Kiel

Working Papers

Kiel Institute
for the World Economy

An Experimental Investigation of Violations of Transitivity in Choice under Uncertainty

by Michael H. Birnbaum and Ulrich Schmidt

No. 1396 | January 2008

Web: www.ifw-kiel.de

Kiel Institute for the World Economy
Duesternbrooker Weg 120
24105 Kiel, Germany

Kiel Working Paper No. 1396

An Experimental Investigation of Violations of Transitivity in Choice under Uncertainty

by

Michael H. Birnbaum and Ulrich Schmidt

January 2008

An Experimental Investigation of Violations of Transitivity in Choice under Uncertainty

Michael H. Birnbaum

Department of Psychology
California State University, Fullerton

Ulrich Schmidt

Department of Economics
University of Kiel
and
Kiel Institute for the World Economy

Mailing Address:

Prof. Michael H. Birnbaum
Dept. of Psychology
CSUF H-830M
Fullerton, CA 92834-6846

Email: mbirnbaum@fullerton.edu

Phone: 714-278-2102

Acknowledgments: We thank the editor and two referees for very helpful comments. Support was received from National Science Foundation Grants, SES 99-86436, and BCS-0129453.

January 2008

An Experimental Investigation of Violations of Transitivity in Choice under Uncertainty

Abstract

Several models of choice under uncertainty imply systematic violations of transitivity of preference. Our experiments explored whether people show patterns of intransitivity predicted by these models. 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 true preferences that does not need to be transitive. Error rate for a choice is estimated from preference reversals between repeated presentations of the same choice. Our results showed that very few people repeated intransitive patterns. We can retain the hypothesis that transitivity best describes the data of the vast majority of participants.

Key words: decision making, errors, regret theory, transitivity

JEL classification: C91, D81

1 Introduction

The most popular theories of decision making under risk and uncertainty assume that people behave as if they compute values (or "utilities") for the alternatives and choose (or at least, tend to choose) the alternative with the highest computed value. This class of models includes expected utility theory (EU), cumulative prospect theory (CPT), prospective reference theory (PRT), transfer of attention exchange (TAX), gains decomposition utility (GDU) and many others (Luce, 2000; Marley & Luce, 2005; Starmer, 2000; Tversky & Kahneman, 1992; Wu, Zhang, & Gonzalez, 2004; Viscusi, 1989). Although these models can be compared by means of special experiments testing properties that distinguish them (Birnbaum, 1999, 2005a, 2005b; Camerer, 1989, 1992; Harless & Camerer, 1994; Hey & Orme, 1994), 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 & Hazen, 1991; Anand, 1987), 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 & Sugden (1982) and Bell (1982) as well as Fishburn's (1982) Skew-symmetric bilinear utility], Bordley's (1992) expectations-based Bayesian variant of Viscusi's PRT model, the priority heuristic model (Brandstaetter, Gigerenzer, & Hertwig, 2006), context-dependent model of the gambling effect (CDG, Bleichrodt & Schmidt, 2002) and context- and reference- dependent utility (CRU, Bleichrodt & Schmidt, 2005).

If one could show that people systematically violate transitivity, it means that the first class of models must be either rejected or revised if they hope to describe human behavior. A number of previous studies attempted to test transitivity (Birnbaum, Patton, & Lott, 1999; Loomes, Starmer, & Sugden, 1989, 1991; Loomes & Taylor, 1992; Humphrey, 2001; Starmer, 1999; Starmer & Sugden, 1998; Tversky, 1969). However, these studies remain controversial; there is not yet consensus that there are situations that produce substantial violations of transitivity (Luce, 2000; Iverson & Falmagne, 1985; Iverson, Myung, & Karabatsos, 2006; Regenwetter & Stober, 2006, Sopher & Gigliotti, 1993; Stevenson, Busemeyer, & Naylor, 1991). Among others, 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 paper is to empirically test patterns of intransitivity that are predicted by two models (i.e. regret theory and majority rule), using an "error" model that has the promise to be neutral with respect to the issue of transitivity and which seems plausible as a description of repeated choices. A second feature of our experimental design is that it does not confound event-splitting effects (violations of coalescing) with the tests of transitivity. If, on the one hand, violations persist when these factors are controlled, models that predicted those violations gain credibility. On the other hand, the absence of violations when these factors are controlled would be consistent with theories that assume transitivity. Note that our study is only devoted to the actual behavior of subjects, we do not consider whether violations of transitivity can be rational or not.

The rest of this paper is organized as follows. The next section describes predictions of regret theory and majority rule with respect to transitivity and reviews earlier experimental studies. Section 3 presents the error model. Design and results of experiments are reported in Sections 4 and 5. Section 5 concludes that despite powerful tests, we find little evidence of systematic violation of transitivity.

