Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

A Tailor-Made Test of Intransitive Choice

Aurélien Baillon, Han Bleichrodt, Alessandra Cillo

Working Paper n. 496

This Version: August, 2013

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

A Tailor-Made Test of Intransitive Choice

AURÉLIEN BAILLON
Erasmus School of Economics, PO Box 1738, 3000 DR Rotterdam, Netherlands, baillon@ese.eur.nl

HAN BLEICHRODT
Erasmus School of Economics, PO Box 1738, 3000 DR Rotterdam, Netherlands, bleichrodt@ese.eur.nl

ALESSANDRA CILLO
Bocconi University and Igier, Via Roentgen 1, 20136 Milan, Italy, alessandra.cillo@unibocconi.it

August 2013

Abstract

We performed a new test of transitivity based on individual measurements of the main intransitive choice models in decision under uncertainty. Our test is tailor-made and, therefore, more likely to detect violations of transitivity than previous tests. In spite of this, we observed only few intransitivities and we could not reject the hypothesis that these were due to random error. A possible explanation for the poor predictive performance of the intransitive choice models is that they only allow for interactions between acts, but 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 significantly better than the nontransitive choice models. We conclude that descriptively realistic models need to allow for within-act interactions, but may retain transitivity.

Subject classifications: Utility/preference: Estimation. Decision analysis: Risk.
Area of review: Decision Analysis.

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 wide 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. It has been argued that they were primarily caused by random errors and that the actual proportion of intransitive preferences is negligible and not convincing enough to abandon transitive theories.[1]

All previous tests of transitivity faced the problem of choosing the right stimuli. Typically, stimuli were selected in a somewhat haphazard way, 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. It does not account for the extensive heterogeneity in preferences that is usually observed in empirical studies.

An alternative approach, which we adopt in this article, is to select a model of nontransitive choice, measure its parameters for each individual separately and then use these measurements to select the individual-specific stimuli that will produce choice cycles according to the model. This "tailor-made approach" has not been used as yet because there were no methods to measure intransitive choice models. It was widely believed that the possibility of intransitive choice excluded the existence of real-valued functions representing these choices.

Broadly speaking, two types of nontransitive choice models can be distinguished. The first type explains intransitive choice models through the notion of regret. Examples include the regret theories of Bell (1982, 1983), Loomes and Sugden 1982, 1987), Fishburn's (1982) SSB theory and, more recently, the random regret minimization model (Chorus 2012), which is used in transport modeling.

The second type of model explains intransitive choice through the notion of similarity (Leland 1994, Leland 1998, Mellers and Biagini 1994, Rubinstein 1988). The intuition is that subjects pay less attention to dimensions that are similar and give weight instead to dissimilar dimensions. A limitation of these models is that similarity judgments are of a dichotomous nature: there is a (not clearly specified) threshold above which subjects take stimuli into account and below which they become inconsequential. Recently, Loomes (2010) proposed a new model, the perceived relative argument model (PRAM), which allows continuous rather than dichotomous similarity judgments. PRAM is the most parsimonious intransitive model available today. A model that is related to the similarity models is González-Vallejo's (2002) proportional difference model. This model extends the deterministic similarity models by adding a stochastic term which reflects decision error.

We propose a general measurement method that allows quantifying both types of nontransitive models. We apply it in an experiment on decision under uncertainty and derive subject-specific tests of intransitivity. In spite of using subject-specific tests, we found little evidence of intransitive cycles and we could not reject the null hypothesis that those that were observed were due to random error.

Our data also indicated that 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 nontransitive choice models, explains these deviations through interactions between the acts in the choice set. The second approach, of which prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992) is the best-known exponent, excludes such between-act interactions, but permits within-act interactions through the ranking of the outcomes of an act. Our method can also quantify prospect theory. Prospect theory performed significantly better than the nontransitive theories in predicting our choice data. 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 given up.

2. Theory

2.1. A general nontransitive additive model

Consider a decision maker who faces uncertainty, modeled through a set $\mathcal{S}$ of possible *states of the world*. 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, …, n.

Let $\succcurlyeq$ 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 a nontransitive additive model (Bouyssou 1986, Fishburn 1990, Fishburn 1991)

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

The functions φ in Eq. (1) are real-valued, they are strictly increasing in their first argument and strictly decreasing in their second argument, and they satisfy *symmetry*: for all x_i,y_i, $-\varphi(x_i,y_i) = \varphi(y_i,x_i)$. Symmetry implies that $\varphi(0,0) = 0$. 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 *concave*, for all $x_i \succcurlyeq y_i \succcurlyeq z_i$, $\varphi(x_i,z_i) \leq \varphi(x_i,y_i) + \varphi(y_i,z_i)$, or *convex*, for all $x_i \succcurlyeq y_i \succcurlyeq z_i$, $\varphi(x_i,z_i) \geq \varphi(x_i,y_i) + \varphi(y_i,z_i)$. 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 $\varphi(x_i,y_i) = u(x_i) - u(y_i)$ expected utility results. Of course, in that case no intransitivities can occur. Another special case of Eq.(1) is regret theory (Bell 1982, Loomes and Sugden 1982) where $\varphi(x_i,y_i) = Q(u(x_i) - u(y_i))$. The *utility function* u is then an interval scale, unique up to scale and unit and the function Q, which is unique up to unit, reflects the impact of regret. Convexity of φ then implies convexity of Q, which is called *regret* *aversion*. A closely related model is the random regret minimization model (Chorus 2012). 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 Eq.(1) if an interstate uniformity axiom is added, which is plausible in decision under uncertainty.

2.2. The perceived relative argument model

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

$$X \succcurlyeq Y \Leftrightarrow \psi(b_X, b_Y) \geq \xi(u_Y, u_X). \tag{2}$$

In Eq. (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 they 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 Eq. (3) α is a person-specific variable whose value may vary from one individual to another. To capture the common violations of expected utility, α should be negative. A comparable idea underlies the similarity models of Leland (1994, 1998), Mellers and Biagini (1994), and Rubinstein 1988). 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. PRAM allows more diverse applications of the 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.[2] Then u_Y denotes in utility terms the advantage that Y has over X and u_X denotes 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

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

Expected utility is the special case of PRAM where $\alpha = 0$ and $\delta = 1$. If $\delta > 1$ then whichever is the bigger 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 Section 3 we show that PRAM predicts intransitive cycles when $\delta > 1$ and that their likelihood increases with the value of δ.[3]

2.3. The proportional difference model

González-Vallejo's proportional difference (PD) model is also based on the notion of similarity, but it embeds a deterministic similarity core in a stochastic framework. Consider two acts X = $(p, x_1; 1-p, x_2)$ and Y = $(p, y_1; 1-p, y_2)$ with $x_1 > y_1$, $x_2 < y_2$, $p_1 < ½$. According to the PD model of González-Vallejo (2002, Eq. (3) and the extension to more than two attributes discussed on page 140), X is strictly preferred to Y if and only if

$$\frac{x_1 - y_1}{x_1} - \frac{y_2 - x_2}{y_2} - \frac{1 - 2p_1}{1 - p_1} \geq \tau + \varepsilon. \tag{5}$$

In Eq.(5), τ is the decision maker's decision threshold. González-Vallejo (2002) suggests that τ can depend on the context and on the decision task. However, within tasks τ is constant. The parameter ε is a random noise term with mean zero. Equation (5) says that X will be preferred to Y if the difference between the proportional advantage of X over Y ($\frac{x_1 - y_1}{x_1}$) and the proportional advantage of Y over X ($\frac{y_2 - x_2}{y_2} + \frac{1 - 2p_1}{1 - p_1}$) exceeds the decision threshold plus error.

