# A Tailor-Made Test of Intransitive Choice

AURÉLIEN BAILLON
Erasmus School of Economics, Rotterdam, the Netherlands

HAN BLEICHRODT
Erasmus School of Economics, Rotterdam, the Netherlands

ALESSANDRA CILLO
Bocconi University, Milan, Italy

March 2012

**Abstract**

We performed a new test of intransitive choice based on individual measurements of regret theory, the most influential intransitive theory. 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 cycles and we could not reject the hypothesis that they were due to random error. Moreover, there was little evidence that regret affected people's choices. A possible explanation for the poor predictive performance of regret theory is that, unlike other non-expected utility models, it assumes that preferences are separable over states of nature. Our data suggest that to account for the violations of expected utility event-separability has to be relaxed.

KEYWORDS: Decision under risk, transitivity, regret theory.

ACKNOWLEDGEMENTS: Graham Loomes and Peter Wakker 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.

## 1. Introduction

Transitivity is a fundamental axiom of rational choice. It underlies most normative and descriptive theories of decision making, including expected utility and prospect theory (Savage 1954, Tversky and Kahneman 1992). There is wide agreement that transitivity is normative (for dissenting views see Loomes and Sugden 1982, Anand 1987, and Fishburn 1991), but its empirical status is less clear. Starting with May (1954) and Tversky (1969), several studies have observed that up to 30% of decision makers violate transitivity (Loomes et al. 1991, Loomes and Taylor 1992, Starmer and Sugden 1998, Starmer 1999, Humphrey 2001b, Day and Loomes 2010). Most of this evidence was inspired by regret theory (Bell 1982, Loomes and Sugden 1982), the most influential theory of intransitive choice.[1]

The observed violations of transitivity are controversial, however. It has been argued that they were primarily caused by random errors and that the actual proportion of intransitive preferences was negligible and not convincing enough to abandon transitive theories (Iverson and Falmagne 1985, Sopher and Gigliotti 1993, Luce 2000, Birnbaum and Gutierrez 2007, Birnbaum and Schmidt 2008, Birnbaum 2010, Birnbaum and Schmidt 2010, 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.

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 behavior. All subjects were then confronted with the same stimuli. An obvious drawback of this "one size fits all approach"

[1] Other intransitive models are Fishburn's (1982) skew-symmetric bilinear utility, similarity models (Rubinstein 1988, Leland 1994, Leland 1998), Bordley's (1992) context-dependent generalization of Viscusi's (1989) prospective reference theory, the stochastic difference model (González-Vallejo 2002), Bleichrodt and Schmidt's (2002) context-dependent model, the priority heuristic (Brandstätter et al. 2006), and Loomes' (2010) perceived relative argument model.

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 typically observed in empirical studies.

An alternative approach is to select a model of intransitive choice, elicit its parameters for each individual separately and then to use these to select the individual-specific stimuli that will lead to choice cycles according to the model. This "tailor-made approach" was unavailable until recently 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 utility functions representing these choices (e.g. Regenwetter et al. 2011, p.44).

Bleichrodt et al. (2010) recently developed a method to measure regret theory. In this article we use their method to design individual-specific tests of transitivity. This allows us to test the prediction that decision makers who are averse to regret will display specific violations of transitivity, termed *regret cycles*. Our tests control for event-splitting effects, under which an event gets more weight if it is split into two subevents, which have confounded previous tests of regret (Starmer and Sugden 1993).

In spite of using individual-specific tests designed to uncover regret cycles, we found little evidence of intransitive choice and could not reject the null hypothesis that the observed regret cycles were due to random error. What is more, we observed little evidence that regret affected subjects' choices. Our results suggest that even though regret is an intuitive notion, regret theory does not capture it well. In particular, regret theory's assumption that preferences are separable over events appears empirically unrealistic.

## 2. Regret theory

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, j = 1 ,…, n, and the events $E_j$ partition the state space. The decision maker's problem is to choose 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 the $p_j$ in $(p_1,x_1;\ldots;p_n,x_n)$ and the $p_j$ 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 ~ denote strict preference and indifference. According to regret theory (Loomes and Sugden 1982), 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[2]

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

The function u in Eq. (1) is a real-valued *utility function*, unique up to scale and unit (Fishburn 1982). Q is a real-valued and strictly increasing function that satisfies *symmetry*: for all x, $-Q(x) = Q(-x)$. Symmetry implies that $Q(0) = 0$. Q is a ratio scale, i.e. unique up to the unit of measurement.

To illustrate Eq. (1), suppose that the decision maker chooses $(p_1,x_1;\ldots;p_n,x_n)$ over $(p_1,y_1;\ldots;p_n,y_n)$ and state j occurs. The decision maker then receives $x_j$ but he would have received $y_j$ had he chosen $(p_1,y_1;\ldots;p_n,y_n)$. According to Eq. (1) the psychological pleasure that the decision maker derives from $x_j$ depends not only on the utility of $x_j$, but also on the utility of $y_j$. If $y_j$ is more desirable than $x_j$, the decision maker experiences regret. If $x_j$ is more desirable, he experiences rejoicing. The function Q reflects the impact of regret. If Q is linear then regret

[2] Loomes and Sugden (1987) proposed a more general version of regret theory. The conclusions from our experiments are also valid under this version.

theory is equivalent to expected utility. The distinctive predictions of regret theory result from the convexity of Q(x) for all positive values of x, which is termed *regret aversion* in the literature. Together with Q(0) = 0, convexity implies that for all positive utility amounts x,y, $Q(x+y) \geq Q(x) + Q(y)$.

It is well-known that people who are regret averse can display intransitive choice behavior (e.g. Loomes and Sugden 1982, p.815). An example is provided in the next Section. Obviously if Q is linear, the special case of expected utility, intransitivities are impossible. Intransitivities are generated by the curvature of Q.

