Management Science

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

A Quantitative Measurement of Regret Theory

Han Bleichrodt, Alessandra Cillo, Enrico Diecidue,

MANAGEMENT SCIENCE
Vol. 56, No. 1, January 2010, pp. 161–175
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5601 | 0161
DOI 10.1287/mnsc.1090.1097

A Quantitative Measurement of Regret Theory

Han Bleichrodt
Erasmus School of Economics, Erasmus University, 3000 DR Rotterdam, The Netherlands, bleichrodt@ese.eur.nl

Alessandra Cillo
IESE Business School, 08034 Barcelona, Spain, alessandra.cillo@unibocconi.it

Enrico Diecidue
Decision Sciences Area, INSEAD, 77305 Fontainebleau Cedex, France, enrico.diecidue@insead.edu

This paper introduces a method to measure regret theory, a popular theory of decision under uncertainty. Regret theory allows for violations of transitivity, and it may seem paradoxical to quantitatively measure an intransitive theory. We adopt the trade-off method and show that it is robust to violations of transitivity. Our method makes no assumptions about the shape of the functions reflecting utility and regret. It can be performed at the individual level, taking account of preference heterogeneity. Our data support the main assumption of regret theory, that people are disproportionately averse to large regrets, even when event-splitting effects are controlled for. The findings are robust: similar results were obtained in two measurements using different stimuli. The data support the reliability of the trade-off method: its measurements could be replicated using different stimuli and were not susceptible to strategic responding.

Key words: regret theory; utility measurement; decision under uncertainty
History: Received May 20, 2008; accepted August 3, 2009, by George Wu, decision analysis. Published online in *Articles in Advance* November 6, 2009.

1. Introduction

Regret theory (Bell 1982, Loomes and Sugden 1982) is an important theory of decision under uncertainty. The theory has intuitive appeal, being based on the notion that people care not only about what they get but also about what they might have gotten had they chosen differently. There exists a large literature in psychology showing the importance of regret in shaping people's preferences under risk (e.g., Larrick 1993, Zeelenberg et al. 1996, Zeelenberg 1999). Regret theory has a relatively simple structure, being based on two functions only: a utility function capturing attitudes toward outcomes, and a function capturing the impact of regret. In spite of its simple structure, regret theory can account for many of the observed deviations from expected utility. The key to explaining these deviations is that decision makers are regret averse, the psychological intuition that people are disproportionally averse to large regrets. The distinguishing feature of regret theory is that it predicts violations of transitivity. These violations are a consequence of regret aversion. They have been confirmed experimentally (e.g., Loomes et al. 1991) and cannot easily be accommodated by the other main nonexpected utility theories, including prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992).

Regret theory also has real-world implications and can explain field data that are incompatible with expected utility. For example, Barberis et al. (2006), Gollier and Salanié (2006), and Muermann et al. (2006) apply regret to financial decisions, Braun and Muermann (2004) to insurance decisions, Perakis and Roels (2008) to the newsvendor model, and Feliz-Ozbay and Ozbay (2007) and Engelbrecht-Wiggans and Katok (2008) to auctions. A reason for this surging interest in regret is the difficulty that other non-expected utility models have in explaining some of the field data. Take the disposition effect, for instance, the widely documented finding that investors appear reluctant to realize losses but are eager to realize gains. It was commonly believed that the disposition effect could be explained by loss aversion, but Hens and Vlcek (2006) and Barberis and Xiong (2009) have shown that prospect theory can have trouble explaining it. By contrast, regret theory is consistent with the disposition effect provided that people are regret averse (Muermann and Volkman 2007).

To apply regret theory in practical decision analysis requires a feasible method to quantitatively measure it. No such method exists to date.[1] A reason for this absence may be that regret theory allows

[1] Loomes and Sugden (1982) suggest a method, but their method confounds regret aversion and event splitting.

for intransitive preferences. Combining quantification with intransitivities seems contradictory. This paper shows that they can be reconciled. We introduce a method to quantitatively measure regret theory based on Wakker and Deneffe's (1996) trade-off method. The trade-off method was originally introduced to measure utility under prospect theory, but we show that it can also be used to measure utility under regret theory. Hence, the trade-off method is not only robust to probability weighting, but also to intransitivities. The elicited utilities are then used to measure regret. Consequently, the paper shows that regret theory can be measured by well-known measurement techniques, and that its measurement is much easier than has commonly been thought.

Our measurement method is parameter free in the sense that it does not require any assumptions about the regret theory functional. An advantage of avoiding parametric assumptions is that the measurements are not confounded by violations of the parametric assumptions made. A second advantage of our method is that it can be performed at the individual level, taking account of heterogeneity in preferences. Individual measurements tend to vary substantially. To take this heterogeneity into account, decision analysis needs a method that can be applied at the level of the individual decision maker.

We implemented our method in an empirical investigation. The data, for the first time, make it possible to observe regret theory and to address the open empirical questions about whether regret aversion really exists and, if so, how strong it is.[2] Regret aversion has commonly been assumed in the theoretical literature on regret, but as yet there exists no unambiguous evidence that this assumption is justified. Several studies have provided qualitative support for regret aversion (for an overview, see Starmer 2000), but the validity of these studies was challenged by Starmer and Sugden (1993). They showed that almost all the empirical support for regret aversion could be explained by a heuristic called event splitting. According to event splitting, an event with a given probability is weighted more heavily if it is considered as two subevents than if it is considered as a single event (see also Humphrey 1995). Our study controlled for event splitting. Support for regret aversion was observed nevertheless. We included several tests to validate our method and to verify that what we measured was really regret aversion. None of these tests refuted regret aversion.

The data also allow answering a point of criticism that has been raised repeatedly against the trade-off method (e.g., Harrison and Rutström 2008). Because the trade-off method uses chained responses, i.e., previous responses are used as inputs in later questions, critics have pointed out that subjects have an incentive to answer strategically. We tested for strategic responding but found no evidence for it, providing additional support for the robustness of measurements by the trade-off method.

In what follows, §2 reviews regret theory. Section 3 describes our new measurement method. Section 4 describes the design of the study, and §5 its results. Section 6 discusses the main findings, and §7 concludes this paper.

[2] Di Cagno and Hey (1988) measured regret theory using Loomes and Sugden's (1982) method and, hence, their measurements confound regret aversion and event splitting.

2. Regret Theory

Let $\mathcal{S}$ denote a *state space*. Subsets of $\mathcal{S}$ are called *events*. A probability measure P is given over the set of events. An *act* is a function from $\mathcal{S}$ to the set of *outcomes*. Our measurement method, described in §3, requires that outcomes are real numbers. They are money amounts in the study in §4. Because our measurement method only uses acts with two different outcomes, attention will be restricted to such *binary acts*. We shall denote acts as $\alpha_p\beta$, which means that there is an event E with probability p such that α obtains under E, and β obtains under the complement of E. In what follows, if we compare two acts $\alpha_p\beta$ and $\gamma_p\delta$, it is implicitly assumed that p refers to the same event.

A preference relation $\succeq$ is given over the set of binary acts. The conventional notations $\succ$ and $\sim$ are used to denote strict preference and indifference respectively. We assume that higher outcomes are preferred to lower outcomes. For money amounts, this assumption is self-evident.

Consider two acts $\alpha_p\beta$ and $\gamma_p\delta$. The general formulation of regret theory proposed by Loomes and Sugden (1987) postulates a real-valued function Ψ such that

$$\alpha_p\beta \succeq \gamma_p\delta \;\Leftrightarrow\; p\Psi(\alpha,\gamma)+(1-p)\Psi(\beta,\delta)\geq 0. \tag{1}$$

The function $\Psi(\alpha,\gamma)$ can be interpreted as assigning a real-valued index to the net advantage of choosing $\alpha_p\beta$ rather than $\gamma_p\delta$ if event E with probability p obtains. The function Ψ is unique up to scale, i.e., it can be replaced by any other function $\Psi' = a\Psi$, $a > 0$, without affecting preferences, and satisfies the following three restrictions:

(i) The function Ψ is *strictly increasing* in its first argument: for any outcome γ, if $\alpha > \beta$, then $\Psi(\alpha,\gamma) > \Psi(\beta,\gamma)$.

(ii) The function Ψ is *skew symmetric*: for all α and β, $\Psi(\alpha,\beta) = -\Psi(\beta,\alpha)$.

(iii) For all $\alpha > \beta > \gamma$, $\Psi(\alpha,\gamma) > \Psi(\alpha,\beta) + \Psi(\beta,\gamma)$.