3. Predicting intransitivities

3.1 First part

Our procedure for predicting intransitivities consists of three parts. The first part used the trade-off method of Wakker and Deneffe (1996) to elicit a *standard sequence* of money amounts $x_0 ,\ldots, x_5$. Two gauge outcomes R and r (R > r > 0), a probability p, and a starting outcome x_0 were selected and we elicited x_1 such that (p, x_1; 1−p, r) ~ (p, x_0; 1−p, R). According to Eq. (1) this indifference implies that

$$\varphi(x_1,x_0) = \frac{1-p}{p}\varphi(R,r). \tag{6}$$

We then substituted the elicited value of x_1 for x_0 and elicited x_2 such that (p, x_2; 1−p, r) ~ (p, x_1; 1−p, R). This gives

$$\varphi(x_2,x_1) = \frac{1-p}{p}\varphi(R,r). \tag{7}$$

It follows from Eqs. (6) and (7) that $\varphi(x_2,x_1) = \varphi(x_1,x_0)$. Because φ is increasing in its first argument and decreasing in its second, this equality implies that the distance in terms of φ between x_2 and x_1 is equal to the distance between x_1 and x_0. Under regret theory, Eqs. (6) and (7) imply that that $u(x_2) – u(x_1) = u(x_1) – u(x_0)$, i.e. the utility difference between x_2 and x_1 is equal to the utility difference between x_1 and x_0.

Continuing this procedure we elicited indifferences $(p,x_{j+1}; 1–p,r) \sim (p,x_j; 1–p,R)$, j = 0,…,4, and thus obtained a standard sequence for which it is true that successive elements are equally spaced in terms of φ. Because φ is a ratio scale we can set $\varphi(x_1,x_0) = 1$. In regret theory and the additive difference model, we obtain that $u(x_j) – u(x_{j–1}) = u(x_1) – u(x_0)$ because u is an interval scale, we can set $u(x_0) = 0$ and $u(x_5) = 1$. Then $u(x_j) = j/5$, j = 0 ,…, 5.

Under PRAM the indifferences between $(p,x_{j+1}; 1-p,r)$ and $(p,x_j; 1-p,R)$ imply that:

$$\left(\frac{u(x_{j+1}) - u(x_j)}{u(R) - u(r)}\right)^{\delta} = \frac{1-p}{p}. \tag{8}$$

Consequently, $u(x_{j+1}) - u(x_j) = u(x_1) - u(x_0)$, $j = 1,\ldots,4$ and successive elements of the standard sequence are equally spaced in terms of utility. If we scale u such that $u(x_0) = 0$ and $u(x_5) = 1$, then $u(x_j) = j/5$.

Finally, under the PD model

$$\left|\frac{x_{j+1} - x_j}{x_{j+1}} - \frac{R-r}{R} - \frac{1-p-p}{1-p}\right| \leq \tau + \varepsilon. \tag{9}$$

Eq. (9) implies that $\frac{x_{j+1} - x_j}{x_{j+1}}$ should be constant up to random noise for successive elements of the standard sequence, which yields an immediate test of the PD model.

3.2. Second part

In the second part, we selected outcomes x_0, x_3, and x_4 from the standard sequence and elicited z_p such that $(p, x_4; 1-p, x_0) \sim (p, x_3; 1-p, z_p)$. The second part provided us with more information about the function φ in Eq. (1) (and consequently about the function Q in regret theory) and the parameter δ in PRAM. Under Eq. (1) and the chosen scaling $\varphi(x_1,x_0) = \varphi(x_j,x_{j-1})$ the indifferences $(p, x_4; 1-p, x_0) \sim (p, x_3; 1-p, z_p)$ imply that

$$\varphi(z_p,x_0) = \frac{p}{1-p}. \tag{10}$$

Equation (10) defines φ as a function of z_p. The subscript p in z_p serves as a reminder that the value of φ depends on the probability used in the elicitation. By varying p we could measure as many points of φ as desired.

The second part of our procedure also allows measuring the parameter δ in PRAM. According to Eqs. (2) – (4), the indifference $(p, x_4; 1-p, x_0) \sim (p, x_3; 1-p, z_p)$ implies that

$$\left(\frac{u(z_p) - u(x_0)}{u(x_4) - u(x_3)}\right)^{\delta} = (5^*u(z_p))^{\delta} = \frac{p}{1-p}. \tag{11}$$

Equation (11) gives the value of δ. The utility of z_p was generally unknown, but could be estimated using elements of the standard sequence.

3.3. Third part

The measurements of the first part were used in the third part to design the tailor-made tests of transitivity. For example, we asked subjects the following three choices:

(i) A = (⅓, x_4; ⅓, x_4; ⅓, x_2) vs. B = (⅓, x_2; ⅓, x_5; ⅓, x_3).

(ii) B vs. C = (⅓, x_3; ⅓, x_3; ⅓, x_4).

(iii) C vs A.

According to Eq. (1) the comparison between A and B gives

$$A \succcurlyeq B \Leftrightarrow \tfrac{1}{3}^*\varphi(x_4, x_2) + \tfrac{1}{3}^*\, \varphi(x_4, x_5) + \tfrac{1}{3}^*\, \varphi(x_2, x_3) \geq 0. \tag{12}$$

By the symmetry of φ, the chosen scaling, and the properties of the standard sequence, $\varphi(x_4, x_5) = \varphi(x_3, x_4) = \varphi(x_2, x_3) = -\, \varphi(x_1, x_0)$. Hence, Eq. (12) can be written as

$$A \succcurlyeq B \Leftrightarrow \varphi(x_4, x_2) - 2 \geq 0. \tag{13}$$

In other words, the decision maker will choose A over B if (and only if) φ is convex. A similar analysis shows that if φ is convex then the decision maker will choose B over C and C over A. Hence, convex φ entails the cycle ABC (ABC stands for "A chosen over B, B chosen over C, and C chosen over A."). Concave φ implies the preference cycle BCA. In regret theory the cycle ABC is usually called a *regret cycle.* The opposite cycle BCA (which is rarely observed empirically) may then be coined a *rejoice cycle.*

PRAM also predicts the cycle ABC. A = (⅓, x_4; ⅓, x_4; ⅓, x_2) gives a better outcome than B = (⅓, x_2; ⅓, x_5; ⅓, x_3) in the first state (which has probability ⅓) and the utility difference is $u(x_4) - u(x_2)$ =

$\frac{2}{5}$. B gives a better outcome than A in the second and in the third state (which have a joint probability of $\frac{2}{3}$) and the utility difference is $u(x_5) - u(x_4) = u(x_3) - u(x_2) = \frac{1}{5}$. According to Eqs. (2) – (4),

$$A \succcurlyeq B \Leftrightarrow (\tfrac{2}{5}/\tfrac{1}{5})^{\delta} = 2^{\delta} \geq \tfrac{2}{3}/\tfrac{1}{3} = 2. \tag{14}$$