## 3. Predicting intransitivities

Our procedure for predicting intransitivities consisted of three steps. In the first step we used the trade-off method of Wakker and Deneffe (1996) to elicit a *standard sequence* of money amounts $x_0, \ldots, x_5$, such that $u(x_{j+1}) - u(x_j) = u(x_1) - u(x_0)$, j = 1, …., 4. Two gauge outcomes R and r (R > r), a probability p, and a starting outcome $x_0$ were selected and we elicited $x_1$ such that $(p, x_1; 1-p, r) \sim (p, x_0; 1-p, R)$. According to Eq. (1) this indifference implies that

$$Q(u(x_1) - u(x_0)) = \frac{1-p}{p} Q(u(R) - u(r)). \tag{2}$$

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

$$Q(u(x_2) - u(x_1)) = \frac{1-p}{p} Q(u(R) - u(r)). \tag{3}$$

Because Q is strictly increasing, it follows from (2) and (3) that $u(x_2) - u(x_1) = u(x_1) - u(x_0)$. Continuing this procedure we obtained the desired standard sequence. Because u is an interval scale, we could set $u(x_0) = 0$ and $u(x_5) = 1$. Then $u(x_j) = j/5$, j = 0 ,…, 5.

In the second step we measured the regret function Q. 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, $x_0$) ~ (p, $x_3$; 1−p, $z_p$). Because Q is a ratio scale we could set Q(⅕) = 1. Using Eq. (1) and the scaling of u, we thus obtained that

$$Q(u(z_p)) = \frac{p}{1-p}\,. \tag{4}$$

The subscript p in $z_p$ serves as a reminder that the value of Q depends on the probability used in the elicitation. The utility of $z_p$ was generally unknown, but could be estimated using elements of the standard sequence. By varying p we could measure as many points of Q as desired.

The measurements of the first step were used in the third step 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.

The measurements of the second step were used to predict how subjects would choose in (i)-(iii). Suppose a decision maker is regret averse and, therefore has a convex Q. According to Eq. (1) the comparison between A and B yields,

$$A \succcurlyeq B \Leftrightarrow ⅓\, Q(u(x_4) - u(x_2)) + ⅓ * Q(u(x_4) - u(x_5)) + ⅓ * Q(u(x_2) - u(x_3)) \geq 0. \tag{5}$$

By the symmetry of Q and the scaling of u Eq. (5) can be written as

$$A \succcurlyeq B \Leftrightarrow Q(⅖) - 2 * Q(⅕) \geq 0, \tag{6}$$

The right hand side of Eq.(6) is positive if Q is convex. Consequently, the decision maker will choose A over B if and only if Q is convex, i.e. if and only if the decision maker is regret averse. In a similar vein, it can be shown that regret aversion is equivalent to a choice of B over

C and a choice of C over A. Hence, regret aversion entails the *regret cycle* ABC.[3] Contrarily, concave Q, which corresponds to a type of behavior that we may call rejoicing seeking, implies the *rejoicing cycle* BCA.

## 4. Experiment

Section 3 demonstrated that violations of transitivity are closely connected with the non-linearity of the regret function Q, measured in the second step. Our tests were designed such that deviations from linearity should produce intransitive choices according to regret theory. The aim of our experiment was to explore whether this relationship between the convexity [concavity] of Q and the number of regret cycles [rejoicing cycles] could indeed be empirically observed.

### Method

#### *Subjects*

Subjects were 54 students (22 male) aged between 18 and 33 (median age 21). They were paid a base fee of €10 and, in addition, each subject had a 10% chance to be selected to play out one choice for real at the experiment's conclusion.

#### *Procedures*

The experiment was 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 needed in the first two steps of our procedure were elicited through a series of choices that "zoomed in" on subjects' indifference values. This iteration procedure is explained in Appendix A. All choices were presented via a computer interface, an

[3] The notation ABC stands for "A chosen over B, B chosen over C, and C chosen over A."

example of which is depicted in Figure 1. 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 their choice, the next question was displayed. If not, the choice was displayed anew. The confirmation question aimed to reduce the impact of response errors. Acts were presented both in a matrix format and as probability wheels.

**Figure 1: Example of the computer interface**

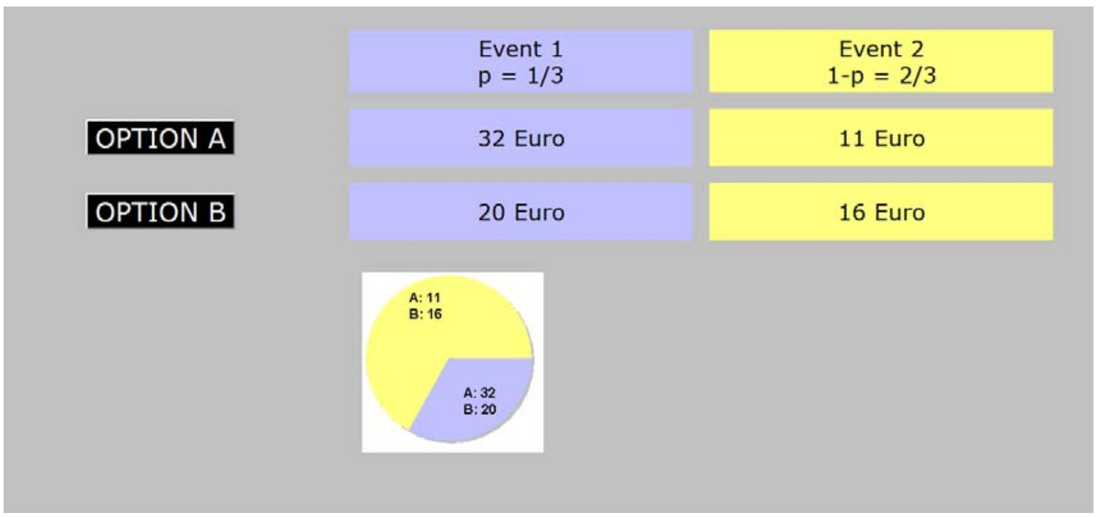



Table 1 summarizes the questions asked in the first two steps of our method. We used the same stimuli as in one of the two measurements of Bleichrodt et al. (2010) to test the robustness of the measurements.