This property was labeled *convexity* in Loomes and Sugden (1987). It is also referred to as *regret aversion*. Skew symmetry entails that for all outcomes α, $\Psi(\alpha,\alpha)=0$. Expected utility is the special case of regret theory in which $\Psi(\alpha,\beta)=u(\alpha)-u(\beta)$, and u is a von Neumann–Morgenstern utility function. Fishburn's (1982) skew-symmetric bilinear theory resembles the general form of regret theory but assumes that preferences are defined over prospects and probability distributions over outcomes, rather than over acts.

Bell (1982, 1983) and Loomes and Sugden (1982) considered a restricted form of (1) in which

$$\Psi(\alpha,\beta)=Q(u(\alpha)-u(\beta)). \tag{2}$$

Bell (1982) referred to u in Equation (2) as a *value function* measuring strength of preference, or incremental value. Loomes and Sugden (1982) referred to u as a *choiceless utility function*, which reflects the utility the decision maker would derive from an outcome x if he experienced it without having chosen it. That is, both Bell (1982) and Loomes and Sugden (1982) refer to introspection.[3] Our method explained in §3 is entirely choice based and does not assume any other primitives like incremental value or choiceless utility. To emphasize this point, we will refer to u as a utility function in what follows.

The function Q in Equation (2) is strictly increasing and has the following symmetry property, implied by skew symmetry: for all α, $-Q(\alpha)=Q(-\alpha)$. Regret aversion, which generates the distinctive predictions of regret theory, implies that Q is convex. It follows from the properties of difference measurement that u is unique up to scale and location (Krantz et al. 1971). Because we can replace Ψ by any function that is a positive multiple of Ψ, we can replace Q by $Q'=aQ$, $a>0$. Because of symmetry of Q, $Q(0)=0$.[4]

3. Measurement Method

This section explains how regret theory can be measured. Even though the measurement method can be used to elicit the general regret model, Equation (1), as explained below, the intuition behind the method is clearer if we consider Equation (2), and the focus will be on this model in what follows.

The method consists of two parts. In the first part, the trade-off method (Wakker and Deneffe 1996) is used to elicit a *standard sequence* of outcomes $\{x_0,\ldots,x_k\}$ that are equally spaced in utility units, i.e., the elements of the standard sequence are such that $u(x_{j+1})-u(x_j)=u(x_1)-u(x_0)$ for all j in $\{1,\ldots,k-1\}$. Hence, the first part elicits the function u, which has the properties of Bell's (1982) incremental value function but without assuming the primitive of incremental value. In the second part the standard sequence is used to elicit the function Q.

It is easily verified, by substituting $\Psi(x_{j+1},x_j)$ for $Q(u(x_{j+1})-u(x_j))$ in the exposition below, that in the general regret theory of Loomes and Sugden (1987) the standard sequence $\{x_1,\ldots,x_k\}$ is such that $\Psi(x_{j+1},x_j)=\Psi(x_1,x_0)$ for all j in $\{1,\ldots,k-1\}$. The second part then uses the standard sequence to elicit Ψ for pairs of elements of the standard sequence. This shows that our method can indeed be used to elicit the general regret model, Equation (1).

3.1. Measurement of u

We start by selecting a probability p with $0<p<1$, two *gauge outcomes* G and g with $G\succ g$, and a *starting outcome* x_0. Then, we elicit the outcome x_1 for which the decision maker is indifferent between $x_{1_p}g$ and $x_{0_p}G$. This indifference yields by Equation (2):

$$pQ(u(x_1)-u(x_0))+(1-p)Q(u(g)-u(G))=0. \tag{3}$$

Rearranging and using the symmetry of Q gives

$$Q(u(x_1)-u(x_0))=\frac{1-p}{p}Q(u(G)-u(g)). \tag{4}$$

We then determine the outcome x_2 for which the decision maker is indifferent between $x_{2_p}g$ and $x_{1_p}G$. Writing out this indifference gives

$$Q(u(x_2)-u(x_1))=\frac{1-p}{p}Q(u(G)-u(g)), \tag{5}$$

and thus, $Q(u(x_1)-u(x_0))=Q(u(x_2)-u(x_1))$. Because Q is strictly increasing, it follows that $u(x_2)-u(x_1)=u(x_1)-u(x_0)$. We proceed by eliciting indifferences $x_{j+1_p}g\sim x_{j_p}G$, and in so doing we obtain a *standard sequence* $\{x_1,\ldots,x_k\}$ for which $u(x_{j+1})-u(x_j)=u(x_1)-u(x_0)$, $j=1,\ldots,k-1$. Because u is unique up to scale and location, the utility of two outcomes can be arbitrarily chosen. We set $u(x_0)=0$ and $u(x_k)=1$. It then follows that $u(x_j)=j/k$ for $j=0,\ldots,k$.

3.2. Measurement of Q

We use the standard sequence elicited in the first part to measure Q. The function Q was scaled such that $Q(1/k)=1$. Because $u(x_{j+1})-u(x_j)=1/k$ for $j=0,\ldots,k-1$, it follows that $Q(u(x_{j+1})-u(x_j))=1$ for any two successive elements of the standard sequence. We chose a probability p and three elements of the elicited standard sequence, namely, x_4, x_3, and

[3] In the appendix to their paper, Loomes and Sugden (1982) explain, however, how u can be derived solely from choices.

[4] Some authors used a slightly different expression for regret theory in which the utility of obtaining α and not β is equal to $u(\alpha)+R(u(\alpha)-u(\beta))$, where R is a regret-rejoice function. This formulation of regret theory is equivalent to Equation (2) when we define $Q(u(\alpha)-u(\beta))=u(\alpha)-u(\beta)+R(u(\alpha)-u(\beta))-R(u(\beta)-u(\alpha))$. Convexity of Q corresponds to R being decreasingly concave.

Table 1 Measurement Method

	Assessed quantities	Indifference	Implication	Stimuli: Measurement 1	Stimuli: Measurement 2
Part 1	$x_1, \ldots, x_5$	$x_{j+1_p} g \sim x_{j_p} G$, $j = 0, \ldots, 4$	$u(x_{j+1}) - u(x_j) = u(x_1) - u(x_0)$	$p = 1/3$ $G =$ €16 $g =$ €11 $x_0 =$ €20	$p = 1/2$ $G =$ €17 $g =$ €13 $x_0 =$ €20
Part 2	$z_1, \ldots, z_4$	$x_{4_{p_j}} x_0 \sim x_{3_{p_j}} z_j$, $j = 1, \ldots, 4$	$Q(u(z_j)) = p_j/(1-p_j)$	$p_1 = 1/4$ $p_2 = 2/5$ $p_3 = 3/5$ $p_4 = 3/4$	$p_1 = 1/4$ $p_2 = 2/5$ $p_3 = 3/5$ $p_4 = 3/4$

x_0, and then determined the outcome z for which the decision maker was indifferent between $x_{4_p} x_0$ and $x_{3_p} z$. By (2), this gives

$$pQ(u(x_4) - u(x_3)) + (1-p)Q(u(x_0) - u(z)) = 0$$
$$\Leftrightarrow \quad Q(u(z)) = p/(1-p). \tag{6}$$

By varying p, we can determine as many points of Q as deemed desirable. The utility of z will, in general, be unknown. It can be approximated from the known utility of elements of the standard sequence using interpolation.

An alternative way to measure Q that does not require approximating utilities is to select an element x_j of the standard sequence and to elicit the probability p for which the decision maker is indifferent between $x_{4_p} x_0$ and $x_{3_p} x_j$. By (2), this gives

$$pQ(u(x_4) - u(x_3)) = (1-p)Q(u(x_j) - u(x_0))$$
$$\Leftrightarrow \quad p = (1-p)Q(j/k)$$
$$\Leftrightarrow \quad Q(j/k) = \frac{p}{1-p}. \tag{7}$$

This *probability elicitation procedure* was used in extensive pilot sessions, but we finally decided not to use it. The procedure used in the study, the *outcome elicitation procedure*, turned out to have important advantages over the probability elicitation procedure. First, the outcome elicitation procedure uses the same response scale in both parts of the measurement, whereas the probability elicitation procedure uses different response scales: outcomes in the first part and probabilities in the second part. It is well known that different response scales prime different aspects of a decision problem, a finding known as *scale compatibility* (Tversky et al. 1988). Second, the outcome elicitation procedure turned out to be less susceptible to response error. The measurement of Q by the probability elicitation procedure was unstable for p close to zero or to one. Even small response errors then had strong effects on the elicited value of Q. Moreover, the effect of response error was asymmetric for p close to 0 or 1. For example, if a subject's true indifference probability was equal to 0.97, then there was much more room for error "to the left" of 0.97 than "to the right" of 0.97, leading to a downward bias in the elicited indifference probability and, hence, to a downward bias in Q. Finally, because Q is a convex transform of p in the probability elicitation procedure, response error can create artificial support for regret.[5]

A potential problem for both methods for measuring Q is error propagation. The trade-off method is a chained method, and errors made in earlier questions cumulate in later questions. Previous studies that explored the impact of error propagation on trade-off measurements and the values derived from these concluded that its impact was small (Bleichrodt and Pinto 2000, Abdellaoui et al. 2005). We performed a similar simulation exercise, which confirmed that error propagation was not a major cause of concern in this study either.

