Theory Dec. (2010) 69:599–614
DOI 10.1007/s11238-009-9147-1

Testing transitivity in choice under risk

Michael H. Birnbaum · Ulrich Schmidt

Published online: 6 May 2009

Abstract Recently proposed models of risky choice imply systematic violations of transitivity of preference. This study explored whether people show the predicted intransitivity of the two models proposed to account for the certainty effect in Allais paradoxes. In order to distinguish "true" violations from those produced by "error," a model was fit in which each choice can have a different error rate and each person can have a different pattern of preferences that need not be transitive. Error rate for a choice is estimated from preference reversals between repeated presentations of the same choice. Results showed that few people repeated intransitive patterns. We can retain the hypothesis that all participants were transitive.

Keywords Transitivity · Errors · Gambling effect · Reference points

1 Introduction

The most popular theories of risky decision making assume that the decider computes a value (or "utility") for each alternative and chooses (or at least, tends to choose) the alternative with the highest value. This class of models includes expected utility theory (EU), cumulative prospect theory (CPT), prospective reference theory (PRT),

M. H. Birnbaum
Department of Psychology, California State University, Fullerton, USA

U. Schmidt (✉)
Department of Economics, University of Kiel, Olshausenstr. 40, 24098 Kiel, Germany
e-mail: us@bwl.uni-kiel.de; u1366@gmx.de

U. Schmidt
Kiel Institute for the World Economy, Kiel, Germany

transfer of attention exchange (TAX), gains decomposition utility (GDU), and many others (Luce 2000; Marley and Luce 2005; Starmer 2000; Tversky and Kahneman 1992; Wu et al. 2004; Viscusi 1989). Although these models can be compared by means of special experiments testing properties that distinguish them (Birnbaum 1999, 2004, 2005a, c), they all share, in common, the property of transitivity.

Transitivity is the property that if a person prefers alternative A to B, and B to C, then that person should prefer A to C. If a person systematically violates this property, it should be possible to turn that person into a "money pump" if the person were willing to pay a little to get A rather than B, something to get B rather than C, something to get C rather than A and so on, ad infinitum. Most theoreticians, but not all (Fishburn 1991, 1992; Bordley and Hazen 1991), conclude that it would not be rational to violate transitivity.

Despite such seemingly "irrational" implications of violating transitivity, some descriptive theories imply that people can in certain circumstances be induced to violate it. Models that violate transitivity include the lexicographic semi-order (Tversky 1969; see also Leland 1994), the additive difference model [including regret theory of Loomes and Sugden (1982) and Fishburn (1982) Skew-symmetric bilinear utility], Bordley (1992) expectations-based Bayesian variant of Viscusi's PRT model, the priority heuristic model (Brandstaetter et al. 2006), context-dependent model of the gambling effect (CDG, Bleichrodt and Schmidt 2002) and context- and reference-dependent utility (CRU, Bleichrodt and Schmidt 2007).

If one could show that people systematically violate transitivity, then it means that the first class of models must be either rejected or modified to allow such effects. Models that can violate transitivity provide a basis for designing experiments to test transitivity. This study will explore violations predicted by models of Bleichrodt and Schmidt (2002, 2007).

A number of previous studies attempted to test transitivity (Birnbaum et al. 1999; Loomes et al. 1989, 1991; Loomes and Taylor 1992; Humphrey 2001; Starmer 1999; Starmer and Sugden 1998; Tversky 1969). However, these studies remain controversial; there is not yet consensus that there are situations that produce systematic violations of transitivity (Luce 2000; Iverson and Falmagne 1985; Iverson et al. 2006; Regenwetter and Stober 2006; Sopher and Gigliotti 1993; Stevenson et al. 1991). A problem that has frustrated previous research has been the issue of deciding whether an observed pattern represents "true violations" of transitivity or might be due instead to "random errors."

The purpose of this article is to empirically test patterns of intransitivity that are predicted by CDG and CRU, using an "error" model that has the promise to be neutral with respect to the issue of transitivity and which seems plausible as a descriptive model of the variability of repeated choices. The lotteries we employ are similar to those used in previous studies; however we present them only in terms of probabilities with no reference to states of the worlds. For our lotteries, CDG and CRU models make the same pattern of predicted violation given parameters chosen to describe well-known phenomena in risky decision making. Interestingly, the direction of violations is opposite to that implied by regret theory.

The remainder of this article is organized as follows. The next section describes CDG and CRU models and shows their predicted pattern of violation of transitivity;

the third section describes the error model; the fourth section describes the experiments; the fifth section presents the results, which show that transitive models can be retained for our data, and the sixth section discusses the implications.

2 Theoretical predictions

Our experimental design involves variations of the three lotteries presented in Table 1 where p and q are probabilities, and $a > b > c$ are monetary consequences. Related to the literature on preference reversals, A can be regarded as "\$-bet," B as "p-bet," and C is cash.