**Table 1: Summary of the measurement of u and Q**

| | Elicited outcome | Indifference | Implication |
|---|---|---|---|
| First step | $x_j$, $j = 1,\ldots,5$ | $(⅓, x_j; ⅔, 11) \sim (⅓, x_{j-1}; ⅔, 16)$ | $u(x_j) = j/5$ |
| Second step | $z_p$, $p = ¼, ⅖, ⅗, ¾$ | $(p, x_4; 1-p, x_0) \sim (p, x_3; 1-p, z_p)$ | $Q(u(z_p)) = \frac{p}{1-p}$ |

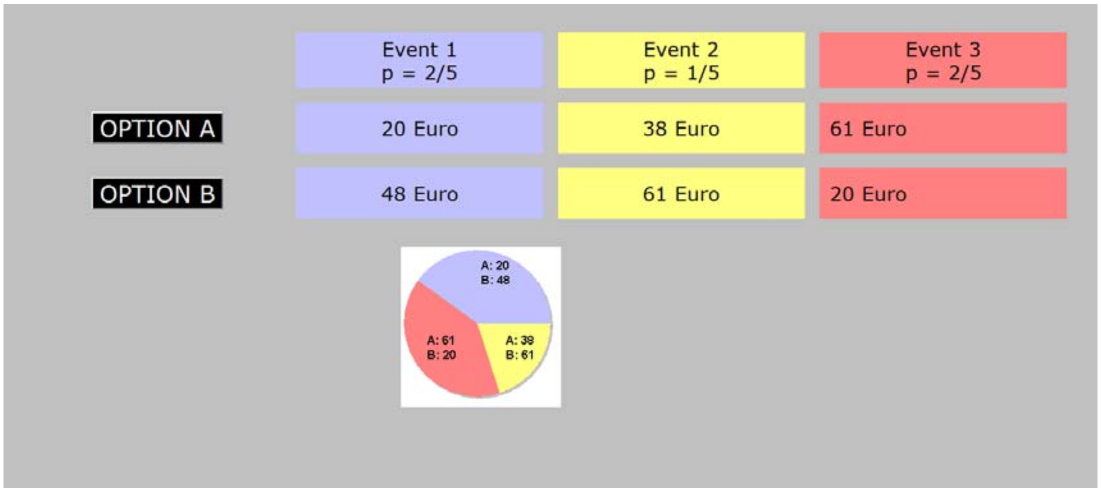
The choices used in the third step are summarized in Table 2. For each triple, regret aversion predicts the intransitive cycle ABC. We used two sets of tests, in tests 1 to 7 the probabilities of the different outcomes were the same and equal to ⅓, and in tests 8 to 14 the probabilities differed and were equal to ⅕, ⅖, and ⅖. We used these two sets to have more tests of transitivity and 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.

**Table 2: 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 that part of the support for regret theory could be explained by event-splitting effects that occur when the same outcome is received under two different states. An event with a given probability is weighted more heavily when it is split into two subevents than when it is presented as a single event (Starmer and Sugden 1993, Humphrey 1995). However, Starmer and Sugden (1998) and Humphrey (2001b) concluded that the regret cycles that they observed could not be explained by event-splitting effects.

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




The design of our study was such that event-splitting effects could not affect the results. As shown in Figure 2, the number of states was always the same and subevents were not combined or split. Humphrey (2001a) found that event-splitting effects were primarily due to a preference for more positive outcomes. In our tests the number of positive outcomes was always the same. Moreover, there is some evidence that the outcome zero can have an especially negative impact. We therefore avoided the outcome zero in our tests and used $x_0$ = €20.

Prior to the actual experiment, subjects answered two training questions. Following these questions, we elicited the outcome x such that (⅓, x; ⅔, 11) ~ (⅓, 40; ⅔, 16). These questions were not used in the final analyses and only served to monitor for any confusion about the experimental instructions. If so, the experimenter explained the task again.

Each experimental session started with the elicitation of the elements of the standard sequence because these were inputs in the other two steps. The order of the other steps, the elicitation of the regret function and the transitivity tests, was randomized. So either the regret function was elicited first or the transitivity tests were performed first. The measurement of the standard sequence had to be performed in a fixed order, but the order of the choices in the other steps was random. Throughout the experiment, we counterbalanced which option was A and which was B. We also varied between subjects in which event column (right or left) the stimuli changed during the bisection process. Hence, for half the subjects the change occurred in the left column and for the other half in the right column.

To test for response error, 13 choices were repeated. At the conclusion of the first part, we repeated the third choice of the iteration process for two randomly selected questions. At the conclusion of the second part we repeated the third choice of the iteration process for $z_{2/5}$, $z_{3/5}$, and $z_{3/4}$. We chose to repeat the third choice because subjects were generally close to their indifference value in this choice and response errors were more likely. We also repeated 8 randomly determined choices from the third part.

We finally repeated the entire elicitation of $x_1$, the first element of the standard sequence, and $z_{1/4}$. These repetitions served two purposes. First, they gave insight into the error made in the elicited indifference values. Second, a difference between the two elicited values of $x_1$ would indicate strategic responding (Bleichrodt et al. 2010), an often-raised objection against the trade-off method. In the trade-off method earlier answers are used as inputs in subsequent questions. By overstating their answers subjects could make later questions more attractive. Because we used a choice-based elicitation procedure, subjects 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 $x_1$ should exceed the original measurement.