Table 1 summarizes the measurement method used. The second column shows the assessed quantities in the two parts, the third column shows the indifferences used to elicit these quantities, the fourth column shows the implications of the elicited indifferences under regret theory, and the final two columns show the stimuli that were used in the study, which is reported in the next section.

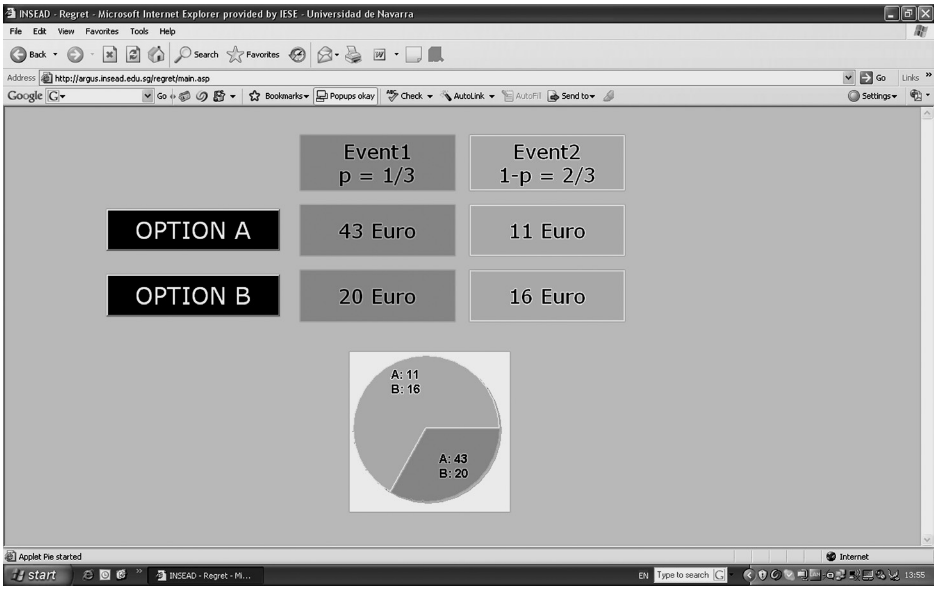
4. Study

4.1. Subjects and Stimuli

Subjects were 55 students (33 female, mean age of subjects was 20.9 years) from different faculties of the Pompeu Fabra University, Barcelona, Spain. Before the actual study, we tested and fine tuned the design in extensive pilot sessions.

[5] For example, if the subject's true indifference value of p is 0.70 and, hence, his true value of Q is 2.33, then an error of -0.05 will give $Q = 1.86$, but an error of $+0.05$ will give $Q = 3$. Consequently, symmetric response errors have asymmetric effects on Q, biasing the aggregate value of Q in the direction of more convexity.

Figure 1 Example of a Screen Faced in the First Part



Source. Microsoft product screen shot reprinted with permission from Microsoft Corporation.

4.2. Incentives

Subjects were paid a flat fee of €15 for their participation. In addition, six subjects were selected randomly to play one of their choices for real at the end of the session. Regret theory predicts that this *random incentive procedure* reveals subjects' true preferences in the sense that no differences will be observed with choices that are always played out for real. Empirical evidence supports this claim (Starmer and Sugden 1991, Beattie and Loomes 1997, Cubitt et al. 1998).

4.3. Procedure

The study was computer run in individual interview sessions lasting 55 minutes on average. Each session started with an explanation of the task, read aloud by the interviewer. Subjects could see this explanation on their computer screen. Then a practice question followed. After this practice question, the interviewer explained the random incentive procedure. The introduction ended with another practice question.

Subjects were not asked directly for their indifference values. Instead, indifference values were determined through a series of binary choices. Each binary choice corresponded to an iteration in a bisection process, described in the appendix. Figure 1 gives an example of the way the choice questions were presented. Subjects faced two options, neutrally labeled A and B, and were asked to choose between these options by clicking on their preferred option. Indifference was not allowed. After they had made their choice, subjects were asked to confirm their choice. If so, they moved on to the next question. If not, they faced the same question again. The confirmation question was intended to reduce the impact of response errors.

Both options always yielded different outcomes under the two events to avoid that event splitting would confound the results. The two events had different colors to make their distinction vivid. To remind subjects of the sizes of the probabilities of the events and to induce them to pay attention to the probabilities, we included a pie chart with the colored areas of the pie corresponding to the sizes of the probabilities involved. To control for order effects, we randomly counterbalanced what was option A and what was option B throughout.

Table 1 shows the stimuli used. Two measurements of regret theory were performed to test the robustness of the findings. If our method is valid and people behave according to regret theory, then the elicited utility functions and the elicited Q functions should be the same in the two measurements. The first measurement used probability 1/3 and gauge outcomes €16 and €11; the second measurement used probability 1/2 and gauge outcomes €17 and €13. The probabilities 1/3 and 1/2 were chosen because the pilot sessions showed that subjects found these easy to understand. The starting outcome x_0 was equal to €20 in both measurements. The two measurements were interspersed to prevent subjects from forming a match that would guide answers and to stimulate subjects to pay attention to the probability dimension. To distinguish the two measurements, we will use the superscripts 1 and 2 to denote the first and second measurements, respectively, in what follows.

To measure utility, we elicited two standard sequences $\{x_0^i, x_1^i, \ldots, x_5^i\}$, $i = 1, 2$. Hence, $k = 5$ in our study. As explained in §3, the measurement method amounted to finding values x_{j+1}^i so that

$x^i_{j+1_p}g^i$ and $x^i_{j_p}G^i$, $i = 1, 2$, were equivalent. To measure Q, we elicited two sequences of money amounts $z^i_1, \ldots, z^i_4$, $i = 1, 2$, that led to indifference between $x^i_{4_p}x_0$ and $x^i_{3_p}z^i_j$. Four values were used for p: $p = 1/4$, $p = 2/5$, $p = 3/5$, and $p = 3/4$. These four indifferences gave rise to four values of Q: $Q(u(z^i_1)) = 1/3$, $Q(u(z^i_2)) = 2/3$, $Q(u(z^i_3)) = 3/2$, and $Q(u(z^i_4)) = 3$. In contrast with the first part, the choice-based elicitations of the second part did not have to be performed consecutively and were randomly interspersed.

4.4. Validity of the Measurements

4.4.1. Internal Validity. We included two tests of the internal validity of the measurements. A problem in experiments is that people's preferences are often imprecise, and they make errors in reporting their responses. To test for response error, 12 choice questions were repeated. At the end of the first part, subjects were presented again with the third choice in the bisection process of four randomly chosen iterations, two for the first measurement of the regret model and two for the second measurement. The third choice was repeated because the value of the stimulus in the third choice was generally close but not equal to the elicited indifference value. Repeating a choice for which preference was obvious would inflate the support for consistency. At the end of the second part, we repeated the third choice in the bisection process of all eight iterations.

A second concern was strategic responding in the first part of the study. A potential problem of using real incentives in the trade-off method is its chained nature. Answers to utility elicitation questions are used as inputs in subsequent questions. This may induce subjects to respond strategically. By overstating values in the first stages of the measurement of u, subjects could increase the attractiveness of the options that they faced in later stages of the elicitation. This danger is particularly acute in matching tasks, where subjects directly state their indifference values, and these feature in subsequent elicitation questions. The danger of strategic responding was reduced in our study by using a choice-based elicitation procedure in which the indifference values used in subsequent questions never really appeared in the process of deriving them. This makes it hard to notice the chained nature of the questions. To test for strategic responding, we repeated the elicitations of x^1_1 and x^2_1 at the end of the first part of the study. Even when subjects understood the chained nature of the questions and responded strategically, they could not realize this during the elicitations of x^1_1 and x^2_1, because these elicitations did not use previous answers. If strategic responding were a problem, the repeated elicitations should produce higher values of x^1_1 and x^2_1 than the original elicitations.