It will be shown below that CDG and CRU models both imply the intransitive pattern, $A \succ B$, $B \succ C$, and $C \succ A$. Bordley's (1992, p. 135) intransitive variant of PRT implies the opposite intransitive pattern, $B \succ A$, $A \succ C$, and $C \succ B$. Regret theory also implies this opposite pattern if probabilities in Table 1 denoted events (states of the world), as shown by Loomes et al. (1991).

Before presenting the models, let us introduce some notation. Lotteries are denoted by capital letters A, B, C; X is the set of all the pure consequences with elements of X being denoted by a, b, c. Formally, an element of X is a lottery that yields one consequence with a probability of one. Probabilities are denoted by p and q, so $p(a)$ is the probability of consequence a in lottery A.

2.1 The context-dependent model of the gambling effect (CDG)

The CDG presupposes that the decision maker employs a different cognitive process when choosing between two risky lotteries from that used when choosing between a risky lottery and a sure consequence. In the latter case, it is assumed that people are more risk averse because the sure outcome makes the risk involved in the risky lottery more salient. The CDG postulates two distinct utility functions, u and v, such that

$$A \succ B \Leftrightarrow \begin{cases} \sum_{i=1}^{n} p(a_i)u(a_i) > \sum_{i=1}^{n} p(b_i)u(b_i) & A \wedge B \notin X \\ \sum_{i=1}^{n} p(a_i)v(a_i) > \sum_{i=1}^{n} p(b_i)v(b_i) & A \vee B \in X \end{cases} \tag{1}$$

The hypothesis that subjects are more risk averse when choosing between a risky and a riskless lottery implies that v is a concave transformation of u. In contrast with other models of the gambling effect by Fishburn (1980), Schmidt (1998) and Diecidue et al. (2004), CDG does not imply violations of first-order stochastic dominance but it does allow violations of transitivity.

Table 1 Design of lotteries used to test transitivity

Lottery	p	q	$1-p-q$
A	a	0	0
B	b	b	0
C	c	c	c

Note: $a > b > c > 0$

The lotteries in Table 1 can violate transitivity under CDG. When choosing between A and B the utility function u is employed while the choices between A and C and B and C are determined by the utility function v. Suppose $B \succ A$ according to the utility function u. Assuming v is more concave than u, we cannot have $A \succ C$ and $C \succ B$ under v. This means the cycle implied by regret theory is ruled out under CDG. Now suppose instead that $A \succ B$ according to utility function u. Again, v is more concave than u so we can have $B \succ C$ and $C \succ A$ which shows that the opposite cycle is admissible under CDG.

2.2 Context- and reference-dependent utility (CRU)

The CRU is based upon Sugden (2003) model of subjective expected utility with state-dependent reference point. Reference points had been discussed by Markowitz (1952), Edwards (1954), and used by Kahneman and Tversky (1979) prospect theory to accommodate evidence that behavior of subjects is driven by gains and losses relative to a reference point and not by final wealth positions as in expected utility theory. Suppose there are n states of the world and let a_i be the consequence of lottery (or act) A in state i. The (subjective) probability of state i is denoted by p_i. Moreover, there is a reference point r_i for every state i. The utility of lottery A in Sugden's model is now given by

$$V(A) = \sum_{i=1}^{n} p_i u(a_i, r_i), \tag{2}$$

where $V(A)$ represents the context- and reference-dependent utility of gamble A. The reference point is often assumed to be equal to initial wealth. Sugden's model generalizes prospect theory by allowing the reference point to be state dependent. In contrast to prospect theory, however, there is no probability weighting. This latter distinction limits the descriptive power of Sugden's model since it cannot explain the typical Allais paradoxes.

The CRU generalizes Sugden's model by allowing the reference point to be context dependent, by which is meant that the reference point may differ in different choice situations. More formally, we have

$$A \succ B \Leftrightarrow \sum_{i=1}^{n} p_i u[a_i, r_i(A, B)] > \sum_{i=1}^{n} p_i u[b_i, r_i(A, B)] \tag{3}$$

This is the most general expression of CRU, where $r_i(A, B)$ is the reference level for state i in this choice. Special forms can be obtained by specific hypotheses on functional forms of the utility function and on how reference point depends on the choice situation. The hypothesis put forward by Bleichrodt and Schmidt (2007) and pursued in this article is that when the initial endowment is zero, the reference point is the maximum of the lowest outcomes of the two gambles in a choice. Assuming that utility depends on the difference of outcome and reference point, we get $u[a_i, r_i(A, B)] = u[a_i - \max\{\min_i(a_i), \min_i(b_i)\}]$. Bleichrodt and Schmidt (2007) showed that with this specification, CRU can explain many classic deviations from expected utility such as Allais paradoxes, preference reversals, and the disparity between willingness-to-pay and willingness-to-accept.

For the lotteries in Table 1, we have under CRU $A \succ B$ as in expected utility theory if $pu(a) > (p+q)u(b)$. However, for the choices between B and C and A and C, the maximum of minimal outcomes is c such that $B \succ C \Leftrightarrow (p+q)u(b-c) + (1-p-q)u(-c) > 0$ and $C \succ A \Leftrightarrow 0 > pu(a-c) + (1-p)u(-c)$.