It follows from (14) that $A \succ B$ if and only if $\delta > 1$. PRAM with $\delta > 1$ also predicts that $B \succ C = (\frac{1}{3}, x_3; \frac{1}{3}, x_3; \frac{1}{3}, x_4)$ (because $u(x_4) - u(x_3) = u(x_1) - u(x_0)$), and that $C \succ A$. Hence, unless $\delta = 1$, when 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. The parameter α drops out of Eq. (14) because the probabilities of the states in which the outcomes differ between the two acts sum to 1 and $1^{\alpha} = 1$.[4]

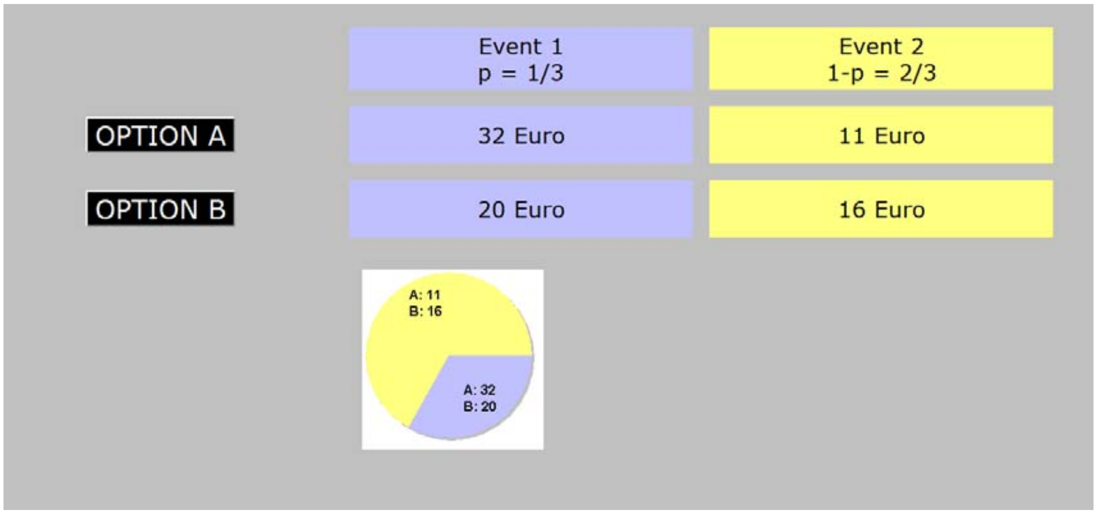
4. Experiment

The previous Section showed that violations of transitivity are closely connected with non-linearity of the function φ and with the parameter δ in PRAM. The aim of our experiment was to explore whether these relations could indeed be observed empirically. In addition, the first part of our method yielded a test of the proportional difference model.

Subjects

Subjects were 54 students (22 male) aged between 18 and 33 (median age 21) from Erasmus University Rotterdam coming from various academic backgrounds. They were paid a €10 show-up fee. In addition, each subject had a 10% chance to play out one of his 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. This selection was performed in the presence of the subject using the Excel random number generator. The subject then played his preferred option in the selected choice with his payoffs determined by another randomly drawn number.

Figure 1. Example of the computer interface used in the experiment.



Procedures

The experiment was computer-run[5] and administered in sessions of two subjects with one experimenter present. Sessions lasted 55 minutes on average. 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 Appendix A. Figure 1 gives an example of the choices subjects faced. Subjects were asked to choose between two acts, A and B, by clicking on their preferred option. They were then asked 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. The confirmation question aimed to reduce the impact of response errors. Acts were presented both in a matrix format and as pie charts with the sizes of the pies corresponding with the sizes of the probabilities of the events. We varied between questions what was Option A and what was Option B. We also varied between 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 and 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 1. Summary of the first two parts of our measurement procedure.

	Elicited outcome	Indifference
First part	x_j, j = 1,…,5	$(⅓, x_j; ⅔, 11) \sim (⅓, x_{j-1}; ⅔, 16)$
Second part	z_p, p = ¼,⅖,⅗,¾	$(p, x_4; 1-p, x_0) \sim (p, x_3; 1-p, z_p)$

Table 2 shows the choices used in the third part to test transitivity. For each triple, convex φ predicts the intransitive cycle ABC. Obviously, all special cases of Eq.(1) (regret, SSB, additive difference model) that we discussed in Section 2 make the same prediction. We used two sets of tests. In tests 1 to 7 the probabilities of the different outcomes were all equal to ⅓. In tests 8 to 14 the probabilities differed and were equal to ⅕, ⅖, and ⅖. We used these two sets to explore whether using different probabilities affected the results. Figure 2 gives an example of the presentation of the choices in the third part of the experiment. Transitivity tests 1, 2, 3, 6, 8, and 9 also tested PRAM. The other transitivity tests had three unequal utility differences and Loomes (2010) did not explain how to analyze such cases under PRAM.

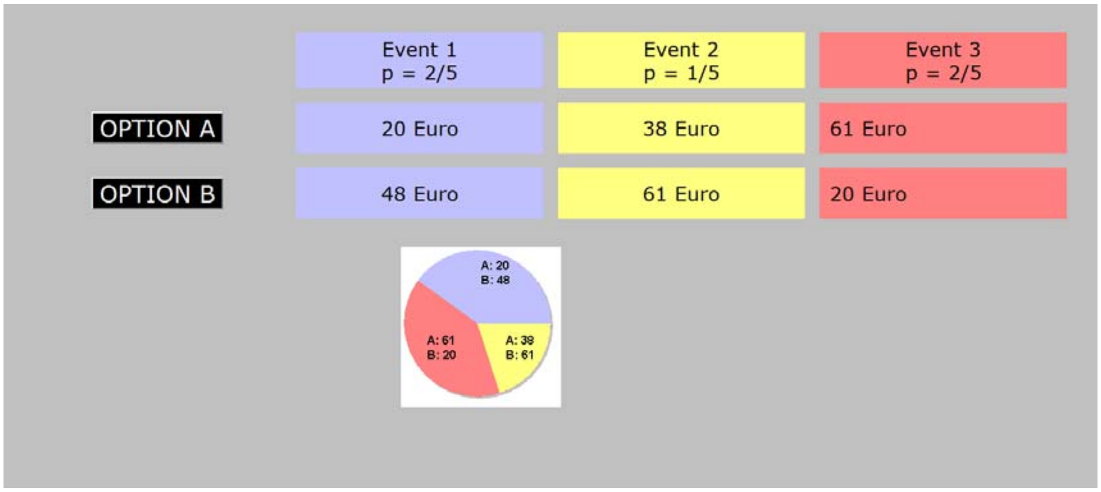
Table 2. The fourteen choice triples used to test transitivity.