4.4.2. External Validity. Two tests of external validity were also included. First, regret theory was measured twice using different stimuli. In the first part of the study, the stimuli in the two measurements were close, but because elements from the elicited standard sequences were used in the second part, the stimuli could differ considerably in the measurement of Q. For example, for subject 10 we used the prospects 110_p20 and 79_pz in the first measurement of Q and the prospects 60_p20 and 55_pz in the second measurement of Q. If the two measurements produced systematically different results, then this would undermine the external validity of the findings. It would imply that people did not behave according to regret theory because new choices could not be predicted by the elicited utility and Q functions.

We included another test of whether subjects behaved according to regret theory. It is possible that even when our findings were consistent with regret theory, people really behaved according to another theory, which happened to make similar predictions. The most obvious candidate is prospect theory, by now the dominant descriptive theory of decision under uncertainty. The trade-off method can also be used to elicit utility under prospect theory, and, as it turned out, our data were not only consistent with regret theory with a convex Q function but also with prospect theory with an inverse S-shaped probability weighting function. To distinguish between regret theory and prospect theory, we elicited the values of z^i_c, $i = 1, 2$, that made subjects indifferent between $x^i_{4_{1/2}}x_0$ and $x^i_{3_{1/2}}z^i_c$. Under regret theory, this indifference implies that

$$(1/2) * Q(u(x^i_4) - u(x^i_3)) = (1/2) * Q(u(z^i_c) - u(x_0)), \quad (8)$$

and, given the scaling of Q and the properties of u, we should observe that $z^i_c = x^i_1$. Under prospect theory, however, observing $z^i_c = x^i_1$ implies that

$$\begin{aligned} &w(1/2)u(x^i_4) + (1 - w(1/2))u(x_0) \\ &\quad = w(1/2)u(x^i_3) + (1 - w(1/2))u(x^i_1), \end{aligned} \quad (9)$$

where w denotes the probability weighting function. Given the scaling of u, (9) implies that $w(1/2) = 1/2$. It could of course be that the inverse S-shaped probability weighting functions happened to satisfy $w(1/2) = 1/2$, i.e., that their point of inflection was 1/2, but this is not in line with the data.[6] Consequently, these consistency questions help to distinguish between regret theory and prospect theory. Large differences between z^i_c and x^i_1 would indicate that subjects did not behave according to regret theory.

[6] Assuming prospect theory, the elicited answers determined $w(1/4)$, $w(2/5)$, $w(3/5)$, and $w(3/4)$. These values suggest that the point of inflection was lower than 1/2 in both measurements.

4.5. Analyses

We will present aggregate and individual data for u and Q. In the analysis of the aggregate data, we focus on the results based on the means and only occasionally report the median results in cases in which they led to additional insights. Significance of differences was tested both parametrically and nonparametrically. Because these tests always led to the same conclusions, only the parametric results are reported. Differences between proportions were tested by the binomial test.

4.5.1. Analysis of u. To investigate curvature of u at the individual level, three classifications were used. The first classification was based on the evolution of the slope of the utility function. For each subject, we computed the differences

$$\Delta_{gh,\,lm} = (x_g - x_h) - (x_l - x_m), \qquad g > h,\; g > l,\; g - h = l - m, \quad (10)$$

with all x_j, $j = g, h, l, m$, elements of the elicited standard sequence. Because $g - h = l - m$, it follows from the construction of the standard sequence that $u(x_g) - u(x_h) = u(x_l) - u(x_m)$. Concavity implies that the slope of the utility function decreases, and consequently, a positive value of $\Delta_{gh,\,lm}$ corresponds to a concave part of u. Likewise, a negative value of $\Delta_{gh,\,lm}$ corresponds to a convex part of u, and a value of zero to a linear part. Twenty values of $\Delta_{gh,\,lm}$ could be observed for each subject. To account for response error, a subject's utility function was classified as concave, convex, or linear if at least 50% of the values of $\Delta_{gh,\,lm}$ were positive, negative, or zero, respectively.[7] Otherwise, a subject was left unclassified.

The second classification was based on the area under the utility function. The domain of u was normalized to $[0, 1]$ through the transformation $(x_j^i - 20)/(x_5^i - 20)$, $j = 0, \ldots, 5$, $i = 1, 2$. If utility is linear, then the area under the normalized utility function is equal to $1/2$. If utility is concave (convex) then the area under the normalized utility function exceeds (is less than) $1/2$. Subjects were classified as concave (convex, linear) depending on whether the area under the normalized utility function exceeded (was less than, was equal to) $1/2$.

To smooth out irregularities, the data were also analyzed under specific parametric assumptions about utility. Two parametric families were examined: the power family and the exponential family. Both families are widely used in decision theory. The results were similar for the two families. To facilitate comparability with other studies on utility measurement, only the results for the power family are reported in what follows.

The *power family* is defined by x^r for $r > 0$, by $\ln(x)$ for $r = 0$, and by $-x^r$ for $r < 0$. We scaled the utility function such that $u(20) = 0$ and $u(x_5) = 1$. It is well known that $r < 1$ corresponds to concave utility, $r > 1$ to convex utility, and $r = 1$ to linear utility.

Estimations were by nonlinear least squares for both the mean and the median data and for each subject separately. The parametric estimates were used to obtain another, parametric classification of the individual utility functions. Using the standard errors (SEs) of the coefficients, a subject was classified as concave (convex) if his power coefficient was statistically significantly smaller (larger) than 1. The linear classification was not used because a coefficient that does not significantly differ from 1 does not imply that a linear function fits the data particularly well.

4.5.2. Analysis of Q. To compute Q, the utilities of the elicited values z_j^i, $i = 1, 2$, $j = 1, \ldots, 4$, had to be determined. This was done through linear interpolation. To check the robustness of the findings, we also used interpolation through the estimated power coefficients. All conclusions were the same under power interpolation, and we will focus on the results under linear interpolation in what follows.

To investigate the curvature of Q at the individual level, we computed the differences

$$\nabla_{gh,\,lm} = (Q(g/5) - Q(h/5)) - (Q(l/5) - Q(m/5)), \qquad g > h,\; g > l,\; g - h = l - m, \quad (11)$$

with $g, h, l, m \in \{0, \ldots, 5\}$. Because $g - h = l - m$, it follows that a positive value of $\nabla_{gh,\,lm}$ corresponds to a convex part of Q, a negative value to a concave part, and a value of zero to a linear part. Twenty values of $\nabla_{gh,\,lm}$ could be observed. A subject was classified as convex if at least 50% of the values of $\nabla_{gh,\,lm}$ were positive, and as concave if at least 50% of the values were negative.

To analyze the shape of Q at the individual level, the individual Q functions were normalized to domain $[0, 1]$ by dividing by $u(z_4^i)$, $i = 1, 2$, and to range $[0, 1]$ by dividing by $Q(u(z_4^i)) = 3$. If the area under the normalized Q function was less than (equal to, greater than) $1/2$ a subject was classified as convex (linear, concave). To check for robustness, we also fitted the individual Q functions parametrically through a power function and classified subjects based on their estimated power coefficients.

5. Results

One subject was excluded because she did not understand the tasks. This left 54 subjects in the final analyses.

[7] Similar criteria were used by Fennema and van Assen (1998), Abdellaoui (2000), and Etchart-Vincent (2004).

The internal validity of the measurements was good. In the first part of the study, 71.7% of the repeated choices were consistent in the first measurement of the regret model, and 78.3% in the second measurement. These rates are comparable to previously observed consistency rates (for a review, see Stott 2006). The consistency in the second part of the study was higher: 86.7% of the repeated choices were consistent in the first measurement, and 87.6% in the second measurement.[8]

No evidence of strategic responding was observed. Strategic responding predicts that the repeated elicitation of x_1^i, $i = 1, 2$, will exceed the original elicitation. In the first measurement of the regret model, the mean of the repeated elicitation of x_1^1 was indeed slightly higher than that of the original elicitation (37.3 versus 34.6) but the difference was not significant ($p = 0.41$). The difference was caused by one outlier, who did not display this behavior in the second measurement of the regret model, suggesting that his response was due to error. The median of the repeated elicitation was actually lower than that of the original elicitation (31.3 versus 32.3).

In the second measurement of the regret model, both the mean and the median of the repeated elicitation were slightly lower than those of the original elicitation. The mean was 28.3 in the repeated elicitation and 28.5 in the original elicitation, and the medians were 25 and 26, respectively. The differences were not significant ($p = 0.75$). The results on the external validity of the measurements are reported below.