2.3 Numerical predictions

Our experiment is devised to test these predicted violations of transitivity in CDG and CRU. In our first series, for example, $A =$ (\$100, 0.5; \$0), $B =$ (\$50, 0.9; \$0), and $C =$ \$37 for sure. The cycle $A \succ B$, $B \succ C$, and $C \succ A$ is implied by both CDG and CRU, given plausible values of the parameters for these models.

For the special case of CDG given by $u(a) = a^\alpha$ and $v(a) = a^\beta$, the cycle is implied when $\alpha > 0.64$, and $0.47 < \beta < 0.6$, which are compatible with previously published results. For instance, in the standard common ratio effect (Kahneman and Tversky 1979), it has been found that (\$3000, 1) $\succ$ (\$4000, 0.8; \$0) and (\$4000, 0.2; \$0) $\succ$ (\$3000, 0.25; \$0); this result is implied by CDG when $\beta < 0.78$ and $\alpha > 0.78$.

For CRU we approximate the utility function as follows:

$$u(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda\, |x|^\alpha & x < 0 \end{cases} \tag{4}$$

Assuming $\lambda = 2.2$, the same intransitive cycle is implied if $0.64 < \alpha < 1.04$. These parameters are also realistic given previous data (Tversky and Kahneman 1992); the above-mentioned common ratio effect is predicted by CRU for $\lambda = 2.2$ when $\alpha > 0.78$.

3 Transitivity and error models

Testing transitivity with fallible data has been a controversial topic. One approach has been to test weak stochastic transitivity (Tversky 1969). Weak stochastic transitivity (WST) is the property that if $P(A, B) > 1/2$ and $P(B, C) > 1/2$, then $P(A, C) > 1/2$. Tversky (1969) concluded that weak stochastic transitivity was systematically violated by some, but not all, of his participants. Iverson and Falmagne (1985) criticized the statistical test used by Tversky; they reanalyzed Tversky's data and concluded that WST could not be rejected if each participant were allowed to have a different true preference order. However, Iverson et al. (2006) applied a neo-Bayesian analysis to the same data with the conclusion that some of Tversky's participants were indeed intransitive. Brandstaetter et al. (2006) proposed a priority heuristic model that is intransitive, which they found gives a reasonably accurate account of previously published data, including the pattern observed by Tversky (1969). Regenwetter and Stober (2006) argued that WST is not the way to test transitivity since it, however, can be violated even if each "true" pattern is transitive. They tested instead the triangle inequality and concluded that transitivity can be retained for Tversky's data.

In the economics literature, a similar debate has occurred concerning whether phenomena predicted by regret theory, such as predicted violations of transitivity, are "true" or can be attributed to noise or "error" (Birnbaum 2004, 2005b; Humphrey 2001; Loomes et al. 1991; Sopher and Gigliotti 1993; Starmer and Sugden 1998; Starmer 1999). Models of error have been discussed by Carbone and Hey (2000), Harless and Camerer (1994), Hey (2005), Hey and Orme (1994), Luce (1994), Sopher and Gigliotti (1993), Thurstone (1927), and others.

Because we plan to test transitivity, we think it best to use a model of error that is neutral with respect to transitivity. Models of Thurstone (1927), Busemeyer and Townsend (1993), Hey and Orme (1994) implicitly assume or imply transitivity in the absence of error. Our approach is described by Birnbaum (2005b), and is similar to that of Harless and Camerer (1994) and Sopher and Gigliotti (1993). Whereas Harless and Camerer (1994) assumed that matched choices have the same error probability, Sopher and Gigliotti allowed the error rates in different choices to be unequal. Birnbaum (2005b) improvement over those articles is to use repeated presentations of the same choice in order to unambiguously estimate error rates for different choices.

Consider a choice between A and B that is presented twice to the same participants. Some people will choose A both times, some will choose B both times, some will switch from A to B, and some switch from B to A. The model assumes that the probability of switching from A to B is given as follows:

$$P(AB) = p(1-e)e + (1-p)(1-e)e = e(1-e) \tag{5}$$

where p is the probability that a person "truly" prefers A over B and e is the error rate for this choice. Those people who truly prefer A over B have correctly reported their preference the first time and made an error the second time, whereas those who "truly" prefer B have also made one error and one correct response. Notice that this model implies that the probability of switching from A to B equals the probability of switching from B to A, and this value is independent of p.

When there are three gambles A, B, and C, there are eight possible response patterns for paired choices, shown in Table 2. We assume that each person can have a different "true" preference pattern, which may or may not be transitive; these are listed in Table 2.