Test	Act A	Act B	Act C
1	$(⅓, x_2; ⅓, x_2; ⅓, x_2)$	$(⅓, x_0; ⅓, x_3; ⅓, x_3)$	$(⅓, x_1; ⅓, x_1; ⅓, x_4)$
2	$(⅓, x_3; ⅓, x_3; ⅓, x_2)$	$(⅓, x_1; ⅓, x_4; ⅓, x_3)$	$(⅓, x_2; ⅓, x_2; ⅓, x_4)$
3	$(⅓, x_4; ⅓, x_4; ⅓, x_2)$	$(⅓, x_2; ⅓, x_5; ⅓, x_3)$	$(⅓, x_3; ⅓, x_3; ⅓, x_4)$
4	$(⅓, x_4; ⅓, x_3; ⅓, x_2)$	$(⅓, x_1; ⅓, x_5; ⅓, x_3)$	$(⅓, x_2; ⅓, x_2; ⅓, x_5)$
5	$(⅓, x_5; ⅓, x_1; ⅓, x_2)$	$(⅓, x_2; ⅓, x_3; ⅓, x_3)$	$(⅓, x_3; ⅓, x_0; ⅓, x_5)$
6	$(⅓, x_4; ⅓, x_2; ⅓, x_1)$	$(⅓, x_0; ⅓, x_4; ⅓, x_3)$	$(⅓, x_2; ⅓, x_0; ⅓, x_5)$
7	$(⅓, x_4; ⅓, x_2; ⅓, x_1)$	$(⅓, x_0; ⅓, x_5; ⅓, x_2)$	$(⅓, x_1; ⅓, x_1; ⅓, x_5)$
8	$(⅕, x_0; ⅖, x_1; ⅖, x_5)$	$(⅕, x_2; ⅖, x_3; ⅖, x_2)$	$(⅕, x_4; ⅖, x_0; ⅖, x_4)$
9	$(⅕, x_0; ⅖, x_3; ⅖, x_3)$	$(⅕, x_2; ⅖, x_5; ⅖, x_0)$	$(⅕, x_4; ⅖, x_2; ⅖, x_2)$
10	$(⅕, x_1; ⅖, x_2; ⅖, x_4)$	$(⅕, x_3; ⅖, x_5; ⅖, x_0)$	$(⅕, x_5; ⅖, x_1; ⅖, x_3)$
11	$(⅕, x_1; ⅖, x_2; ⅖, x_4)$	$(⅕, x_3; ⅖, x_5; ⅖, x_0)$	$(⅕, x_4; ⅖, x_1; ⅖, x_3)$
12	$(⅕, x_1; ⅖, x_2; ⅖, x_4)$	$(⅕, x_3; ⅖, x_5; ⅖, x_0)$	$(⅕, x_5; ⅖, x_2; ⅖, x_2)$
13	$(⅕, x_0; ⅖, x_2; ⅖, x_4)$	$(⅕, x_2; ⅖, x_5; ⅖, x_0)$	$(⅕, x_4; ⅖, x_2; ⅖, x_2)$
14	$(⅕, x_0; ⅖, x_0; ⅖, x_5)$	$(⅕, x_2; ⅖, x_4; ⅖, x_0)$	$(⅕, x_4; ⅖, x_0; ⅖, x_3)$

Previous studies have found evidence of event-splitting effects (also called coalescing) that 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 is typically weighted more heavily when it is split into two subevents than when it is presented as a single event. To avoid that event-splitting effects would confound the results, acts always had the number of states was always the same and subevents were not combined or split. Humphrey (2001) found that event-splitting effects were mainly caused by a preference for more positive outcomes. In our tests the number of positive outcomes was always the same. Humphrey (2001) also found that the outcome zero can have an especially negative impact. We therefore avoided the outcome zero in our tests and used x_0 = €20 instead.

The transitivity tests used acts with three different outcomes. On the other hand, the parameters of the general nontransitive model and PRAM were elicited using binary acts. Potential issues with predicting choices over three-outcome acts from preferences over binary acts are discussed in the concluding section.

Figure 2. Example of a choice question in the third part of the experiment.



Prior to the actual experiment, subjects answered two training questions. After these questions, we elicited the outcome x that led to indifference between (⅓, x; ⅔, 11) and (⅓, 40; ⅔, 16). These questions were not used in the final analyses and only served to monitor for confusion about the experimental instructions.[6]

Each experimental session began with the elicitation of the elements of the standard sequence because these were used as inputs in the other two parts. The order of the second and the third part was random. 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.

To test for response error, 13 choices were repeated. 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 $z_{2/5}$, $z_{3/5}$, and $z_{3/4}$. Subjects were generally close to indifference in the thrid 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_1, the first element of the standard sequence, and $z_{1/4}$. These repetitions gave insight into the errors in the elicited indifference values. A difference between the two elicited values of x_1 might also signal strategic responding, a potential drawback 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 they 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_1, because this question did not use previous responses. If subjects answered strategically in the remaining questions and overstated their indifference values, the repeated measurement of x_1, which could be affected by strategic responding, should exceed the original measurement, which could not be affected by strategic responding.

Analysis

Testing axioms of measurement theory, such as transitivity, requires accounting for the inherently unreliable nature of choice behavior. As has been pointed out by Iverson and Falmagne (1985), there is no unique way of doing so and several approaches have been put forward. We used several approaches to account for response error. To explore whether violations of transitivity were due to error, we used

Birnbaum’s true and error model (Birnbaum and Gutierrez 2007, Birnbaum 2011). Let p_{ABC} denote the probability of a true pattern ABC, p_{ABA} the probability of a true pattern ABA, etc. Let e_1, e_2, and e_3 denote the probabilities of making errors in the choices between A and B, B and C, and A and C respectively. According to the true and error model the probability P(ABC) of observing the pattern ABC is:

$$P(ABC) = p_{ABC}(1-e_1)(1-e_2)(1-e_3) + p_{ABA}(1-e_1)(1-e_2)e_3 + \ldots + p_{BCA}e_1e_2e_3. \qquad (15)$$

Let f_i denote the observed and $\hat{f}_i = n*P(i)$ the predicted frequency of pattern i where n is the number of subjects and i = ABC, ABA,….,BCA. Parameters were estimated to minimize $\chi^2 = \sum_{j=1}^{8}(f_i - \hat{f}_i)^2/\hat{f}_i$. Without further restrictions the model is underidentified, but following Birnbaum (2011) we used the preference reversals in the repeated questions to estimate the errors e_1, e_2, and e_3 in the different tests. The errors could be different across choices. This addresses criticism raised against the error model of Harless and Camerer (1994), which assumed the same error rate for all choices. As Loomes (2005) points out, assuming an equal error rate is unrealistic as errors are more likely in some choices than in others. For example, if one act clearly dominates another, hardly anyone makes errors.

A different approach to account for the stochastic nature of experimental data was developed by Iverson and Falmagne (1985). Their probabilistic model is based on the realization that measurement axioms restrict the set of admissible choice probabilities to a union of convex polyhedrons. They then derive likelihood ratio tests to test the axiom in question. Regenwetter et al. (2011) use the same idea. A drawback of this likelihood ratio approach is that it requires many replications, typically at least 20, of each choice. For our method, which is based on tailor-made, theory-driven tests, this would require over 1500 choices in total for each subject, a number that is clearly infeasible. The likelihood ratio approach is conservative (see Iverson and Falmagne, 1985, p.144). It assumes transitivity as the null hypothesis and has relatively low power. On the other hand, the true and error model is neutral towards transitivity and does not assume beforehand that transitivity holds. In other words, if we cannot reject transitivity by the true and error model then it is unlikely that we would have rejected transitivity by the likelihood ratio approach. Only when we would reject transitivity we should be worried that this conclusion might be different under the likelihood ratio approach.