5.1. Utility

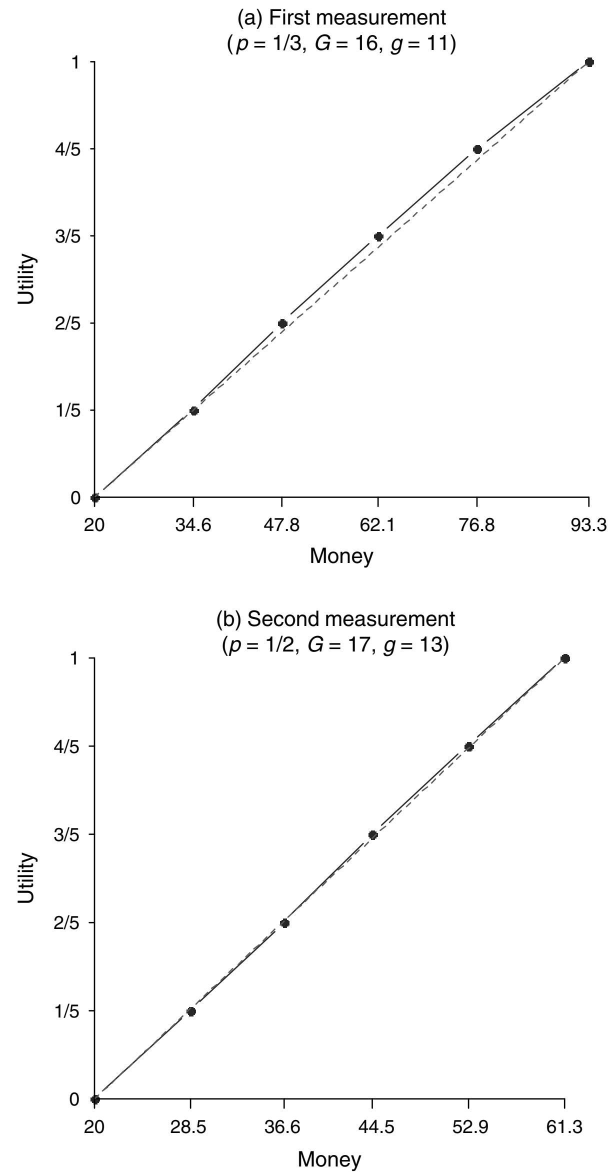
5.1.1. Aggregate Findings. Figure 2 shows the elicited utility functions based on the mean data. The dotted line is drawn for comparison and represents the case of linear utility. The figure shows that at the aggregate level utility was close to linear in both measurements. The differences between the step sizes $(x_{j+1} - x_j)$ and $(x_j - x_{j-1})$, for $j = 1, \ldots, 4$, give an indication of the shape of utility. Concavity corresponds to increasing step sizes, convexity to decreasing step sizes, and linearity to constant step sizes. The horizontal axes of Figure 2 show that the step sizes were close, in agreement with linearity of u at the aggregate level. The differences between the step sizes were not significant ($p = 0.13$ in the first measurement, and $p = 0.83$ in the second measurement).

Parametric fitting confirmed that the elicited utility functions were close to linear. The fitted power coefficients based on the mean data were 0.98 (SE = 0.015) in the first measurement and 1.01 (SE = 0.006) in the second measurement. Both power coefficients did not differ significantly from 1, the case corresponding to linear utility. Fitting based on the median data revealed comparable concavity in the first measurement (power coefficient = 0.98; SE = 0.026), but more concavity in the second measurement (power coefficient = 0.94; SE = 0.018). The latter coefficient was significantly different from 1 ($p = 0.02$).

The parametric fittings also indicated that the two utility functions were similar. The estimated power coefficients were close and not significantly different ($p = 0.08$ based on the mean data, and $p = 0.18$ based on the median data).

The absence of significant concavity of utility may be surprising to decision analysts used to concave utility and risk aversion. It should be kept in mind

Figure 2 Elicited Utility Functions Based on Mean Data



[8] Consistency did not differ much across the different questions; therefore, we pooled the data.

though that under nonexpected utility, risk aversion and concave utility are no longer equivalent. Under regret theory, part of people's attitude toward risk is captured by the regret function Q. It is well known that studies that estimate utility under expected utility find too much utility curvature (Fennema and van Assen 1998). Under nonexpected utility, less utility curvature is typically observed. The estimated power coefficients are close to those reported in other studies measuring utility by the trade-off method (Abdellaoui 2000, Abdellaoui et al. 2005, Schunk and Betsch 2006) and also to some studies measuring nonexpected utility using other methods (Tversky and Kahneman 1992, Abdellaoui et al. 2008, and one of the estimates in Abdellaoui et al. 2007a). However, the estimates of Gonzalez and Wu (1999), Abdellaoui et al. (2007b), and two other estimates in Abdellaoui et al. (2007a) were lower, indicating more concavity of utility.

5.1.2. Individual Data. The individual data produced more evidence for concavity of utility. In the nonparametric classifications of utility, this evidence was modest. Based on the evolution of the slope, 32 subjects were concave in the first measurement and 20 subjects were convex. In the second measurement, these numbers were 27 concave, 21 convex, and 2 linear. The remaining subjects could not be classified. Based on the area under the normalized utility function, 30 subjects had concave utility, and 24 subjects had convex utility in the first measurement. In the second measurement, these numbers were 27 concave, 23 convex, and 4 linear. The proportion of concave subjects was never significantly different from the proportion of convex subjects.

The parametric classification showed stronger support for concavity: In the first measurement, 30 subjects had significantly concave utility and 10 subjects had significantly convex utility; in the second measurement, 31 subjects had significantly concave utility and 12 subjects had significantly convex utility. Hence, what we observed was that the number of concave subjects was about the same as in the nonparametric classification, but the number of convex subjects was considerably lower. These results suggest that at least part of the convexity observed in the nonparametric classifications was due to noise. In both measurements, the proportion of concave subjects was significantly higher than the proportion of convex subjects ($p = 0.002$ in the first measurement, and $p = 0.004$ in the second measurement).

The conjecture that the observed convexity was primarily due to noise is supported if we compare the classifications of the subjects across the two measurements. Table 2 illustrates. The table shows that only one subject was significantly convex in both measurements of the regret model. By contrast, 18 subjects were significantly concave in both measurements.

Table 2 Parametric Classification of Subjects

	Second measurement			
	Concave	Convex	Unclassified	Total
First measurement				
Concave	18	6	6	30
Convex	6	1	3	10
Unclassified	7	5	2	14
Total	31	12	11	54

The individual data confirmed that elicited utility was similar across the two measurements, providing support for the external validity of the measurements of utility. Table 2 shows that few subjects were significantly concave in one measurement and significantly convex in the other. Moreover, the measures of utility curvature did not differ significantly between the two measurements. The mean area under the normalized utility curve was 0.506 in the first measurement and 0.501 in the second measurement ($p = 0.65$). The mean of the individual power coefficients was 1.00 in both measurements ($p = 0.91$). The medians were 0.95 in both measurements.[9] The correlation between the power function coefficients in the two elicitations was, however, low. The Pearson correlation was −0.13, and the Spearman rank correlation was −0.15. This low correlation reflects the error in measuring the power function coefficients. Neither differs significantly from 0 ($p > 0.10$ in both cases).

5.2. Regret

5.2.1. Selection of Subjects. One of the external tests examined to what extent subjects deviated from the predictions of regret theory by comparing z_c^i with x_1^i, $i = 1, 2$. If these differed strongly, then subjects violated regret theory. The question of what constitutes a strong difference is obviously somewhat arbitrary. Even if subjects behaved fully in accordance with regret theory, z_c^i and x_1^i could differ due to error and rounding. People make errors when responding to choice questions. A difference between z_c^i and x_1^i could also be due to the elicitation procedure used, which only faced subjects with integer numbers. The final indifference value was rounded to the next integer. Based on previous experience with the trade-off method, we took a difference of

[9] One pilot session involved 56 subjects in which utility was elicited by a similar method as used in the experiments reported in this paper except that we used larger stimuli: $G = €100$, $g = €40$, and $x_0 = €200$. The value of p was set equal to 1/3. In this pilot, we also found linear utility at the aggregate level. The mean of the individual power estimates was 1.00 (median 0.94) and did not differ significantly from the power estimates that we found in the two experiments reported in this paper ($p = 0.99$ and $p = 0.92$ compared with the first and the second measurements, respectively).

half the average step size in the first two elicitations of the standard sequence as being attributable to error and rounding. By this criterion, seven subjects were excluded because their responses indicated that they did not behave according to regret theory, even allowing for error. The conclusions presented below were not sensitive to using smaller or wider error bounds, although, obviously, the number of excluded subjects varied with the size of the permitted error bounds.