The probability of showing the intransitive pattern 000 is as follows:

$$\begin{aligned} P(000) = {} & p_{000}(1-e_1)(1-e_2)(1-e_3) + p_{001}(1-e_1)(1-e_2)e_3 \\ & + p_{010}(1-e_1)e_2(1-e_3) + p_{011}(1-e_1)e_2e_3 \\ & + p_{100}e_1(1-e_2)(1-e_3) + p_{101}e_1(1-e_2)e_3 \\ & + p_{110}e_1e_2(1-e_3) + p_{111}e_1e_2e_3 \end{aligned} \tag{6}$$

where $P(000)$ is the probability of showing the observed intransitive pattern in the data; p_{000} is the probability that a person has 000 as her "true" pattern; and e_1, e_2,

Table 2 Patterns of choice

Notation	Preference pattern	Preference order
000	$A \succ B; B \succ C; C \succ A$	Intransitive
001	$A \succ B; B \succ C; C \prec A$	$A \succ B \succ C$
010	$A \succ B; B \prec C; C \succ A$	$C \succ A \succ B$
011	$A \succ B; B \prec C; C \prec A$	$A \succ C \succ B$
100	$A \prec B; B \succ C; C \succ A$	$B \succ C \succ A$
101	$A \prec B; B \succ C; C \prec A$	$B \succ C \succ A$
110	$A \prec B; B \prec C; C \succ A$	$C \succ B \succ A$
111	$A \prec B; B \prec C; C \prec A$	Intransitive

The pattern predicted by CRU and CDU is 000

and e_3 are the probabilities of making an "error" in expressing preference on the three choices. There are seven other equations like the above for the other seven observed patterns.

This study uses two types of replications: the same choice can be presented exactly the same way, or it can be presented with the positions of the two gambles counterbalanced. These features permit the estimation of error terms. With two replications, there are 64 possible response patterns (8×8), and the equations (as in Expression 6) can be expanded to allow for up to six errors or correct reports.

This "true and error" model with replications is neutral with respect to the issue of transitivity. The transitive model is a special case of this model in which parameters representing true probabilities of intransitivity are fixed to zero; i.e., $p_{000} = p_{111} = 0$. We can test transitivity by comparing the fit of the transitive model to the general model in which all the parameters are free.

4 Method

Participants chose between gambles by viewing the choices via the Internet and clicking a button beside the gamble in each choice they would rather play. Gambles were described in terms of containers holding 100 tickets from which one would be chosen at random to determine the prize. They were displayed as in the following example:

```
Which do you choose?
  A:  50 tickets to win $100
      50 tickets to win $0
OR
  B:  win $45 for sure
```

There were 20 choices. The first two assessed risk aversion. The other 18 were composed of three series of six choices each, each designed to test predictions of CRU and CDG. Each series was composed of three choices testing transitivity with each choice counterbalanced for position (first or second gamble). The 18 were intermixed with order restricted so that no two trials from the same group of three would appear on successive trials.

Table 3 shows the lotteries of Series I, II, and III. The difference between Trials 5 and 20 Trials 8 and 17, and Trials 11 and 14 is just the (first or second) positioning

Table 3 Tests of transitivity, showing trial numbers for Series I

Code	Trial	Choice: First gamble	Choice: Second gamble	% Second gamble: Series I ($b = 50$, $c = 37$)	Series II ($b = 53$, $c = 33$)	Series III ($b = 55$, $c = 30$)
AB	5	A : 50 to win \$100 50 to win \$0	B : 90 to win \$b 10 to win \$0	67	70	70
BC	8	B : 90 to win \$b 10 to win \$0	C : \$c for sure	33	28	23
CA	11	C : \$c for sure	A : 50 to win \$100 50 to win \$0	48	58	57
BA	20	B : 90 to win \$b 10 to win \$0	A : 50 to win \$100 50 to win \$0	33	31	27
CB	17	C : \$c for sure	B : 90 to win \$b 10 to win \$0	70	79	76
AC	14	A : 50 to win \$100 50 to win \$0	C : \$c for sure	51	44	39

Note that Trials 5, 8, and 11 are the same as 20, 17, and 14, respectively, but counterbalanced for position. Predicted pattern of intransitivity is $A \succ B \succ C \succ A$ for CRU and CDU

of lotteries. Consider Series I: Trial 5 was a choice between $A = (\$100, 0.5; \$0)$ and $B = (\$50, 0.9; \$0)$, in Trial 8, B is compared with $C = \$37$ for sure, and Trial 11 is a choice between C and A. In the notation of Table 1, $a = \$100$, $b = \$50$, and $c = \$37$. Series II and III were the same as Series I, except that $b = \$53$, and $c = \$33$ in Series II; and $b = \$55$ and $c = \$30$ in Series III.

Two groups of participants were tested. A group of college undergraduates performed the 20 choices twice, separated by four other intervening tasks that required about 20 min. The 127 college students were tested in labs containing Internet-connected computers. They participated as one option to fulfill an assignment in lower division psychology. Of these, 54% were female; 87% were 20 years or younger, and no one was older than 26.

A second group of 162 participants was recruited via the Web, who participated with the understanding that one would be chosen to receive the prize of one of their 20 chosen gambles. Out of the 162 Web recruits, 66% were female; 35% were 20 years or less, and 15% were over 40.