2 Integrative Contrast Models: Regret Theory and Majority Rule

Regret theory and majority rule are both special cases of the integrative contrast model, which can be written as follows:

$$A \succ B \Leftrightarrow \sum_{i=1}^{n} \phi(E_i)\psi(a_i, b_i) \qquad (1)$$

where $A \succ B$ denotes *A* is preferred to *B*, a_i and b_i are the subjective values of the consequences of *A* and *B* for state of the world E_i, $\phi(E_i)$ is the subjective probability of event E_i, and ψ maps a contrast in consequences for this state of the world (or dimension) into a preference between the gambles. It is assumed that $\psi(a, b) = -\psi(b, a)$ and $\psi(a, b) = 0$ if $a = b$. When probabilities are specified, it is assumed that $\phi(E_i) = p_i$, where p_i is the probability of E_i.

According to *regret theory* (Loomes & Sugden, 1982; Bell, 1982), people compare the prizes for each state of the world and make choices in order to minimize regret. For example, suppose:

$$\phi(E_i)\psi(a_i, b_i) = p_i(x_i - y_i)^3 \qquad (2)$$

where p_i is the probability that state of the world *i* occurs; x_i and y_i are the cash payoffs of *A* and *B* in this state of the world, respectively. Note that in this case, large differences in payoff produce extra large regrets (i.e. the regret function is convex), as proposed by regret theory. Note as well that the cubic function retains the signs (directions) of the regrets.

Consider Choices 11, 5, and 13 of Table 1. Loomes, Starmer, & Sugden (1991) reported that these choices produced the greatest percentage of intransitive cycles (28%, see p. 437). In addition, this set was chosen because the observed incidence of this intransitive cycle exceeded the frequency of the most common transitive preference pattern that differed from it by only one choice. According to Equations 1 and 2, $\sum_{i=1}^{n} p_i(x_i - y_i)^3 = -18.7$, so $B \succ A$; $\sum_{i=1}^{n} p_i(x_i - y_i)^3 = -8.3$, so $C \succ B$; however, $\sum_{i=1}^{n} p_i(x_i - y_i)^3 = 45.2$, so $A \succ C$, violating transitivity.

Insert Table 1 about here.

The *majority rule model* (sometimes called the *most probable winner* model) is also a special case of Equation (1) in which the contrast functions are as follows:

$$\psi(a_i, b_i) = \begin{cases} 1, & a_i \succ b_i \\ 0, & a_i = b_i \\ -1, & a_i \prec b_i \end{cases} \quad (3)$$

According to this model, people should prefer *A* to *B* (it has higher values on two of the three dimensions), they should prefer *B* to *C*, and *C* to *A*, for the same reasons. Thus, majority rule also predicts violations of transitivity, but of the opposite pattern from that predicted by regret theory. [In this case, $\sum_{i=1}^{n} p_i \psi(a_i, b_i) = 0.4$, $\sum_{i=1}^{n} p_i \psi(b_i, c_i) = 0.4$, yet $\sum_{i=1}^{n} p_i \psi(a_i, c_i) = -0.2$; therefore, $A \succ B$, $B \succ C$, but $C \succ A$.]

A problem in previous empirical tests of regret theory is that certain confounds were present in those studies (Humphrey, 2001; Starmer & Sugden, 1998). Probably the most important problem was that different forms of the gambles were used in different choices. *A* and *B* were presented for comparison as three-branch gambles (as illustrated in Choice 11 of Table 1, $A = (\$10, 0.4; \$3, 0.3; \$3, 0.3)$, $B = (\$7.5, 0.4; \$7.5, 0.3; \$1, 0.3)$). However, the so-called choice between *B* and *C* was actually presented in a form in which the two upper branches of *B* and C were coalesced, creating two new gambles, $B' = (\$7.5, 0.7; \$1, 0.3)$, $C' = (\$5, 0.7; \$5, 0.3)$). The so-called choice between $C = (\$5, 0.4; \$5, 0.3; \$5, 0.3)$ and *A* was presented with the two lower branches coalesced, creating two other new gambles, C'' and A'' where $C'' = (\$5, 0.4; \$5, 0.6)$, and $A'' = (\$10, 0.4; \$6, 0.3)$. According to the transitive TAX model (e.g., Birnbaum & Navarrete, 1998), with parameters estimated from previous data, splitting and coalescing of branches could account for the apparent violations of transitivity. According to TAX with parameters estimated from prior research we have $U(A)$ = 4.33, $U(B)$ = 5.33; $U(C)$ = 5.00; $U(B')$ = 3.79; $U(C')$ = 5.00, $U(A'')$ =

5.01; $U(C'')$ = 5.00. Thus, this TAX model implies that $A \prec B$, $B' \prec C'$, and $C'' \prec A''$, which is intransitive only if we assume coalescing.