There were four more subjects for whom z_c^1 and x_1^1 differed by more than 0.75 times the average step size in the first two elicitations of the standard sequence, but z_c^2 and x_1^2 were (approximately) equal. For these subjects, we interpreted the difference between z_c^1 and x_1^1 as reflecting error, mainly caused by overstating x_1^1.[10] They were deleted from the analysis of Q in the first measurement of the regret model, but not in the second, where we had no reason to believe that error had driven their responses.

Several responses violated *monotonicity*, requiring that prospects become more attractive if the probability of the best outcome increases. Monotonicity implied that if $p > q$ and $x^i_{4_p}x_0 \sim x^i_{3_p}z$ and $x^i_{4_q}x_0 \sim x^i_{3_q}z'$, then $z > z'$. Regret theory predicts that $z > z'$ through the increasingness of Q.[11] Allowing for the same error size as above, 5.3% (4.6%) of the responses in the first (second) measurement of the regret model violated monotonicity. These rates of monotonicity violations are relatively low compared with other studies in the literature. One subject violated monotonicity several times and in both measurements. He was excluded from the analyses. The other subjects violated monotonicity only once. For these subjects, the violating response was deleted, but the other responses were retained.

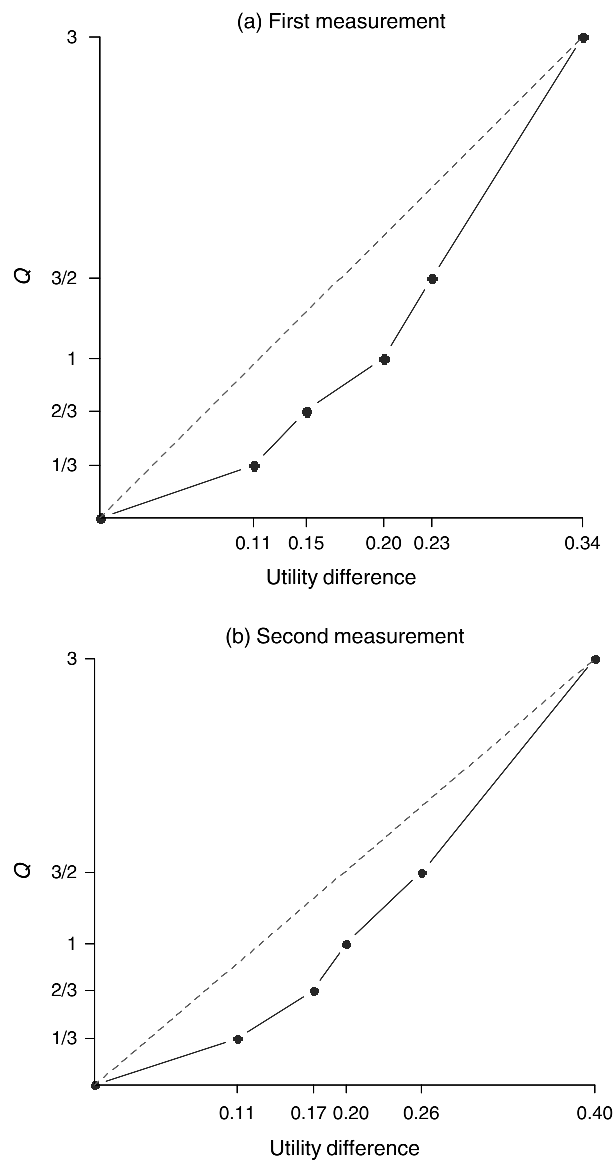
For the subjects remaining in the analyses (42 in the first measurement of the regret model and 46 in the second measurement), z_c^i and x_1^i, $i = 1, 2$, did not differ significantly. Mean values were 30.7 and 32.2 in the first measurement of the regret model ($p = 0.15$) and 25.9 and 26.7 in the second measurement, respectively ($p = 0.23$).[12] Under prospect theory, these numbers imply that $w(1/2) = 0.47$ in both measurements. This degree of probability weighting is less than what is commonly observed in the literature. Both probability weights did not differ significantly from 1/2 ($p > 0.10$ in both cases).

[10] This interpretation was confirmed by their other responses in the elicitation of Q, which were all (much) smaller than x_1^1.

[11] This also holds true in the more general regret model of Loomes and Sugden (1987), which assumes increasingness of Ψ in its first argument.

[12] With all subjects included, the evidence was more ambiguous. Values of z_c^1 and x_1^1 differed significantly by the parametric t-test ($p = 0.04$), but not by the nonparametric Wilcoxon test ($p = 0.10$). Values of z_c^2 and x_1^2 did not differ significantly ($p > 0.10$ in both tests).

Figure 3 Elicited Q Functions Based on Mean Data



5.2.2. Aggregate Results. Figure 3 shows the Q functions obtained in the two measurements. The dotted lines represent linearity of Q, the special case of expected utility. As predicted by regret theory, Q was convex in both measurements. The degree of convexity was slightly more pronounced in the first measurement than in the second measurement. The estimated power coefficients based on the mean data were 1.89 (SE = 0.047) in the first measurement and 1.73 (SE = 0.037) in the second measurement. The power coefficients differed significantly at the 5% level ($p = 0.02$).

Under regret theory, the elicited utility differences in the two measurements should be the same. This follows from the increasingness of Q. We could not reject the hypothesis that the utility differences were equal and, consequently, the hypothesis that subjects behaved in line with regret theory ($p > 0.10$ in all tests). The results were not affected by the use of linear interpolation. The utility differences were also not significantly different when interpolation by the estimated power coefficients was used instead ($p > 0.10$).

5.2.3. Individual Analysis. At the individual level, convexity was clearly the dominant pattern. Based on the evolution of the slope of Q, 30 (36) subjects were convex and 10 (15) were concave in the first (second) measurement. Based on the area under the standardized regret curve, 37 (41) subjects were convex and 4 (5) concave in the first (second) measurement. The proportion of convex subjects was always significantly higher than the proportion of concave subjects ($p < 0.01$ in all tests).

The classifications were consistent across the two measurements. According to the evolution of the slope, 22 subjects were convex in both measurements. Only two subjects were concave in both measurements. According to the area under the normalized curve, 34 subjects were convex in both measurements, and 2 subjects were concave in both measurements.

The parametric classification confirmed the predominance of regret aversion and the consistency across the two measurements. In the first (second) measurement 35 (35) subjects had a power coefficient exceeding 1, corresponding to convex Q, and 6 (11) subjects had a power coefficient smaller than 1, corresponding to concave Q. Twenty-seven subjects were convex in both measurements, and only two subjects were concave in both measurements.

The measurements of Q were more noisy than those of u as reflected in larger standard errors of the estimated power coefficients. Because of these and the relatively low number of observations, many power coefficients failed to reach statistical significance. In the first (second) measurement, 10 (19) subjects were significantly convex, and only 1 (1) subject was significantly concave. Of the 10 subjects who were significantly convex in the first measurement, 9 were also significantly convex in the second measurement. No subject was significantly concave in both measurements. The proportion of convex subjects was significantly larger than the proportion of concave subjects in all comparisons ($p < 0.001$).

The finding that fewer subjects were significantly convex in the first measurement of the regret model than in the second measurement reflects that the data from the first measurement were more noisy. It is not clear why this was, because the two measurements were randomly interspersed. The finding is the more puzzling because, by the nature of the measurement method used, errors have more impact the smaller the step size of the elicited standard sequence. The step size was, however, considerably smaller in the second measurement than in the first, as shown in Figure 2.

The measures of the curvature of Q were similar across the two measurements, providing support for the external validity of the measurements. It was already mentioned that most subjects were classified the same in the two measurements. Moreover, the mean areas under the normalized Q curves were close: 0.301 in the first measurement and 0.299 in the second measurement, with the difference being insignificant ($p = 0.94$). The means of the individual power coefficients were also close, and their difference was not significant: 1.73 in the first measurement and 1.65 in the second measurement ($p = 0.71$). The medians of the individual power coefficients were even closer: 1.53 in the first measurement and 1.51 in the second measurement. The correlations between the power function coefficients were low, reflecting the error in measuring these: the Pearson correlation coefficient was 0.02, and the Spearman rank correlation coefficient was 0.15. Neither differs significantly from 0 ($p > 0.10$ in both cases).

5.3. Decreasing or Increasing Convexity?

It is of interest to explore how the degree of convexity of Q evolves with the size of the utility differences. Do deviations from linearity, i.e., deviations from expected utility, occur primarily for low utility differences or for larger utility differences? Under regret theory, the more convex a subject's Q function, the more likely the subject is to display intransitive choice behavior. Empirical studies often have difficulty uncovering intransitive choice patterns (e.g., Birnbaum and Schmidt 2008). One explanation could be that the differences in payoffs are too small to give rise to regret effects. That is, it might be that deviations from expected utility primarily occur when the difference in attractiveness, i.e., the difference in utility, is large enough. This section explores this possibility.