Complete materials can be examined at the following URLs: http://psych.fullerton.edu/mbirnbaum/decisions/Schmidt_Ulli.htm; http://psych.fullerton.edu/mbirnbaum/decisions/Schmidt_Ulli_1.htm.

5 Results

The percentages of people who chose the second gamble in each pair are shown in Table 3 for all the three series. If most people followed the predictions of CRU or CDG

Table 4 Response patterns for Series I

Pattern	First rep			Second rep			First and second
	#5, 8, 11	#20, 17, 14	Both	#5, 8, 11	#20, 17, 14	Both	Combined
000	2	4	1	2	2	0	0
001	20	20	10	35	26	23	8
010	6	7	5	7	10	6	4
011	8	9	5	4	6	3	3
100	35	35	25	30	36	26	15
101	27	30	16	19	24	14	7
110	22	18	13	28	18	15	7
111	7	4	1	2	5	2	1
Total	127	127	76	127	127	89	45

Responses to trials #20, 17, and 14 have been reflected to correct for the counterbalancing of position. Only one person showed the predicted pattern of intransitivity on both sets of Choices 5, 8, and 11, and 20, 17, and 14 in the First replicate. However, 25 repeated the transitive pattern $B \succ C \succ A$

model, with the parameters we assumed, the choice proportions should be less than 0.5 in the first three rows of Table 3 (Choices 5, 8, and 11) and greater than 0.5 in the second three rows (Choices 20, 17, and 14) of the same table. Instead, we see that the majority prefers ($50, 0.9; $0) over ($100, 0.5; $0) in Choices 5 and 20, so the modal choices are transitive; similar results were obtained for $b = \$53$ and $55 in Series II and III, respectively.

Table 4 shows the number of people who showed each pattern of preference on Choices 5, 8, and 11, on (reflected) Choices 20, 17, and 14, and on both of these sets of trials on Replicates 1 and 2 of Series I. Patterns 000 and 111 are intransitive; the other six patterns are transitive. The pattern of violations predicted by CRU and CDG is 000, and the 111 pattern would be consistent with Bordley (1992). Only one person repeated the predicted pattern 000 of intransitivity in the first replicate and no one repeated this same pattern on the second replicate of Series I. Only one person repeated the opposite intransitive pattern, 111, in both replicates of Series I.

Tables 5 and 6 show the same analysis of response patterns as Table 4 for Series II and III, respectively. Data from the sample recruited via the Web are shown in Table 7 for all the three series. Results in Tables 5, 6, and 7 are all quite similar to those in Table 4: very few people showed intransitive orders, and almost no one repeated intransitive patterns (only two in Table 5, two in Table 6, and four in Table 7).

Table 8 shows how the data of Series I are partitioned for the fit of the true and error model. Data are partitioned into the number of people who showed each pattern repeatedly (both), and the average number who showed each pattern on either the first three choices of Table 3 or the second three choices of Table 3 (with positions reversed), but not both. By construction, these 16 frequencies are mutually exclusive and sum to the number of participants (127).

When the true and error model is fit to the data of each replicate of each series separately, it is found that intransitive patterns are estimated to be low in probability, as one might expect from the small numbers who exhibited these patterns repeatedly.

Table 5 Response patterns for Series II

Pattern	First rep			Second rep			First and second
	#10, 4, 7	#19, 15, 13	Both	#10, 4, 7	#19, 15, 13	Both	Combined
000	2	1	0	3	1	1	0
001	24	23	16	31	32	26	11
010	5	5	4	5	7	4	3
011	5	9	3	2	2	0	0
100	27	36	20	21	32	20	15
101	41	39	29	35	36	28	17
110	17	13	9	26	16	15	5
111	6	1	1	4	1	0	0
Total	127	127	82	127	127	94	51

Each entry is the number of participants who showed each choice combination

Table 6 Response patterns in Series III

Pattern	First rep			Second rep			First and second
	#6, 3, 12	#18, 9, 15	Both	#6, 3, 12	#18, 9, 15	Both	Combined
000	2	0	0	2	1	1	0
001	25	24	15	33	27	23	12
010	5	5	3	5	5	4	3
011	4	7	2	1	3	1	1
100	34	26	24	24	25	20	14
101	39	44	30	37	46	33	21
110	14	19	13	23	19	17	9
111	4	2	1	2	1	0	0
Total	127	127	88	127	127	99	60

Table 7 Response patterns in Web participants

Pattern	Series I			Series II			Series III		
	#5, 8, 11	#20, 17, 14	Both	#10, 4, 7	#19, 15, 13	Both	#6, 3, 12	#18, 9, 15	Both
000	2	4	0	1	2	0	4	0	0
001	20	25	14	30	30	21	22	24	18
010	11	8	5	9	7	4	6	8	3
011	8	7	4	8	5	0	5	5	2
100	44	37	31	36	37	25	37	36	28
101	37	36	24	46	50	35	59	59	48
110	31	35	24	23	21	14	19	19	13
111	7	8	3	7	8	1	7	8	1
Total	160	160	105	160	160	100	159	159	113

Notes: Each choice was repeated only with position counterbalanced. Totals do not sum to number of participants (162) due to occasional skipped items