Both the true and error model and the likelihood ratio approach make strong independence assumptions. Equation (15) shows that the true and error model assumes independence of errors within each transitivity test. The likelihood ratio approach makes the even stronger assumption that each choice is independent. We could relax these assumptions in the analysis of the relation between φ and δ and the number of choices made according to Eq. (1) and PRAM, respectively, by estimating mixed effects models. In these models errors can be correlated across choices within subjects.

Finally, to estimate the relation between the number of intransitive cycles and the shape of φ and the value of δ, we used zero-inflated Poisson regression and the hurdle model to account for the relatively large number of zeros that we observed. The reported tests are those of the model that fitted best by the Akaike Information Criterion. Standard errors were always Huber-White errors to allow for general error terms.

5. Results

Consistency

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 those observed in previous research (Stott 2006). There was no difference between the two measurements of x_1 and $z_{¼}$ (paired t-test, $p = 0.42$ for x_1 and $p = 0.52$ for $z_{¼}$). The absence of a difference between the two measurements of x_1 indicates that measurements using the trade-off method were not affected by strategic responding.

Part 1: elicitation of the standard sequence

For two subjects, the repeated measurement of x_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 probably reflected confusion and we, therefore, excluded them from the remaining analyses.[8]

Table 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, $p = 0.08$).

The data are inconsistent with the proportional difference model of González-Vallejo (2002). The prediction of this model that $\frac{x_{j+1} - x_j}{x_{j+1}}$ is constant up to random noise for different j could clearly be rejected (repeated measures ANOVA, $p < 0.01$).

Table 3: Mean values of the elicited standard sequence.

x_0	x_1	x_2	x_3	x_4	x_5
20	32.55 [28.38,33.50]	45.05 [36.38,48.50]	60.39 [46.62,60.75]	74.95 [55.88,75.50]	89.57 [66.50,85.50]
$\frac{x_{j+1} - x_j}{x_{j+1}}$	0.357 [0.295,0.403]	0.251 [0.191,0.295]	0.222 [0.183,0.255]	0.174 [0.137,0.185]	0.153 [0.132,0.171]

Note: interquartile ranges in square brackets.

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 about equal. 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), which indicated concave utility ($p < 0.01$).

Figure 3. Utility based on the mean data.

Second part: elicitation of φ and δ

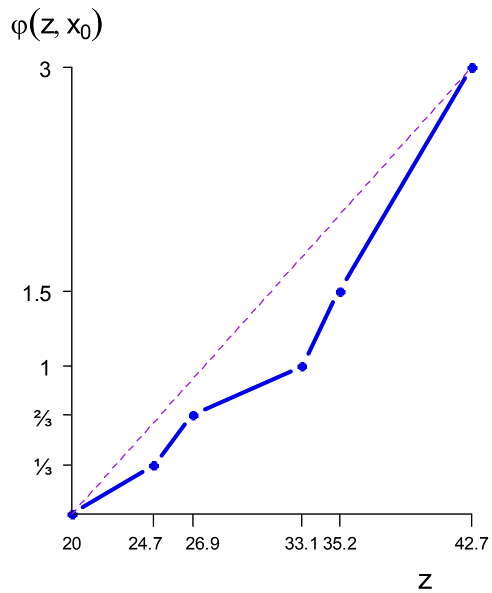
For two subjects, the original measurement of $z_{1/4}$ exceeded the repeated measurement by more than three times the standard deviation and we excluded these subjects.

Figure 4 shows that the estimated φ functions based on the mean data were convex. Parametric estimation showed that this was not due to random error. The estimated power coefficient using the pooled data was 1.67 (standard error = 0.23), which differed significantly from 1 ($p < 0.01$).

At the individual subject level, we also found evidence for convex φ. We computed for each subject the area under the φ-function minus the area under the diagonal. If φ is convex then this area is negative, if φ is concave it is positive. According to this (nonparametric) classification 35 subjects had a convex φ and 15 subjects had a concave φ.

Under regret theory, our findings are similar to those of Bleichrodt et al. (2010). This suggests that our findings are robust and are not due to peculiarities in the data.

Figure 4. The function φ based on the mean data.



The mean of the individual estimates of δ in the PRAM model was 1.57. However, only 11 subjects had δ significantly larger than 1, while 10 subjects had δ was significantly smaller than 1. For the other subjects we could not reject the null that δ was equal to 1 at the 10% level.

Predicting cycles

Table 3 shows the response patterns for the 14 transitivity tests. The two intransitive patterns are shaded and the final row shows the total proportion of cycles for each of the 14 tests. 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 around 20%. The latter two studies also observed that the cycle ABC, which is predicted by Eq. (1) with convex φ and by PRAM, was more common than the opposite cycle BCA. We did not find this asymmetry.

The above conclusions also held when we took the individual subject as the unit of analysis. There were 18 subjects who exhibited no cycles, 15 subjects who exhibited at least one cycle only of the type ABC, 12 subjects who exhibited at least one cycle only of the type BCA, and 6 subjects who cycled in both directions.

According to Eq. (1), ABC cycles should increase with the convexity of φ. We found no evidence for this: the coefficient of the measure of convexity on the number of ABC cycles even had the wrong (negative) sign ($p = 0.05$).[9] There was no evidence either that BCA cycles increased with the concavity of φ ($p = 0.79$). On the other hand, we did find evidence that the number of ABC cycles increased with the value of δ, as predicted by PRAM ($p = 0.01$).

The dearth of intransitive cycles was not due to response errors. The estimated parameters of the true and error model indicated that the proportion of ABC cycles differed from 0 in only 4 tests and even then they were low (0.3%, 1.1%, 2.8%, and 9.2%). The estimated proportion of BCA cycles differed from 0 in 4 tests (1.6%, 3.2%, 4.1%, 6.2%). What is more, the fit of a purely transitive model (i.e. a model in which we imposed the restriction that the proportions of ABC cycles and of BCA cycles are equal to zero) was not significantly worse than that of the unrestricted model, meaning that we could not reject the null hypothesis of transitivity. The best fit to the data was obtained using a transitive model with a common error in all choices. We could not reject this model in any of the 14 tests at a significance level of 1%.

Table 3 presents the results of all subjects, including those for whom no intransitivities were predicted because φ did not deviate from linearity or δ did not differ from 1. Tversky (1969) already concluded that only a minority of subjects displays intransitive choice behavior. We therefore repeated the analysis for those subjects for whom φ deviated significantly from linearity (based on estimated power coefficients) and for those for whom δ was significantly larger than 1. These are the subjects who are most likely to display intransitive choice behavior. We excluded the subjects for whom δ was significantly lower than 1 because their behavior contradicted PRAM.

Table 3. The proportions of subjects displaying each of the eight possible response patterns in the transitivity tests.