Two measures of the change in convexity were used. The first compared the increase in slope between $Q(0)$ and $Q(u(z_2))$ with the increase in slope between $Q(u(z_2))$ and $Q(u(z_4))$. The difference between the slope between $Q(u(z_2))$ and $Q(u(z_1))$ and the slope between $Q(u(z_1))$ and $Q(0)$ was compared with the difference between the slope between $Q(u(z_4))$ and $Q(u(z_3))$ and the slope between $Q(u(z_3))$ and $Q(u(z_2))$.[13] If the former was larger than the latter, the subject was *decreasingly convex*; if it was smaller, the subject was *increasingly convex*.

The second measure was based on the (normalized) area between the diagonal and Q. The area between $Q(x_0)$ and $Q(z_2)$ was compared with the area between

[13] That is, we compared $(Q(u(z_2)) - Qu(z_1))/(u(z_2) - u(z_1)) - Q(u(z_1))/(u(z_1))$ and $(Q(u(z_4)) - Q(u(z_3)))/(u(z_4) - u(z_3)) - (Q(u(z_3)) - Q(u(z_2)))/(u(z_3) - u(z_2))$.

$Q(z_2)$ and $Q(z_4)$. Decreasing convexity means that the former exceeds the latter. Increasing convexity corresponds to the latter exceeding the former.

The results based on both measures were broadly similar and indicated no unequivocal support for either decreasing or increasing convexity of Q. Of the subjects who had a convex Q in the first (second) measurement of the regret model, 20 (17) were decreasingly convex, and 12 (20) were increasingly convex, based on the evolution of slope. Based on the area between the diagonal and Q, 17 (18) subjects were decreasingly convex, and 16 (19) subjects were increasingly convex in the first (second) measurement. Decreasingly convex subjects tended to deviate more from linearity, but the difference was generally insignificant.

6. Discussion

This paper has introduced a new and tractable method to measure regret theory. Using regret theory in practical decision analysis requires such a method, but it did not exist as yet. We have shown that our method is feasible and valid in the sense that two measurements produced similar results. The method made it possible for the first time to observe regret theory and thereby to answer the open empirical questions of whether regret aversion really exists and how pronounced it is. This paper has also provided new evidence on the validity of the trade-off method. We showed that the trade-off method is robust to violations of transitivity and is not vulnerable to strategic responding.

The data revealed substantial regret aversion. The function Q, which captures regret, was convex both at the aggregate level and for most subjects even when event splitting was controlled for. Moreover, the observed convexity was robust. It appeared in both measurements of the regret model that we performed using different stimuli. What is more, the degree of regret aversion was similar in the two measurements.

One might retort that even though the two measurements were close, this may have been caused by something other than regret. We observed no evidence for this conjecture. Two tests were included to explore whether our subjects behaved according to regret theory. The tests particularly tried to separate regret aversion and probability weighting, an important cause of deviations from expected utility modeled by prospect theory. Both tests indicated that a substantial majority of the subjects behaved in line with the predictions of regret theory.

The data provided support for the validity and reliability of the trade-off method. First, different elicitations using different stimuli yielded similar results. Second, the two tests of strategic responding that we included showed no evidence of strategic responding. Previous studies already showed that the impact on the trade-off method of error propagation, another potential drawback of using chained responses, was small (Bleichrodt and Pinto 2000, Abdellaoui et al. 2005). Taken together, the available evidence suggests that the trade-off method is a reliable method for measuring utilities.

At the aggregate level, utility was close to linear. This finding is consistent with earlier studies using the trade-off method. At the individual level, however, more evidence for concave utility was found. Subjects who were concave in one measurement generally were also concave in the other measurement. By contrast, only one subject was significantly convex in both measurements, suggesting that part of the observed convexity was due to noise.

A possible heuristic that subjects could have applied in the trade-off method would be to select a fixed difference between x_j and x_{j+1} and to choose one option if the difference was less than this value and the other if it exceeded this value. Applying this heuristic consistently throughout the elicitation would result in linearity of u. There was no indication that subjects applied this heuristic: none of the subjects displayed constant differences between successive elements of the standard sequence. The heuristic did not show up in the classification of the individual subjects either, because hardly any subjects were classified as having linear utility. It could of course be that subjects did not base their answers solely on this heuristic but that it nevertheless biased their responses in the direction of linearity of u. Abdellaoui et al. (2007a) compared the utilities elicited by the trade-off method with those of other elicitation methods that are less vulnerable to this heuristic and observed no significant differences.

A potential drawback of the study performed in this paper is the use of forced choices: subjects always had to choose one of the options, and indifference was not allowed. This was done to stimulate subjects to think hard about the choices and to avoid "lazy responses": if indifference were allowed, subjects might just state it because it is an easy option that reduces cognitive effort. The problem with the use of forced choices is that if subjects are truly indifferent, they still have to choose one of the options. To encourage breaking ties in a random manner, we varied what was option A and what was option B. Hence, breaking ties by consistently choosing A or B did not affect the findings. If subjects nevertheless used a specific rule to break ties, then the use of forced choices may have biased the results. It turned out that selecting a specific rule for breaking ties had no effect on u. For the measurement of Q, however, it might have had an effect, but its overall direction is ambiguous.

One plausible rule for breaking ties is choosing the less risky option given that most subjects were risk averse. This will bias Q in the direction of convexity on $[0, 1/5]$. On $[1/5, \rightarrow)$, on the other hand, the bias will be in the direction of concavity.

Regret theory allows for intransitive behavior. The more Q deviates from linearity, the more a decision maker is prone to display intransitive choices. Previous studies had to use ingenious designs to find intransitive choice patterns (Loomes et al. 1991) and often were unable to uncover these (Birnbaum and Schmidt 2008). An advantage of the method of this paper is that it is applied at the individual level, and thereby allows for a straightforward identification of individuals who are particularly prone to intransitive behavior. Once intransitive individuals are identified, regret theory immediately predicts choices that are likely to lead to intransitive preferences. For instance, suppose a standard sequence x_0, x_1, x_2, x_3 is elicited, and Q is found to be convex. Let $(1/3{:}x_3,\ 1/3{:}x_2,\ 1/3{:}x_0)$ denote the prospect that gives x_3 with probability 1/3, x_2 with probability 1/3, and x_0 with probability 1/3. Then, regret theory predicts the following intransitive cycle: $(1/3{:}x_2,\ 1/3{:}x_1,\ 1/3{:}x_2) \succ (1/3{:}x_3,\ 1/3{:}x_2,\ 1/3{:}x_0) \succ (1/3{:}x_1,\ 1/3{:}x_3,\ 1/3{:}x_1) \succ (1/3 : x_2,\ 1/3 : x_1,\ 1/3{:}x_2)$. Testing whether these cycles indeed obtain and are more likely for subjects with a more convex Q is a topic worthy of future research.

7. Conclusion

Nearly 30 years after its introduction, regret theory remains a popular model of decision under uncertainty. There seems to be a recent increase in the interest in regret theory with various theoretical papers using regret in decision theory. This paper is the first to make regret theory quantitatively observable. It shows that a quantitative measurement of regret theory is feasible and can be performed at the individual level and by familiar measurement tools. The measurements have good internal and external validity. Showing that such a quantification and smooth measurement are possible in spite of the absence of transitivity is the main contribution of this paper.

Substantial regret aversion was found both at the aggregate and at the individual level even when event splitting was controlled for. Regret aversion was similar in the two measurements that we performed. The trade-off method was robust to violations of transitivity and a valid method for measuring utility. It was not affected by strategic responding.

The provision of a tractable and valid method that takes account of preference heterogeneity will hopefully foster future applications of regret theory in decision analysis and will facilitate research into intransitive preferences.

Acknowledgments

The authors are grateful to Aurélien Baillon, Graham Loomes, Kirsten Rohde, Stefan Trautmann, Peter P. Wakker, George Wu, an associate editor, and four referees for their comments on previous drafts of this paper; to Dinesh Kunchamwar and Alpika Mishra for writing the computer program used in the experiment; and to Arthur Attema for his help in running the experiment. Han Bleichrodt's research was supported by grants from The Netherlands Organization for Scientific Research (NWO). Alessandra Cillo's research was supported by the IESE Research Division and by the Spanish Ministry of Science and Innovation, Project ECO2008-05155/ECON, and Enrico Diecidue's research was supported by grants from the INSEAD research and development department and the INSEAD Alumni Fund.