Table 8 Data partitioned into conjunction and union excluding conjunction (Series I)

Pattern	First rep		Second rep	
	Both	Union-Conj	Both	Union-Conj
000	1	2	0	2
001	10	10	23	7.5
010	5	1.5	6	2.5
011	5	3.5	3	2
100	25	10	26	7
101	16	12.5	14	7.5
110	13	7	15	8
111	1	4.5	2	1.5
Total	76	51	89	38

These are the data to which the model is fit, to minimize the chi-square between predicted and obtained frequencies. The 16 entries must sum to the number of participants, so there are 15 degrees of freedom in the data

Table 9 Parameters estimated from true and error model (Series I)

Pattern	First rep #5, 8, 11 and #20, 17, 14		Second rep # 5, 8, 11 and 20, 17, 14	
	Full model	Transitive	Full model	Transitive
000	0.006	(0)	0.000	(0)
001	0.154	0.170	0.265	0.264
010	0.055	0.057	0.063	0.069
011	0.064	0.063	0.029	0.021
100	0.313	0.315	0.288	0.285
101	0.239	0.214	0.163	0.203
110	0.160	0.181	0.173	0.158
111	0.009	(0)	0.020	(0)
χ^2	6.51	7.60	3.15	8.57

Values are estimates of probability of each "true" preference pattern. Values in parentheses are fixed. Estimated error terms are 0.09, 0.08, and 0.10 in the first repetition, and 0.10, 0.08, and 0 in the second repetition. Although one or two participants may have been systematically intransitive in Series I, a good fit is still obtained when we assume there were no intransitive participants

Parameter estimates for Replicates 1 and 2 of Series I are shown in Table 9. According to the model, fewer than 2% of the participants were intransitive. Both the transitive and full models provided satisfactory fits to the data. Similar results (not shown) were obtained with Series II and III and for the Web data.

The Chi-Squares (indices of lack of fit) for unconstrained and transitive models are shown in Table 10. None of these values is significant. The difference in fit between the transitive and free (allowing intransitivity) models is not significant in any of the nine tests ($\alpha = 0.05$), indicating that we can retain the hypothesis that no one was truly intransitive [$P(000) = P(111) = 0$] in Series I, II, and III of Lab or Web data.

The error terms (e_1, e_2, e_3) can be estimated directly from the number of participants who reverse preferences between repetitions. This method may not be optimal to minimize the test statistic, but it has the advantage that the estimated error terms are independent of assumptions concerning transitivity. Between the first and second

Table 10 Chi-squares indices of (lack of) fit for transitive model and unconstrained model

Data set	Model (degrees of freedom)		
	Transitive (7)	All free (5)	Difference (2)
Series I rep 1	7.60	6.51	1.08
Series I rep 2	8.57	3.15	5.42
Series II rep 1	4.42	4.13	0.30
Series II rep 2	5.05	2.16	2.88
Series III rep 1	6.45	5.16	1.30
Series III rep 2	4.84	2.87	1.97
Series I Web	8.50	2.58	5.92
Series II Web	3.72	3.72	0.00
Series III Web	3.45	3.30	0.15

Notes: Critical values of chi-square with 7, 5, and 2 degrees of freedom are 14.1, 11.1, and 6.0 for $\alpha = 0.05$, respectively. None of the values is significant, indicating that the transitive model provides a satisfactory description of the data

repetitions, the average rate of agreement over all of the 20 choices was 85.6%. Within replicates, the agreement between the same choices with the order reversed was 88.5%. The correlation between these two estimates of "error" was 0.63, indicating individual differences in the error rate; some people are less reliable than others. Assuming 87% agreement, there are 13% preference reversals $[2e(1-e)]$, which corresponds to an average error rate of $e = 0.07$.

For Series I Replicate 1, there were 24 who reversed preferences between Trial 5 and 20, 25 who reversed preferences between Trials 8 and 17, and only nine who reversed preferences between Trials 11 and 14. These correspond to error rates of $e_1 = 0.11$, $e_2 = 0.11$, and $e_3 = 0.04$. These estimates assume nothing about transitivity. With the error rates fixed to these values, $\chi^2 = 7.09$ for the model with probabilities of all the sequences free, and $\chi^2 = 8.71$ for the transitive model. Results for the other series and replicates yielded similar conclusions: the success of the transitive model does not depend on its freedom to estimate error rates to optimize fit to transitivity.

We can illustrate the power of the statistical tests by adding hypothetical participants who repeated the predicted 000 pattern. We again fit data of Replicate 1 of Series I, holding the error rates fixed to the values estimated from observed preference reversals between replicates, and we added 1, 2, 3, 4, or 5 hypothetical people who repeated the predicted 000 pattern. The χ^2 values are 13.44, 21.9, 33.03, 46.07, and 60.47 for 1, 2, 3, 4, and 5 added cases, respectively. Had there been just three people who repeatedly showed the predicted pattern of CRU and CDG, we could have rejected the purely transitive model in favor of the hypothesis that a small percentage of people are truly intransitive. When people are relatively consistent in their preferences, as they were in this study, estimated error rates are small, and the test statistic is very sensitive to cases where people show a repeated pattern violating the model.