Test Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ABC	3.7	3.7	5.6	1.9	3.7	0	0	5.6	3.7	1.9	5.6	7.4	1.9	9.3
ABA	24.1	22.2	13.0	9.3	13.0	29.6	20.4	7.4	13.0	18.5	11.1	3.7	5.6	11.1
ACC	9.3	13.0	13.0	20.4	5.6	5.6	11.1	5.6	35.2	13.0	25.9	31.5	37.0	3.7
ACA	25.9	29.6	20.4	24.1	9.3	18.5	33.3	9.3	20.4	25.9	29.6	22.2	16.7	5.6
BBC	9.3	3.7	11.1	13.0	16.7	11.1	5.6	16.7	9.3	11.1	11.1	9.3	5.6	25.9
BBA	16.7	11.1	20.4	16.7	33.3	13.0	9.3	29.6	9.3	14.8	9.3	13.0	9.3	11.1
BCC	9.3	9.3	13.0	9.3	7.4	13.0	11.1	18.5	7.4	9.3	3.7	11.1	14.8	27.8
BCA	1.9	7.4	3.7	5.6	11.1	9.3	9.3	7.4	1.9	5.6	13.0	1.9	7.4	5.6
ABC + BCA	5.6	11.1	9.3	7.4	14.8	9.3	9.3	13.0	5.6	7.4	18.6	9.3	9.3	14.8

Note. The light shaded patterns are the intransitive patterns. The pattern ABC is predicted by Eq.(10 with φ convex and by PRAM. The final row shows for each test the total proportion of subjects who displayed an intransitive pattern.

There were 10 subjects for whom φ deviated significantly from linearity. For these subjects intransitive cycles remained rare (none of these subjects displayed more than 3 intransitive cycles where 14 were predicted), but the probability of ABC cycles now increased with the convexity of φ ($p = 0.05$) and the probability of BCA cycles with the concavity of φ ($p < 0.01$), as predicted by Eq. (1). The estimated proportion of ABC cycles now differed from 0 in 7 out of 14 tests and for one of the tests we could reject transitivity ($p = 0.04$).

For 11 subjects δ significantly exceeded 1. In agreement with PRAM, the number of intransitive cycles increased with δ ($p < 0.01$) when we only included these subjects in the analyses. On the other hand, even for those subjects, we could not reject transitivity in any of the tests.

The lack of support for intransitivity might be caused by the omission of other variables that affected subjects' choices. A candidate is the difference in expected value between the acts. Even though expected value should theoretically play no role, subjects may have used it as a heuristic in their choices. Expected value may also have had affected choice because subjects made errors in the first part of the experiment and the "true" utility difference between elements of the standard sequence was not always equal. For example, if subjects reported a value of x_1 that was too low then this could explain why they chose the common pattern ABA instead of ABC in the first test. Only act C has states in which x_1 obtains and the downward error in x_1 would be reflected in a relatively low expected value of C. However, the effect of expected value was nonsignificant in all tests.[10]

Another robustness check that we performed also concerned the effects of response error and imprecision in the measured standard sequences. Errors in the elements of the standard sequence may have caused subjects to deviate from the predictions of the general additive nontransitive model (Eq. (1)) and PRAM in the third part of the experiment. The impact of response error is most relevant for outcomes that are close in terms of utility. If the utility difference is larger, cycles will probably show up anyhow. We therefore divided the transitivity tests into those choices in which there was at least one state of nature in which the utility difference between the outcomes exceeded one half[11] and those for which this was not so. Again, this did not affect the results. Using other utility thresholds than ½ had no effect either.

The observed lack intransitivity does not mean, however, that expected utility held. Expected utility is the special case of Eq. (1) with φ linear and the observed convexity of φ violates expected utility. Moreover, the utility difference between successive elements of the standard sequence is also constant under expected utility. Because expected utility evaluates probabilities linearly, subjects should be indifferent in all 42 choices of the third part except for random error. Consequently, the eight response patterns reported in Table 3 should be equally likely under expected utility. However, the true and error model rejected the null of equal proportions in all but one of the tests, indicating that expected utility did not hold.

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 all nontransitive models is that they allow for interactions between acts. Even though we could reject expected utility, we observed little evidence of intransitive cycles and, hence, of interactions between acts. This raises the question whether deviations from expected utility are better explained through interactions within acts. In this Section, we 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, Quiggin 1982). We assume without loss of generality that all acts $(p_1,x_1;\ldots;p_n,x_n)$ are *rank-ordered*, i.e., $x_1 \geq \ldots \geq x_n$. Prospect theory evaluates acts $(p_1,x_1;\ldots;p_n,x_n)$ as

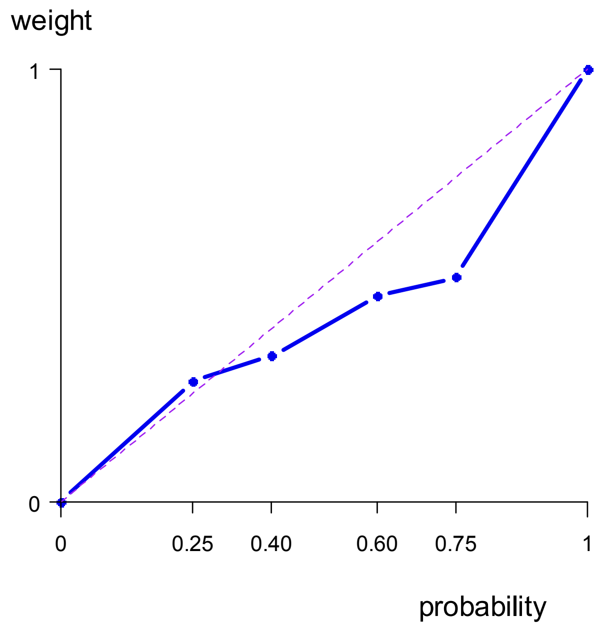
$$\sum_{j=1}^{n} \pi_j u(x_j). \tag{16}$$

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 (16) shows that 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 Eq. (16)

and Figure 3 illustrates what it looks like. In Section 5 we found that if we assume power utility, the pooled estimate of the power coefficient is equal 0.87, which is about the same as the estimate obtained by Tversky and Kahneman (1992).

Figure 5: The probability weighting function based on the mean data



From the indifference $(p, x_4; 1-p, x_0) \sim (p, x_3; 1-p, z_p)$, elicited in the second part, we obtain that

$$w(p) = \frac{u(z_p)}{u(x_4) - u(x_3) + u(z_p)}. \tag{17}$$

The utility of z_p was generally unknown, but could be estimated through interpolation from the utility function measured in the first part. This gave the probability weights for p = ¼, p = ⅖, p = ⅗, and p = ¾. 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 is inverse S-shaped, overweighting small probabilities and underweighting larger probabilities, in agreement 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)[12] was 0.57, which is close to the value of 0.61 obtained by Tversky and Kahneman (1992).