After making all choices, subjects opened a ticket drawn from a nontransparent bag containing 10 tickets, one of which was a winning ticket. If the subject had a non-winning ticket, he received €10 and the experiment was over. If he had the winning ticket, we randomly selected the choice to be played for real. This selection was performed in front of the subject. The subject then played his preferred option in the selected choice where his payoffs were determined by another random draw.

***Analysis***

We computed two measures to determine a subject's curvature of utility. First we measured the area under the subject's normalized utility function, which is obtained by dividing all elements of the standard sequence by $x_5$, its final element. If the area under the curve exceeds 0.5, utility is concave, if it is less than 0.5, utility is convex.

We also fitted a power function to each subject's utility function. Concave utility corresponds with a power coefficient less than 1, convex utility with a power coefficient exceeding 1. We only classified a subject as concave [convex] if the power coefficient was smaller [larger] than 1 at the 5% significance level.

To measure the regret function we needed to know the utilities of $z_{1/4}$, $z_{2/5}$, $z_{3/5}$, and $z_{3/4}$ for each subject. Because these were generally not part of the elicited standard sequences, their utility was estimated through interpolation using the elements of the standard sequences. We both used linear interpolation and interpolation based on the fitted individual power coefficients. The results were similar and we will only report the results based on linear interpolation.

Curvature of the regret function was determined using the same two measures as for utility: the area under the normalized regret function and the estimated power coefficient. For the classification based on the power coefficients we used again the criterion that a subject's regret function was convex [concave] if the estimated power coefficient was larger [smaller] than 1 at the 5% level. The scaling point $Q(\frac{1}{5}) = 1$ was often an outlier in the individual measurements of Q. We therefore analyzed the data both with and without this point. The results were similar unless stated otherwise.

To test the robustness of our findings we analyzed the data both using the original and the repeated measurement of $z_{\frac{1}{4}}$. We hypothesized that the latter might be less subject to error, because the question occurred later in the experiment and subjects might have had the opportunity to learn more about their preferences for the kind of acts used to measure Q. However, the results were similar and we will only report those computed using the original measurement of $z_{\frac{1}{4}}$.

We analyzed the transitivity tests both for all subjects and for those subjects who made the same choice in at least 75% of the replications in the third part. Reversing a choice in more than a quarter of the choices might indicate confusion or random responding and we, therefore, explored whether the more consistent subjects were more likely to behave according to regret theory. However, this hypothesis was not supported; we observed no differences between the two samples and we will only report the results for the full sample.

We used Birnbaum's true and error model (Birnbaum and Gutierrez 2007, Birnbaum 2011) to explore whether possible violations of transitivity were due to response error. This model is neutral towards transitivity and does not assume beforehand that transitivity holds. Let $p_{ABC}$ denote the probability of a true pattern ABC, $p_{ABA}$ the probability of a true pattern ABA, etc. Then according to the true and error model the probability P(ABC) of observing the pattern ABC is:

$$P(ABC) = p_{ABC}(1-e_1)(1-e_2)(1-e_3) + p_{ABA}(1-e_1)(1-e_2)e_3 + \ldots + p_{BCA}e_1e_2e_3, \qquad (7)$$

where $e_1$, $e_2$, and $e_3$ are the probabilities of making errors in the choices between A and B, B and C, and A and C respectively. 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,…. Parameters are 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. It is of interest to note that the errors can be different across choices. The model therefore addresses a point of criticism that was raised against the error model of Harless and Camerer (1994) who 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 commits an error.

To analyze the relation between regret aversion and the number of choices made according to regret theory we estimated a random effects probit model with subject-specific errors. This allows individual errors to be correlated across choices. Subjects often follow a common strategy in answering the various choices and this automatically implies that the different choices are not independent. We also estimated a mixed model with subject as a random factor and choice as a fixed factor and a SUR (seemingly unrelated regression) model in which the impact of regret was choice-specific. The advantage of the SUR model is that individual preferences are allowed to vary across choices, as in the error models of Loomes and Sugden (1995) and Regenwetter et al. (2011), but unlike these models it does not assume that choices are made independently.

## 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 $z_{¼}$). $z_{¼}z_{¼}$ The absence of a difference between the two measurements of $x_1$ suggests that measurements using the trade-off method are unaffected by strategic responding.

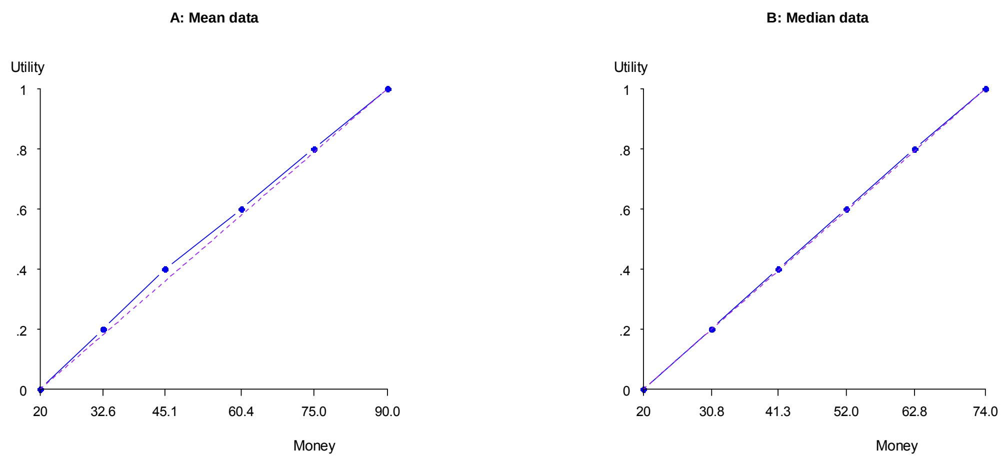
### *Utility*

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 are likely to reflect confusion and we, therefore, excluded these subjects from the remaining analyses.[4]

Utility was close to linear both for the pooled and for the individual data. Figure 3 shows the utility function based on the mean and the median data with the dotted line indicating linear utility. Consistent with linear utility, we could not reject the hypothesis that the differences between successive elements of the standard sequences were equal (repeated measures ANOVA, $p = 0.08$).[5] On the other hand, the estimated power coefficients of 0.96 (se = 0.01) for the median data and 0.83 (se = 0.02) for the mean data indicated some concavity of utility. Both coefficients differed from 1, the case of linear utility ($p = 0.005$ for the median data and $p < 0.001$ for the mean data).