6 Discussion and conclusions

Our data provide no significant evidence that people are systematically intransitive for the choices we tested, which were chosen based on predictions of CRU and CDG. We found no significant violations of transitivity of the type predicted by CRU and CDG, nor did we find significant violations of the opposite type, as predicted by Bordley (1992) model. Satisfaction of transitivity does not rule out CRU or CDG, however, owing to those models allowing transitive preferences for this experiment with different parameters. In other words, our results do not rule out intransitivity for choices not yet tested.

The success of transitivity in our data is compatible with findings of Birnbaum and Gutierrez (2007) who tested transitivity using Tversky's (1969) gambles. Brandstaetter et al. (2006) noted that their priority heuristic model implies that people should systematically violate transitivity with Tversky's choices. As we observed in the data of this study, Birnbaum and Gutierrez also found no evidence of systematic intransitivity, contrary to the conclusions of Tversky (1969) and Brandstaetter et al. (2006).[1] With the Tversky gambles, Birnbaum and Gutierrez found that the vast majority of data that were internally consistent (where people agreed with their own choices between repetitions) followed the transitive order matching expected value, transfer of attention exchange, and cumulative prospect theory, which all made the same predictions in that study. Interestingly, this consensus among people occurred despite the fact that Tversky's gambles were designed to have nearly identical expected values.

The data of this study show greater individual differences among people for their transitive orders than found by Birnbaum and Gutierrez. There are four popular transitive patterns in our data: 001, 100, 101, and 110. In this study, we have a sure thing, a highly probable prize and a medium probability prize. The Tversky gambles, in contrast, included no sure thing; perhaps our larger individual differences arise because of this feature of our study. Although people disagreed with each other more in this study than others, people were fairly consistent with their own choices in this study. Participants agreed with their own judgments 87% of the time on average. In Birnbaum (1999), the lab sample agreed with their own judgments 82% of the time. Perhaps this higher internal consistency was facilitated by many repetitions of the same or similar choices in this study.

Other studies using similar structure for the set of gambles (Table 1) argued that violations of transitivity were "real." Loomes et al. (1989, 1991) and Starmer and Sugden (1998) found that the pattern of intransitivity predicted by regret theory was more frequent than the opposite pattern. They used choice problems similar to ours, but presented them using states-of-the-world format. This format allows for statewise comparisons of consequences between lotteries, which is necessary for regret effects to occur. Interestingly, in a states-of-the-world format, regret theory implies exactly the opposite cycles from those predicted by CRU and CDG.

As noted above, the conclusions of Loomes et al. remain controversial, because the reported pattern of asymmetric intransitivities might have resulted from response errors, event-splitting effects, or other complications, rather than from "real" intransitivity (Humphrey 2001; Sopher and Gigliotti 1993; Starmer and Sugden 1998). Nevertheless, our finding of transitivity does not contradict regret theory because that theory does not predict the intransitivity when gambles are presented in the format

[1] The models of Tversky (1969) and Brandstaetter et al. (2006) model do not imply intransitivity in this study. They predict, instead, that majority choices should exhibit the transitive order $C \succ B \succ A$, whereas the observed modal choices in Series II and III are $B \succ A \succ C$.

that we used. Our method of using replications to estimate the errors could be applied in that situation to determine if there are "true" violations of transitivity produced by regret.

Blavatskyy (2003) reported a large incidence of violations of transitivity using lotteries that fit the structure of Table 1. He postulated a heuristic of relative probability comparison (see also Blavatskyy 2006), which implies the same intransitive patterns as do CRU and CDG. In his experiments, about 55% of subjects indeed exhibited these cycles. However, his study is difficult to compare with ours because lotteries were represented by natural frequencies in a sample of nine previous observations, without any specified probability information. His form of presentation may well be crucial to the effect he reported.

The CDG and CRU models, as described here, account for standard paradoxes but they do not account for the "new paradoxes" described by Birnbaum (1999, 2004, 2005a, b, c). They have in common that "sure things" introduce additional considerations to risky decision making that can create intransitivity. Had the predicted pattern of intransitivity been observed, we would have been able to refute a large class of transitive utility models, including Birnbaum's transfer of attention exchange (TAX) model, which accounts for the new paradoxes. This would have encouraged us to revise CDG and CRU theories to explain those phenomena or to revise TAX to incorporate, for example, reference levels that depend on the best lowest consequence, and therefore to account for intransitive preference.

In conclusion, we tested for violations of transitivity that were predicted by two models with parameters chosen to explain common findings. When data are analyzed using an error model in which different people can have different "true" preference patterns, but vary in their responses to the same choices because of "errors," we find no evidence to reject the hypothesis that everyone had a transitive preference order.

Acknowledgments Support was received from National Science Foundation Grants, SES 99-86436, and BCS-0129453.

References

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399–407.

Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes, 95*, 40–65.

