336.

Starmer C, Sugden R (1993) Testing for juxtaposition and event-splitting effects. *J. Risk Uncertainty* 6(3):235–254.

Starmer C, Sugden R (1998) Testing alternative explanations of cyclical choices. *Economica* 65(259):347–361.

Stott HP (2006) Cumulative prospect theory's functional menagerie. *J. Risk Uncertainty* 32(2):101–130.

Swait J, Adamowicz W (2001) The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *J. Consumer Res.* 28(1):135–148.

Tversky A (1969) Intransitivity of preferences. *Psych. Rev.* 76(1):31–48.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* 5(4):297–323.

van Osch SMC, van den Hout WB, Stiggelbout AM (2006) Exploring the reference point in prospect theory: Gambles for length of life. *Medical Decision Making* 26(4):338–346.

Wakker PP, Deneffe D (1996) Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Sci.* 42(8):1131–1150.

Aurélien Baillon is a professor of behavioral economics at the Erasmus School of Economics of Erasmus University Rotterdam. His work focuses on individual decision making under risk and ambiguity. Through both empirical and theoretical studies, his research addresses issues in subjective probability elicitation, models of attitude towards risk and ambiguity, and aggregation of expert opinions.

Han Bleichrodt is a professor of behavioral economics at the Erasmus School of Economics of Erasmus University Rotterdam. His research activities focus on decisions—those made under uncertainty and those made over time and on social choice theory. He has a particular interest in those decisions involving health.

Alessandra Cillo is an assistant professor of decision sciences at Bocconi University Milan. Her research focuses on decisions under uncertainty. Her experimental work in behavioral economics addresses issues in elicitation of utility functions with single or multiple attributes. Her theoretical work focuses on risk value modeling.