To assess the predictive power of prospect theory, we used the utilities and probability weights obtained in the first two parts to predict the choices in the third part. We estimated the weights of probabilities ⅕, ⅓, ⅔, and ⅘ by fitting for each subject the one-parameter Tversky-Kahneman probability weighting function. We could estimate the probability weights based on the obtained parameter estimates.[13]

Prospect theory predicted choices significantly better than the nontransitive models. Prospect theory predicted the correct choice in 55.9% of the choices, which was significantly higher than 50%, the case of random choice ($p < 0.01$). The general additive nontransitive model (Eq. (1) and PRAM predicted the correct choice in 49.6% and 50.3% of the choices. Both proportions do not differ from 50% (both $p > 0.74$). If we only include subjects who differ significantly from expected utility then these proportions drop to 48.3% for Eq. (1) and 47.8% for PRAM (both $p > 0.38$).

While prospect theory predicted choices better than the nontransitive theories, the 55.9% of correctly predicted choices may still appear low. However, for some acts the difference in prospect theory value was close to zero. For such choices even small errors might lead to deviations between predicted and actual choices. Indeed, the difference in prospect theory value had a strong impact in predicting the choices in the third part (mixed-effects model, $p < 0.01$). By contrast, neither the convexity of φ in Eq. (1) nor δ in PRAM affected choices (both $p > 0.30$).

7. Concluding remarks

Intransitive choices were thin on the ground. This conclusion is in line with Birnbaum and Schmidt (2008) and Regenwetter et al. (2011) even though we used tests that were specifically designed to uncover violations of transitivity. Our data violated all nontransitive theories that we tested, including regret theory, Fishburn's (1982) SSB theory, a special case of Tversky's additive difference model, Gonzalez-Vallejo's (2002) proportional difference model, 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.[14] The nontransitive theories did not predict choices and the convexity of φ (and thus regret) and PRAM's parameter δ did not affect choices. We explored several potentially confounding variables and different strategies for modeling the stochastic nature of choice, but the evidence against the nontransitive models was robust. This does not mean that expected utility held: we observed violations of expected utility both in the second and in the third part of our measurements.

One explanation for the lack of support for the nontransitive models might be that we used two-outcome acts in the measurement of the nontransitive models, but three-outcome acts in the tests of intransitivity. Theoretically this 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).

On the other hand, prospect theory did much better in predicting the choices in the third part. This suggests another explanation, namely that the nontransitive models do not describe people's preferences well. Notions like 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 nontransitive 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. The support for prospect theory that we observed suggests that accounting for the violations of expected utility requires giving up event-separability. Allowing for between-act interactions while retaining event-separability is not a viable modeling strategy. This does not imply that between-act interactions play no role in decision under uncertainty. We found little support for the between-act interactions modeled by the nontransitive models, but there may be other ways in which between-act interactions affect choices. For example, they may shape reference points in prospect theory. Prospect theory does not predict how reference point are formed. One possibility is that it they are determined by a comparison between the acts under consideration. Evidence for such a between-act interaction was given by Bleichrodt et al. (2001), Hershey and Schoemaker (1985), 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.

Appendix A. Procedure to measure the indifference values

To elicit the standard sequence of outcomes in the first part of our procedure, outcomes x_{j+1} were elicited through choices between A = (⅓,x_{j+1}; ⅔, 11) and B = (⅓,x_j; ⅔, 16).[15] The outcomes x_j and x_{j+1} were always integer-valued. The initial value of x_{j+1} was a random integer in the interval [$x_{j,}$ x_j +25]. There were two possible scenarios:

(i) If A was chosen we increased x_{j+1} by €25 until B was chosen. We then halved the step size and decreased x_{j+1} by €13. If A [B] was subsequently chosen we once again halved the step size and increased [decreased] x_{j+1} by €6, etc.

(ii) If B was chosen we decreased x_{j+1} by D′= (x_{j+1} – x_j)/2 until A was chosen. We then increased x_{j+1} by D′/2. If A was subsequently chosen then we increased [decreased] x_{j+1} by D′/4, etc.

The elicitation ended when the difference between the lowest value of x_{j+1} for which B was chosen and the highest value of x_{j+1} for which A was chosen was less than or equal to €2. The recorded indifference value was the midpoint between these two values. Table A1 gives an example of the procedure for the elicitation of x_1 through choices between A = (⅓, x_1; ⅔,11) and B = (⅓,20; ⅔,16). In this example, the initial random value for x_1 was 36. The recorded indifference value was the midpoint of 26 and 28, that is, 27.

Table A1. Example of the elicitation of x_1.

Iteration	x_1	Choice
1	36	A
2	28	A
3	24	B
4	26	B

The procedure in the second part was largely similar. We elicited the value of z_p for which indifference held between A = $(p,x_4; 1-p,20)$ and B = $(p,x_3; 1-p,z_p)$[16] where p was one of {¼, ⅖, ⅗, ¾} and x_4 and x_3 were the outcomes of the standard sequence elicited in the first part. We only used integers for z_p and the program ensured that z_p was never equal to x_3 to avoid the possibility of event-splitting effects. The initial stimulus z_p was a random integer in the range $[z_{EV}-3, z_{EV}+3]$ where z_{EV} is the value of z_p that makes A and B equal in expected value with the restriction that z_p could not be less than €20. There were two possible scenarios:

(i) As long as A was chosen we increased z_p by $D = (x_4 - z_{EV})/2$ if $p \leq ½$ and by $D = (x_5 - z_{EV})/2$ if $p > 1/2$. We used a different adjustment for $p \leq ½$ to avoid violations of stochastic dominance. We kept increasing z_p by this amount until B was chosen. Then we decreased z_p by D/2. If A [B] was subsequently chosen we increased [decreased] z_p by D/4, etc. A special case occurred if the difference between z_p and x_4 (for $p \leq ½$) or between z_p and x_5 (for $p > ½$) was less than 5. Then we increased z_p by 10 and subsequently kept increasing z by 5 until B was chosen. Then we decreased z by 3.

(ii) If B was chosen we decreased z_p by $D' = (z - 20)/2$ until A was chosen. We then increased z_p by $D'/2$. If A [B] was subsequently chosen we increased [decreased] z_p by $D'/4$, etc.

The remainder of the procedure was the same as in the elicitation of u. The elicitation ended when the difference between the lowest value of z_p for which B was chosen and the highest value of x_{j+1} for which A was chosen was less than or equal to €2. The recorded indifference value was the midpoint between these two values. Table A2 gives an example of the procedure for the elicitation of $z_{¼}$. In the example, the initial choice was between A = (¼, 61; ¾, 20) and B = (¼, 48; ¾, 26), where 26 was selected as the initial stimulus value from the interval [24.3 –3, 24.3 +3]. The recorded indifference value was 30, the midpoint between 29 and 31.

Table A2. Example of the elicitation of $z_{¼}$ when x_4 = 70 and x_3 = 50.

Iteration	$z_{¼}$	Choice
1	26	A
2	44	B
3	35	B
4	31	B
5	29	A

Endnotes

[1] 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), Birnbaum and Schmidt (2010), and Regenwetter et al. (2011). On the other hand, 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 non-nested models.

[2] Loomes used the letter c to denote this function, but for consistency with the rest of the paper we use the letter u.

[3] In general, whether intransitive choices are observed also depends on α. In our tests PRAM predicts that the probability of cycles increases with δ regardless of the value of α.