Appendix. Bisection Method for Eliciting the Indifference Values

In the measurement of u, x_{j+1} was elicited through choices between $A = x_{j_p} G$ and $B = x_{j+1_p} g$, $j = 0, \ldots, 4$.[14] The initial value of x_{j+1} was a random integer in the interval $[x_j, x_j + 5 * (G - g)]$. There were two possible scenarios:

(i) If A was chosen we increased x_{j+1} by $D = 5 * (G - g)$ until B was chosen. We then decreased x_{j+1} by $D/2$. If A (B) was subsequently chosen we increased (decreased) x_{j+1} by $D/4$, 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 $D'/4$. If A was subsequently chosen then we increased (decreased) x_{j+1} by $D'/8$, etc.

We stopped the elicitation 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 A.1 gives an example of the procedure for the elicitation of x_1 through comparisons between $A = €20_{1/2}€17$ and $B = €x_{1_{1/2}}€13$. In this example, the initial random value for x_1 was 26. We recorded as indifference value the midpoint between 34 and 36, that is, 35.

The procedure in the second part was largely similar. We elicited the value of z for which indifference held between $A = €x_{4_p}€20$ and $B = €x_{3_p}€z$,[15] where p was one of $1/4, 2/5, 3/5, 3/4$ and x_4 and x_3 were the outcomes of the standard sequence elicited in the first part.

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

[14] Note that in the experiment we varied what was option A and what was option B. For clarity of exposition, we keep them the same here.

[15] Again, what was A and what was B varied across the steps in the iteration process, see Footnote 5.

Table A.1 $p = 1/2$: **Example of the Elicitation of** x_1

Iteration	x_1	Choice
1	26	A
2	46	B
3	36	B
4	31	A
5	34	A

Table A.2 **Example of the Elicitation of** z **When** $x_4 = 70$ **and** $x_3 = 50$

Iteration	z	Choice
1	28	B
2	24	A
3	26	A

A and B equal in expected value. There were two possible scenarios:

(i) As long as A was chosen we increased z by $D = (x_4 - z_{EV})/2$ if $p \leq 1/2$; $D = (x_5 - z_{EV})/2$ if $p > 1/2$. We used a different adjustment for $p \leq 1/2$ 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 $p \leq 1/2$) or between z and x_5 (for $p > 1/2$) 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.

(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. We stopped the elicitation 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 A.2 gives an example of the procedure for the elicitation of z through comparisons between $A = €70_{1/4}€20$ and $B = €50_{1/4}€28$, where 28 was selected as the initial stimulus value from the interval $[26.6 - 3, 26.6 + 3]$. We recorded as the indifference value the midpoint between 26 and 28, that is, 27.

References

Abdellaoui, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management Sci.* **46**(11) 1497–1512.
Abdellaoui, M., C. Barrios, P. P. Wakker. 2007a. Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory. *J. Econometrics* **138**(1) 356–378.
Abdellaoui, M., H. Bleichrodt, O. l'Haridon. 2008. A tractable method to measure utility and loss aversion under prospect theory. *J. Risk Uncertainty* **36**(3) 245–266.
Abdellaoui, M., H. Bleichrodt, C. Paraschiv. 2007b. Measuring loss aversion under prospect theory: A parameter-free approach. *Management Sci.* **53**(10) 1659–1674.
Abdellaoui, M., F. Vossmann, M. Weber. 2005. Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Sci.* **51**(9) 1384–1399.
Barberis, N., W. Xiong. 2009. What drives the disposition effect? An analysis of a long-standing preference-based explanation. *J. Finance* **64**(2) 751–784.
Barberis, N., M. Huang, R. H. Thaler. 2006. Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *Amer. Econom. Rev.* **96**(4) 1069–1090.
Beattie, J., G. Loomes. 1997. The impact of incentives upon risky choice experiments. *J. Risk Uncertainty* **14**(2) 155–168.
Bell, D. E. 1982. Regret in decision making under uncertainty. *Oper. Res.* **30**(5) 961–981.
Bell, D. E. 1983. Risk premiums for decision regret. *Management Sci.* **29**(10) 1156–1166.
Birnbaum, M., U. Schmidt. 2008. An experimental investigation of violations of transitivity in choice under uncertainty. *J. Risk Uncertainty* **37**(1) 77–91.
Bleichrodt, H., J. L. Pinto. 2000. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Sci.* **46**(11) 1485–1496.
Braun, M., A. Muermann. 2004. The impact of regret on the demand for insurance. *J. Risk Insurance* **71**(4) 737–767.
Cubitt, R., C. Starmer, R. Sugden. 1998. On the validity of the random lottery incentive system. *Experiment. Econom.* **1**(2) 115–131.
Di Cagno, D., J. D. Hey. 1988. A direct test of the original version of regret theory. *J. Behav. Decision Making* **1**(1) 43–56.
Engelbrecht-Wiggans, R., E. Katok. 2008. Regret and feedback information in first-price sealed-bid auctions. *Management Sci.* **54**(4) 808–819.
Etchart-Vincent, N. 2004. Is probability weighting sensitive to the magnitude of consequences? An experimental investigation on losses. *J. Risk Uncertainty* **28**(3) 217–235.
Feliz-Ozbay, E., E. Y. Ozbay. 2007. Auctions with anticipated regret: Theory and experiment. *Amer. Econom. Rev.* **97**(4) 1407–1418.
Fennema, H., M. van Assen. 1998. Measuring the utility of losses by means of the trade-off method. *J. Risk Uncertainty* **17**(3) 277–295.
Fishburn, P. C. 1982. Non-transitive measurable utility. *J. Math. Psych.* **26**(1) 31–67.
Gollier, C., B. Salanié. 2006. Individual decisions under risk, risk sharing, and asset prices with regret. Working paper, University of Toulouse, Toulouse, France.
Gonzalez, R., G. Wu. 1999. On the form of the probability weighting function. *Cognitive Psych.* **38**(1) 129–166.
Harrison, G. W., E. Rutström. 2008. Risk aversion in the laboratory. J. C. Cox, G. W. Harrison, eds. *Risk Aversion in Experiments. Research in Experimental Economics*, Vol. 12. Emerald, Bingley, UK, 41–196.
Hens, T., M. Vlcek. 2006. Does prospect theory explain the disposition effect? Working paper, University of Zurich, Zurich.
Humphrey, S. J. 1995. Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *J. Risk Uncertainty* **11**(3) 263–274.
Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–291.
Krantz, D. H., R. D. Luce, P. Suppes, A. Tversky. 1971. *Foundations of Measurement*, Vol. 1. Academic Press, New York.
Larrick, R. P. 1993. Motivational factors in decision theories: The role of self-protection. *Psych. Bull.* **113**(3) 440–450.
Loomes, G., R. Sugden. 1982. Regret theory: An alternative theory of rational choice. *Econom. J.* **92**(368) 805–824.
Loomes, G., R. Sugden. 1987. Some implications of a more general form of regret theory. *J. Econom. Theory* **41**(2) 270–287.
Loomes, G., C. Starmer, R. Sugden. 1991. Observing violations of transitivity by experimental methods. *Econometrica* **59**(2) 425–439.
Muermann, A., J. M. Volkman. 2007. Regret, pride, and the disposition effect. Working paper, University of Pennsylvania, Philadelphia.
Muermann, A., O. S. Mitchell, J. M. Volkman. 2006. Regret, portfolio choice, and guarantees in defined contribution schemes. *Insurance, Math., Econom.* **39**(2) 219–229.

Perakis, G., G. Roels. 2008. Regret in the newsvendor model with partial information. *Oper. Res.* **56**(1) 188–203.

Schunk, D., C. Betsch. 2006. Explaining heterogeneity in utility functions by individual differences in decision modes. *J. Econom. Psych.* **27**(3) 386–401.

Starmer, C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *J. Econom. Literature* **38**(2) 332–382.

Starmer, C., R. Sugden. 1991. Does the random-lottery incentive system elicit true preferences? An experimental investigation. *Amer. Econom. Rev.* **81**(4) 971–978.

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

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

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

Tversky, A., S. Sattath, P. Slovic. 1988. Contingent weighting in judgment and choice. *Psych. Rev.* **95**(3) 371–384.

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

Zeelenberg, M. 1999. Anticipated regret, expected feedback and behavioral decision making. *J. Behav. Decision Making* **12**(2) 93–106.

Zeelenberg, M., J. Beattie, J. van der Pligt, N. K. de Vries. 1996. Consequences of regret aversion: Effects of expected feedback on risky decision making. *Organ. Behav. Human Decision Processes* **65**(2) 148–158.