[4]Keeping them in did not affect the conclusions.
[5]The mean [median] step sizes were 12.55, 12.50, 15.35, 14.56, and 14.62 [10.75, 10, 10.50, 10, and 10.25].

**Figure 3: Utility based on the mean and on the median data**




The individual subject data confirmed the absence of a dominant shape of utility. The number of concave subjects did not differ from the number of convex subjects. Based on the area under the utility function, there were 24 subjects with concave utility and 25 subjects with convex utility (binomial test, $p = 1$). Consistent with linear utility, the median area under the normalized utility function was 0.50 (mean area 0.51) and did not differ from 0.50 (Wilcoxon test, $p = 0.30$).

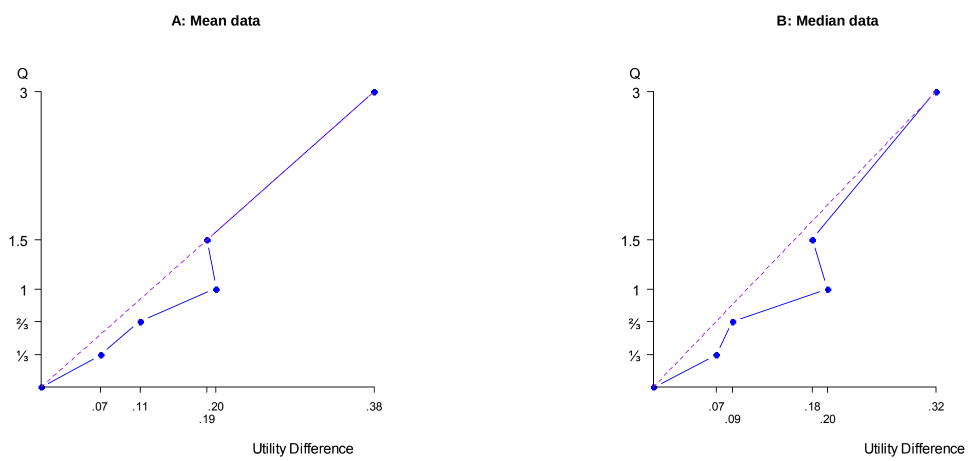
The median of the individual power coefficients was 0.98 (mean 0.97), which did not differ from 1 (Wilcoxon test, $p = 0.88$). There were 16 subjects whose fitted power coefficient was significantly lower than 1, corresponding with concave utility, and 14 subjects for whom it significantly exceeded 1, corresponding with convex utility (binomial test, $p = 0.86$). For the remaining 22 subjects the estimated utility function did not differ from linearity.

***Regret***

For one subject, the original measurement of $z_{¼}$ exceeded the repeated measurement by more than three times the standard deviation, probably reflecting confusion. We therefore excluded this subject from the analyses.

Several subjects violated monotonicity. Whilst most of these violations fell within the margins of error and had no discernible impact on the results, in some cases the violations were too large to be attributable to error and had a substantial effect. We excluded those violations that were larger than the standard deviation of the difference between the original and the repeated measurement of $z_{¼}$. This occurred in 3.8% of the choices. The violations of monotonicity were not obvious to detect for the subjects and the violation rate was low compared with previously observed violations of nontransparent monotonicity (e.g. Birnbaum 2008). No subject violated monotonicity more than once.

**Figure 4: Regret based on the mean and the median data**



The aggregate data were consistent with regret aversion. Figure 4 shows that the estimated regret functions based on the mean and on the median data were convex, in agreement with regret aversion. The pattern was not smooth, however, and the scaling point (⅕,1) was an outlier.

Parametric fitting indicated that regret aversion was not due to chance. The fitted power coefficients are 1.27 (standard error = .07) based on the mean data and 1.39 (standard error = 0.13) based on the median data. Both estimates differ from 1 ($p = .011$ for the mean data and $p = .018$ for the median data), the case of expected utility.

At the individual level the evidence for regret aversion was mixed. Based on the area under the regret function, there was clear evidence for regret aversion. Thirty-eight subjects were regret averse versus 12 rejoicing seeking (binomial test, $p < 0.001$). The median area under the regret curve was 0.41 (mean 0.33), which was different from 0.50 (Wilcoxon test, $p < .001$), the case of expected utility.

Based on the power coefficients, however, the evidence for regret aversion was less obvious. The median of the individual power coefficients was 1.09 (mean = 1.13), which did not differ from 1, the case of expected utility (Wilcoxon test, $p = 0.10$). [7] There were 13 subjects with a convex regret function (power coefficient exceeding 1) and 5 with a concave function. The difference was marginally significant (binomial test, $p = 0.09$). Thirty-three subjects could not be classified.

[6] If we exclude the scaling point (⅕, 1), the median increases to 0.43 and the mean to 0.38.

[7] Using the repeated instead of the original measurement of $z_{.25}$ the median power coefficient increased to 1.15 and was significantly different from 1 at the 5% level.

There was more evidence for regret aversion when the scaling point (⅕, 1) was excluded. Then there were 19 subjects with a convex regret function and 5 with a concave function and the number of regret averse subjects exceeded the number of rejoicing seeking subjects (binomial test, $p = 0.006$).