[4] The parameter α could be measured by adding one question. We could ask for instance for the value of z that led to indifference between (p_1, x_4; p_2, x_0; p_3, x_0) and (p_1, x_3; p_2, z; p_3, x_0). This implies that $\left(\frac{1}{5*u(z)}\right)^{\delta} = \left(\frac{p_2}{p_1}\right)^{(p1 + p2)^{\alpha}}$. The parameter α can immediately be solved from this equation.

[5] The experiment can be found at http://regret.unibocconi.it/.

[6] The mean response was €53.50 (median €50) and all subjects reported a value of x exceeding €40 suggesting that they understood the task.

[7] In the true and error model the rate of preference reversals in choice problem j is equal to $2*e_j*(1 - e_j)$.

[8] Keeping them in did not affect the conclusions.

[9] The results reported here are based on estimating zero-inflated Poisson regressions with robust standard errors.

[10] It was, however, significantly positive in the mixed model.

[11] An example is the choice between (⅓, x_4; ⅓, x_2; ⅓, x_1) and (⅓, x_0; ⅓, x_4; ⅓, x_3). In the first state the utility difference between x_4 and x_0 is ⅘, which exceeds ½.

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

[13] We also estimated the weights based on linear interpolation. The results were similar.

[14] In a recent comment on Loomes (2010), Guo and Regenwetter (2013) also report data challenging PRAM.

[15] In the experiment we varied what was option A and what was option B.

[16] In the experiment we varied which option was A and which B.

Acknowledgements

Michael Birnbaum, Enrico Diecidue, Graham Loomes, Tony Marley, Michel Regenwetter, Ulrich Schmidt, Peter Wakker, and a reviewer provided helpful comments. The research of Aurélien Baillon and Han Bleichrodt was made possible through grants from the Netherlands Organization for Scientific Research (NWO). 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.

References

Abdellaoui, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management Science* **46** 1497-1512.

Bell, D. E. 1982. Regret in decision making under uncertainty. *Operations Research* **30** 961-981.

Bell, D. E. 1983. Risk premiums for decision regret. *Management Science* **29** 1156-1166.

Birnbaum, M. H. 2010. Testing lexicographic semiorders as models of decision making: Priority dominance, integration, interaction, and transitivity. *Journal of Mathematical Psychology* **54** 363-386.

Birnbaum, M. H. 2011. Testing mixture models of transitive preference: Comment on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review* **118** 675-683.

Birnbaum, M. H., R. J. Gutierrez. 2007. Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organizational Behavior and Human Decision Processes* **104** 96-112.

Birnbaum, M. H., U. Schmidt. 2008. An experimental investigation of violations of transitivity in choice under uncertainty. *Journal of Risk and Uncertainty* **37** 77-91.

Birnbaum, M. H., U. Schmidt. 2010. Testing transitivity in choice under risk. *Theory and Decision* **69** 599-614.

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

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

Bleichrodt, H., A. Cillo, E. Diecidue. 2010. A quantitative measurement of regret theory. *Management Science* **56** 161-175.

Bouyssou, D. 1986. Some remarks on the notion of compensation in MCDM. *European Journal of Operational Research* **26** 150-160.

Brandstätter, E., G. Gigerenzer, R. Hertwig. 2006. The priority heuristic: Making choices without trade-offs. *Psychological Review* **113** 409-432.

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

Chorus, C. 2012. Random regret minimization: An overview of model properties and empirical evidence. *Transport Reviews* **32** 75-92.

Fishburn, P. C. 1982. Non-transitive measurable utility. *Journal of Mathematical Psychology* **26** 31-67.

Fishburn, P. C. 1990. Skew symmetric additive utility with finite states. *Mathematical Social Sciences* **19** 103-115.

Fishburn, P. C. 1991. Nontransitive additive conjoint measurement. *Journal of Mathematical Psychology* **35** 1-40.

Fishburn, P. C. 1992. Additive differences and simple preference comparisons. *Journal of Mathematical Psychology* **36** 21-31.

Guo, Y., M. Regenwetter. 2013. A quantitative test of the perceived relative argument model. Commentary on Loomes (Psychological Review, 2010). Working Paper.

Gonzalez, R., G. Wu. 1999. On the form of the probability weighting function. *Cognitive Psychology* **38** 129-166.

González-Vallejo, C. 2002. Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review* **109** 137-155.

Harless, D., C. F. Camerer. 1994. The predictive utility of generalized expected utility theories. *Econometrica* **62** 1251-1289.

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

Humphrey, S. J. 1995. Regret aversion or event-splitting effects? more evidence under risk and uncertainty. *Journal of Risk and Uncertainty* **11** 263-274.

Humphrey, S. J. 2001. Are event-splitting effects actually boundary effects? *Journal of Risk and Uncertainty* **22** 79-93.

Iverson, G., J. Falmagne. 1985. Statistical issues in measurement. *Mathematical Social Sciences* **10** 131-153.

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

Leland, J. W. 1994. Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty* **9** 151-172.

Leland, J. W. 1998. Similarity judgments in choice under uncertainty: A reinterpretation of the predictions of regret theory. *Management Science* **44** 659-672.

Loomes, G. 2005. Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics* **8** 301-323.

Loomes, G. 2010. Modeling choice and valuation in decision experiments. *Psychological Review* **117** 902-924.

Loomes, G., R. Sugden. 1982. Regret theory: An alternative theory of rational choice. *Economic Journal* **92** 805-824.

Loomes, G., R. Sugden. 1987. Some implications of a more general form of regret theory. *Journal of Economic Theory* **41** 270-287.

Loomes, G., C. Starmer, R. Sugden. 1991. Observing violations of transitivity by experimental methods. *Econometrica* **59** 425-439.

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

May, K. O. 1954. Intransitivity, utility, and the aggregation of preference patterns. *Econometrica* **22** 1-13.

Mellers, B. A., K. Biagini. 1994. Similarity and choice. *Psychological Review* **101** 505-518.

Myung, J. I., G. Karabatsos, G. I. Iverson. 2005. A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology* **40** 205-225.

Payne, J. W. 1976. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance* **16** 366-387.

Quiggin, J. 1981. Risk perception and risk aversion among Australian farmers. *Australian Journal of Agricultural Economics* **25** 160-169.

Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3** 323-343.

Regenwetter, M., J. Dana, C. P. Davis-Stober. 2011. Transitivity of preferences. *Psychological Review* **118** 42-56.

Rubinstein, A. 1988. Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory* **46** 145-153.

Sopher, B., G. Gigliotti. 1993. Intransitive cycles: Rational choice or random error? *Theory and Decision* **35** 311-336.

Starmer, C., R. Sugden. 1998. Testing alternative explanations of cyclical choices. *Economica* **65** 347-361.

Starmer, C., R. Sugden. 1993. Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty* **6** 235-254.

Stott, H. P. 2006. Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty* **32** 101-130.

Swait, J., W. Adamowicz. 2001. The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research* **28** 135-148.

Tversky, A. 1969. Intransitivity of preferences. *Psychological Review* **76** 31-48.

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

van Osch, S.,M.C., W. B. van den Hout, A. M. Stiggelbout. 2006. Exploring the reference point in prospect theory: Gambles for length of life. *Medical Decision Making* **26** 338-346.

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