Birnbaum, M. H. (2005a). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty, 31*, 263–287.

Birnbaum, M. H. (2005b). Testing properties of decision making models with fallible data. Working manuscript, available from M. Birnbaum.

Birnbaum, M. H. (2005c). Three new tests of independence that differentiate models of risky decision making. *Management Science, 51*, 1346–1358.

Birnbaum, M. H., & Gutierrez, R. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes, 104*, 97–112.

Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes, 77*, 44–83.

Blavatskyy, P. (2003). Content-dependent preferences in choice under risk: Heuristic of relative probability comparisons. International Institute for Applied Systems Analysis. Interim Report 03-031, http://www.iiasa.ac.at/Publications/Documents/IR-03-031.pdf.
Blavatskyy, P. (2006). Axiomatization of a preference for most probable winner. *Theory and Decision*, *60*, 17–33.
Bleichrodt, H., & Schmidt, U. (2002). A context-dependent model of the gambling effect. *Management Science, 48*, 802–812.
Bleichrodt, H., & Schmidt, U. (2007). Context- and reference-dependent utility. Working Paper, available from U. Schmidt.
Bordley, R. F. (1992). An intransitive expectations-based Bayesian variant of prospect theory. *Journal of Risk and Uncertainty, 5*, 127–144.
Bordley, R., & Hazen, G. B. (1991). SSB and weighted linear utility as expected utility with suspicion. *Management Science, 37*, 396–408.
Brandstaetter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Choices without tradeoffs. *Psychological Review*, *113*, 409–432
Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review, 100*, 432–459.
Carbone, E., & Hey, J. D. (2000). Which error story is best? *Journal of Risk and Uncertainty, 20*(2), 161–176.
Diecidue, E., Schmidt, U., & Wakker, P. P. (2004). The gambling effect reconsidered. *Journal of Risk and Uncertainty, 29*, 241–259.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380–417.
Fishburn, P. C. (1980). A simple model for the utility of gambling. *Psychometrika, 45*, 435–448.
Fishburn, P. C. (1982). Nontransitive measurable utility. *Journal of Mathematical Psychology, 26*, 31–67.
Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty, 4*, 113–134.
Fishburn, P. (1992). Nontransitive preferences and normative decision theory. In J. Geweke (Ed.), *Decision making under risk and uncertainty: New models and empirical findings* (pp. 3–10). Dordrecht: Kluwer.
Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica, 62*, 1251–1290.
Hey, J. D. (2005). Comparing theories: What are we looking for? In U. Schmidt & S. Traub (Eds.), *Advances in public economics: Utility, choice, and welfare: A Festschrift for Christian Seidl* (pp. 213–234). Dordrecht, The Netherlands: Springer.
Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62*, 1291–1326.
Humphrey, S. J. (2001). Non-transitive choice: Event-splitting effects or framing effects? *Economica, 68*, 77–96.
Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences, 10*, 131–153.
Iverson, G. J., Myung, J. I., & Karabatsos, G. (2006). Intransitivity of preference: Revisited. Working Paper, available from Geoffrey J. Iverson, Institute for Mathematical Behavioral Sciences, UCI, Irvine, CA 92697, USA.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.
Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty, 9*, 151–172.
Loomes, G., Starmer, C., & Sugden, R. (1989). Preference reversal: information-processing effect or rational non-transitive choice? *Economic Journal, 99*, 140–151.
Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica, 59*, 425–440.
Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92*, 805–824.
Loomes, G., & Taylor, C. (1992). Non-transitive preferences over gains and losses. *The Economic Journal, 102*, 357–365.
Luce, R. D. (1994). Thurstone and sensory scaling: Then and now. *Psychological Review, 101*, 271–277.
Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates.
Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy, 60*, 151–158.

Marley, A. A. J., & Luce, R. D. (2005). Independence properties vis-à-vis several utility representations. *Theory and Decision, 58*, 77–143.

Regenwetter, M., & Stober, C. (2006). Testing transitivity by testing the triangle inequality instead of weak stochastic transitivity. Edwards Bayesian Research Conference, Fullerton, CA, January, 2006.

Schmidt, U. (1998). A measurement of the certainty effect. *Journal of Mathematical Psychology, 42*, 32–47.

Sopher, B., & Gigliotti, G. (1993). Intransitive cycles: Rational Choice or random error? An answer based on estimation of error rates with experimental data. *Theory and Decision, 35*, 311–336.

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

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

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

Stevenson, M. K., Busemeyer, J. R., & Naylor, J. C. (1991). Judgment and decision-making theory. In M. Dunnette & L. M. Hough (Eds.), *New handbook of industrial-organizational psychology* (pp. 283–374). Palo Alto, CA: Consulting Psychologist Press.

Sugden, R. (2003). Reference-dependent subjective expected utility. *Journal of Economic Theory, 111*, 172–191.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286 (Reprinted in 1994, 101, 266–270).

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.

Wu, G., Zhang, J., & Gonzalez, R. (2004). Decision under risk. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 399–423). Oxford: Oxford University Press.