***Comparison with Bleichrodt et al. (2010)***

We used the same stimuli as in one of the two measurements in Bleichrodt et al. (2010). The results on utility were similar. The elicited standard sequences did not differ between the two studies ($p > 0.10$ in all pairwise tests) and equality of the fitted power coefficients based on the median data could not be rejected either (t-test, $p = 0.48$). However, based on the mean data the power coefficient was lower in our study (0.83 versus 0.98), indicating more concavity (t-test, $p < 0.001$). Based on the area under the utility function, the proportions of concave and convex subjects were similar in the two studies. However, Bleichrodt et al. 2010) found more concavity based on the individual power coefficients.

Regret aversion was somewhat less pronounced in our study. It was comparable based on the median data ($p = 0.44$), but lower based on the mean data ($p < 0.001$). The areas under the regret curves did not differ (Wilcoxon test, $p = 0.11$; t-test, $p = 0.76$) when all points were included, but it was higher in our study, reflecting less regret aversion, when the scaling point (⅕, 1) was excluded (Wilcoxon test, $p = 0.04$). The classification of the individual subjects was similar based on the area under the regret curve, but there was more regret aversion in Bleichrodt et al. (2010) based on the power coefficients.

[8] The median was still 1.09, the mean increased to 1.18.

*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. A first look at the table reveals that cycles were rare. The proportion of cycles is comparable to Birnbaum and Schmidt (2008) and Loomes (2010). It is lower than in Loomes et al. (1991) and Starmer and Sugden (1998) who observed intransitivity rates around 20%. More strikingly, unlike these studies, we did not observe that regret cycles were more common than rejoicing cycles.

**Table 3: Response patterns in the transitivity tests in %**

| 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 |

This conclusion 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 in the direction consistent with regret aversion, 12 subjects who exhibited at least one cycle only in the direction consistent with rejoicing seeking, and 6 subjects who cycled in both directions.

We observed no correlation between the number of regret cycles and regret aversion. This held for both measures of regret aversion, the area under the regret curve and the estimated power coefficients.The dearth of regret cycles was not due to response errors. The estimated parameters of the true and error model indicated that the true proportion of regret cycles differed from 0 in only 3 tests and were low also in those tests (2.8%, 1.1%, and 9.2%). The estimated true proportion of rejoicing cycles differed from 0 in 4 tests (3.2%, 4.2%, 1.6%, and 6.2%). What is more, the fit of a purely transitive model (i.e. a model in which we impose the restriction that the proportions of regret cycles and of rejoicing cycles are equal to zero) was not significantly worse, meaning that we could not reject the hypothesis that everyone was transitive. 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%.[9] The finding that a common error gave a good approximation to the data may have been caused by the use of tailor-made tests, which were similar in structure.

Table 3 presents the results of all subjects, including those who did not deviate from expected utility, i.e. whose elicited Q did not differ from linearity. Excluding these subjects did not affect the proportion of intransitive subjects in a substantive way and there was still no evidence of the predicted asymmetry in the number of regret and rejoicing cycles. Intransitive choice patterns were rare and not asymmetric.

[9] At the 5% we could reject it once (p = 0.045).

This did not mean, however, that expected utility held. Because expected utility is a special case of regret theory, it is also true that the difference in utility between any two successive elements of the standard sequence elicited in the first part of our experiment is constant. Because expected utility evaluates probabilities linearly, subjects should be indifferent in all 42 choices except for random error. Consequently, all 8 response patterns reported in Table 3 should be equally likely if expected utility held. 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.

***Predicting choices***

Even though the transitivity tests provided no support for the existence of regret cycles, it might still be that regret affected subjects' choices. If so, we should observe that the more convex a subject's Q in the second step of our measurement, the more likely he is to choose according to regret aversion. Hence, we should observe a negative relation between the area under the regret curve and the probability of choosing according to regret aversion (the lower the area the more regret averse is a subject). We indeed found this: the correlation was negative ($\rho = -0.20$) and marginally significant ($p = 0.08$ by a one-sided test[10]). When regret aversion was measured by the power coefficient, we should expect a positive correlation, because a higher power coefficient reflects more regret aversion. However, this was not observed; the correlation did not differ from zero.

Table 4 summarizes the results from the random-effects probit estimation. The mixed model and the SUR model gave similar results, but their fit was slightly worse based on the Akaike information criterion (AIC).When regret aversion was measured by the area under Q, the coefficient had the predicted negative sign and was marginally significant ($p = 0.07$, one-sided

[10] We used a one-sided test because regret theory makes a specific prediction about the direction of the correlation.

test). However, when regret was measured by the estimated power coefficient, it had the wrong sign ($p = 0.01$).

**Table 4: Estimation results for the random-effects probit model (standard errors in parentheses)**

| | Intercept | Regret |
|---|---|---|
| Regret measured by area | 0.015<br>(0.056) | −0.188<br>(0.128) |
| Regret measured by power coefficient | −0.005<br>(0.036) | −0.040<br>(0.018) |

The mixed results on regret might be caused by the omission of other variables that affected subjects' choices. One 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. Another possibility is that 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 would have made a mistake and had 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 this error of too low a value of $x_1$ would be reflected in a relatively low expected value of C. However, the coefficient for regret was not affected when the difference in expected value between the two acts in a choice was included as an additional regressor in the probit model. In agreement with regret theory, the impact of expected value was nonsignificant.[11]

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

The final possibility that we considered 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 regret theory 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, regret effects are more likely to show up anyhow. We therefore split the transitivity tests into those choices in which the utility difference between the outcomes exceeded ½ for at least one state and those for which this was not the case. Again, this did not affect the results; the impact of regret was similar in the two subsamples. Using other utility thresholds than ½ had no effect either.

## 6. Discussion

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. It implies that there are no grounds to abandon transitivity as a descriptive axiom of choice. Our data violated regret theory. Not only did we find no evidence that regret cycles were more common than the opposite rejoicing cycles, but

we also found only limited evidence that regret had an impact on individual choices. We explored the impact of several potentially confounding variables, but the evidence against regret theory appeared robust.

One explanation for the lack of support for regret theory is that we used two-outcome acts in the measurement of regret theory, 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). To illustrate, the modal response in each of our tests could be predicted by the following heuristic rule: choose the act that offers the highest minimum outcome, if this minimum is equal choose the one which has the lowest probability of this minimum outcome, if this is also equal choose the one with the highest second lowest outcome. This rule resembles the priority heuristic (Brandstätter et al. 2006), even though the priority heuristic was only formulated for two-outcome acts.

The interpretation of our results as evidence that the transition from two-outcome to three-outcome acts is problematic, is consistent with Gonzalez and Wu (2003) who observed similar problems for prospect theory. They measured prospect theory's weighting and value functions from fitting two-outcome act cash equivalents and applied these to predict cash equivalents for three-outcome acts. They found systematic under-prediction of cash equivalents using cumulative prospect theory (Tversky and Kahneman 1992) and systematic over-prediction using original prospect theory (Kahneman and Tversky 1979). We obtained some indication that prospect theory had similar problems in accounting for our data. Prospect theory with Tversky and Kahneman's (1992) estimated probability weighting function was rejected in all but one transitivity test by the true and error model. Of course, this analysis is only tentative and by imposing one common probability weighting function on all of our subjects it ignores the substantial heterogeneity that was present in our data. A convincing test of prospect theory's predictive power should measure each individual's probability weighting function separately, but unfortunately we had insufficient to do so as for regret.

An alternative explanation for our findings is that regret theory does not describe people's preferences well. Regret is an intuitive notion and evidence from neuroscience suggests that it plays a fundamental role in regulating choice behavior (Camille et al. 2004). However, regret theory may not be the appropriate way to model this. Expected utility assumes both separability across events and transitivity and thereby rules out both within-act interactions and between-act interactions. Nonexpected utility models can be broadly divided into two categories. One category has retained transitivity, but has given up event-separability by assigning significance to the ranking of the outcomes of an act. In other words, this group of theories, which includes prospect theory for example, has allowed within-act interactions, but has ruled out between-act interactions. The other category has given up transitivity (allowed between-act interactions), but has retained event-separability (ruled out within-act interactions). Regret theory belongs to this second category. In regret theory the ranking of the outcomes plays no role. However, as we pointed out above when giving a (priority) heuristic explanation of our results, it appears that the ranking of the outcomes was relevant in our tests.

Regret theory 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 allows between-act interactions on both the probability and the outcome dimension, but retains event-separability. We show in Appendix B that our data violate PRAM as well. They suggest that allowing for between-act interactions while retaining event-separability is not a viable modeling strategy. To account for the violations of expected utility, giving up event-separability appears necessary.

## Appendix A: Method for measuring the indifference values

To elicit the standard sequence of outcomes in the first step 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).[12] 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.

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

**Table A1: p = ⅓: 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 for which indifference held between A = $(p,x_4; 1-p,20)$ and B = $(p,x_3; 1-p,z)$[13] 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 and the program ensured that z was never equal to $x_3$ to avoid the possibility of event-splitting effects. The initial stimulus z was a random integer in the range $[z_{EV}-3, z_{EV}+3]$ where $z_{EV}$ is the value of z that makes A and B equal in expected value with the restriction that z could not be less than €20. There were two possible scenarios:

(i) As long as A was chosen we increased z 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 by this amount until B was chosen. Then we decreased z by D/2. If A [B] was subsequently chosen we increased [decreased] z by D/4, etc. A special case occurred if the difference between z and $x_4$ (for $p \leq ½$) or between z and $x_5$ (for $p > ½$) was less than 5. Then we increased z by 10 and subsequently kept increasing z by 5 until B was chosen. Then we decreased z by 3.

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

(ii) If B was chosen we decreased z by D′=(z − 20)/2 until A was chosen. We then increased z by D′/2. If A [B] was subsequently chosen we increased [decreased] z 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 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. 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 when $x_4$ = 70 and $x_3$ = 50**

| Iteration | z | Choice |
|---|---|---|
| 1 | 26 | A |
| 2 | 44 | B |
| 3 | 35 | B |
| 4 | 31 | B |
| 5 | 29 | A |

**Appendix B: Illustration that the data also violate Loomes' PRAM model.**

Because we only used acts with at most three different states of nature in our experiment, we will explain Loomes' (2010) perceived relative argument model (PRAM) for such acts. Let X = $(p_1, x_1; p_2, x_2; p_3, x_3)$ and Y = $(p_1, y_1; p_2, y_2; p_3, y_3)$. According to PRAM:

$$X \succcurlyeq Y \Leftrightarrow \varphi(b_X, b_Y) \geq \xi(u_Y, u_X). \tag{A1}$$

In Eq. (A1), φ reflects the perceived advantage of X versus Y on the probability dimension, whereas ξ reflects the perceived advantage of 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. Loomes (2010) assumes that

$$\varphi( b_X, b_Y) = (b_Y/b_X)^{(b_X+b_Y)^{\alpha}}, \tag{A2}$$

where α is a person-specific variable whose value may vary from one individual to another. Loomes further assumes that there exists a real-valued utility function u defined over the set of outcomes.[14] Then $u_Y$ denotes the advantage that Y has versus X in terms of utility and $u_X$ denotes the advantage that X has versus Y in terms of utility. 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) does not explain how to define $u_X$ and $u_Y$ when the equality $u(x_i) - u(y_i) = u(x_j) - u(y_j)$ does not obtain for any $i,j \in \{1,2,3\}$ $i \neq j$. Loomes (2010) further assumes that

$$\xi(u_Y, u_x)= (u_Y/u_X)^{\delta}, \text{ where } \delta \geq 1. \tag{A3}$$

In the first step of our experiment, the elicitation of the standard sequences we established indifference between $(p,x_{j+1}; 1-p,r)$ and $(p,x_j; 1-p,R)$. According to PRAM this indifference implies that:

---

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

$$\left(\frac{u(x_{j+1}) - u(x_j)}{u(R) - u(r)}\right)^{\delta} = \frac{1-p}{p}. \tag{A4}$$

Consequently, $u(x_{j+1}) - u(x_j) = u(x_1) - u(x_0)$, $j = 1,\ldots,4$, and the utility difference of successive elements of the standard sequence is also constant under PRAM. Thus, we can scale u such that $u(x_0) = 0$ and $u(x_5) = 1$. It follows that $u(x_j) = j/5$.

Because the elements of the standard sequence are equally spaced in terms of utility, PRAM makes specific predictions in several of our transitivity tests. Consider, for example, the comparison between A and B in our first test of transitivity. According to PRAM,

$$A \succcurlyeq B \Leftrightarrow \left(\frac{u(x_2) - u(x_0)}{u(x_3) - u(x_2)}\right)^{\delta} = (⅖/⅕)^{\delta} = 2^{\delta} \geq ⅔/⅓ = 2. \tag{A5}$$

It follows from (A5) that $A \succcurlyeq B$ if and only if $\delta > 1$. Likewise PRAM predicts that $B \succcurlyeq C$ iff $\delta > 1$, because $u(x_4) - u(x_3) = u(x_1) - u(x_0)$, and that $C \succcurlyeq A$ iff $\delta > 1$. Hence, unless $\delta = 1$ and the subject is indifferent between A, B, and C, we should observe the intransitive cycle ABC. A similar analysis shows that PRAM also predicts the cycle ABC in transitivity tests 2, 3, 6, 8, and 9 if $\delta > 1$. In the other transitivity tests there were three unequal utility differences and, as explained above, it is not clear how to analyze these under PRAM.

As shown in the main text, our data are inconsistent with the existence of systematic ABC cycles. Then the only possibility left is that $\delta = 1$ in which the PRAM model makes the same predictions as expected utility, namely that all response patterns are equally likely. However, when we imposed this restriction in the true and error model we could reject it at the 1% level in 5 out of 6 tests. The only exception was the third transitivity test.

A similar analysis as in the main paper also shows that there is no evidence that subjects for whom $\delta$ was larger were more likely to choose according to PRAM's predictions in individual choices. To see this, note that the second part of our experiment, in which we elicited indifferences $(p,x_4; 1-p,x_0) \sim (p,x_3; 1-p,z_p)$ implies in terms of the PRAM model that

$$(5u(z_p))^{\delta} = \frac{p}{1-p}\,. \tag{A6}$$

In terms of regret theory this indifference implies that

$$Q(u(z_p)) = \frac{p}{1-p}\,, \tag{A7}$$

And in our parametric analysis we estimated Q using a power function

$$Q(u(z_p)) = u(z_p)^{\beta}. \tag{A6}$$

By the uniqueness properties of the power function (Wakker 2008) it follows immediately that the estimations of β and δ should produce similar information. However, in our main analysis we showed subjects for whom β was larger, and who consequently were more regret averse, were not more likely to choose the option predicted by regret theory. It follows that subjects for whom δ was larger were not more likely to choose according to PRAM.

## References


Anand, P. (1987). Are the preference axioms really rational? Theory and Decision, 23, 189-214.

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

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

Birnbaum, M. (2008). New paradoxes of risky decision making. Psychological Review, 115, 463-501.

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. & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semi-order. Organizational Behavior and Human Decision Processes, 104, 96-112.

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

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

Bleichrodt, H. & Schmidt, U. (2002). A context-dependent model of the gambling effect. Management Science, 48, 802-812.

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

Bordley, R. F. (1992). An intransitive expectations-based Bayesian variant of prospect theory. Journal of Risk and Uncertainty, 5, 127-144.

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

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, 1167-1170.

Day, B. & Loomes, G. (2010). Conflicting violations of transitivity and where they may lead us. Theory and Decision, 68, 233-242.

Fishburn, P. C. (1991). Nontransitive preferences in decision theory. Journal of Risk and Uncertainty, 4, 113-134.

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

Gonzalez, R. & Wu, G. (2003). Composition rules in original and cumulative prospect theory. Working Paper University of Michigan, Ann Arbor.

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

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

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

Humphrey, S. J. (2001b). Non-transitive choice: Event-splitting effects or framing effects? Economica, 68, 77-96.

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

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

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

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

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

Loomes, G. (2010). Modeling choice and valuation in decision experiments. Psychological Review, 117, 902.

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

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

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. C. & Taylor, C. (1992). Non-transitive preferences over gains and losses. Economic Journal, 102, 357-365.

Loomes, G. & Sugden, R. (1995). Incorporating a stochastic element into decision theories. European Economic Review, 39, 641-648.

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

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.

Myung, J. I., Karabatsos, G., Iverson, G. I. (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.

Regenwetter, M., Dana, J., Davis-Stober, C. P. (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.

Savage, L. J. (1954). The foundations of statistics. Wiley, New York.

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

Starmer, C. (1999). Cycling with rules of thumb: An experimental test for a new form of non-transitive behavior. Theory and Decision, 46, 141-158.

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

Starmer, C. & Sugden, R. (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. & Adamowicz, W. (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. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5, 297-323.

Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. Journal of Risk and Uncertainty, 2, 235-264.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. Health Economics, 17, 1329-1344.

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