In this paper, we keep all gambles in the same three-branch form to avoid this confound with coalescing. Starmer & Sugden (1998) and Humphrey (2001) recognized this confound and controlled for it by presenting choices in fully split forms or by using a different format for display ("strip") in which gambles were presented in fully coalesced form. But those articles had a second problem; namely, they used asymmetry of different types of intransitivity as evidence of intransitivity. As we show in the next section, such asymmetry is entirely compatible with an error model in which people make occasional "errors" in determining or reporting their preferences, even if they are truly transitive.

3 Error Model

In a study with three choices, *AB*, *BC*, and *CA* (as in Table 1), there are eight possible outcomes: 000, 001, 010, 011, 100, 101, 110, and 111, where 0 denotes choice of the first gamble, and 1 denotes choice of the second gamble. The pattern, 110 denotes preference for the second gamble in the first two choices and preference for the first gamble in the third (i.e., $A \prec B$, $B \prec C$, and $C \succ A$). The pattern, 000, represents intransitive cycle, $A \succ B$, $B \succ C$, and $C \succ A$, and 111 is its reverse cycle, $A \prec B$, $B \prec C$, and $C \prec A$, respectively. Suppose a person has the true pattern, 110, but sometimes makes random "errors" in discovering or reporting her preferences. If so, it takes only one error to produce the 111, but it takes two errors to produce the intransitive pattern, 000.

In the model of Sopher and Gigliotti (1993), the probability that a person exhibits intransitive choices 111 and has the true preference pattern, 110, is given as follows:

$$P(111 \cap 110) = p_{110}(1-e_1)(1-e_2)e_3, \qquad (4)$$

where $P(111 \cap 110)$ is the probability that a person shows the observed intransitive pattern 111 and has the true pattern 110, p_{110} is the probability of the true pattern 110, and e_1, e_2, and e_3 are the probabilities of making errors on the *AB*, *BC*, and *CA* choices, respectively. It is assumed that the errors are independent and that $0 < e_i < ½$. To show this pattern of observed preferences, this person made no errors on the first two choices and made an error in the third choice. In this model, the probability of showing a given data pattern is the sum of eight terms representing the eight possible true patterns and the probability of showing a given data pattern given each true pattern.

For example, the probability of showing the 111 data pattern is given as follows:

$$P(111) = p_{000}e_1e_2e_3 + p_{001}e_1e_2(1-e_3) + \ldots + p_{111}(1-e_1)(1-e_2)(1-e_3) \quad (5)$$

There are seven other equations like the above for the other seven observed data patterns. One can fit this model to the observed frequencies of these patterns, as in Sopher and Gigliotti (1993). The predicted frequencies are given by $\hat{f}_i = n\hat{p}_i$, where n is the number of participants, and $i = 000, 001, \ldots, 111$. Parameters are estimated to minimize $\chi^2 = \sum_{i=1}^{8}(f_i - \hat{f}_i)^2 / \hat{f}_i$. However, there are only 7 degrees of freedom in the 8 observed frequencies (since they sum to the number of participants), and there are 3 error terms and 7 parameters representing the eight probabilities, p_{000}, p_{001}, p_{010}, …, (which sum to 1). This model therefore has more parameters than there are degrees of freedom in the data. Unless we make some arbitrary assumptions, or increase the degrees of freedom in the data, this model is under-determined.

Consider the observed frequencies (“data”) in Table 2 representing 200 participants who made three choices. These data resemble results such as shown in Loomes, et al. (1991, p. 437, Table 4). In order to simplify the model, we might assume that the error rates are all equal (as is done by Harless & Camerer, 1994), in which case we can fit the data perfectly with the assumption that 23% of the participants were intransitive, with the pattern 111. We could also fit the data with the assumption that $p_{000} = p_{111} = 0$ (that everyone is transitive), if we allow unequal errors $\hat{e}_1 = 0.01$, $\hat{e}_2 = 0.09$, and $\hat{e}_3 = 0.31$, as in Sopher and Gigliotti (1993). Both of these models correctly predict that 46 people show the 111 data pattern and no one shows the opposite pattern. Finally, note that when we attempt to fit the transitive model with equal errors, the data no longer fit very well. Thus, if we hope to answer the question of transitivity, we need a way to estimate the error rates that is independent of arbitrary assumptions such as transitivity holds or all errors are equal

Insert Table 2 about here.

The error theory of Hey and Orme (1994) assumes an additive error component, as is assumed in models of Thurstone (1927), Luce (1959; 1994), Busemeyer and Townsend (1993), and others. These models assume perfect transitivity in the absence of error and assume that the probability of errors will be related to the distance on an underlying (transitive) continuum. If we want to test transitivity, rather than assume it, however, we cannot use these error models.

An approach that is neutral with respect to the issue of transitivity has been suggested by Birnbaum (2004b) and applied by Birnbaum and Gutierrez (2007) in testing predicted violations of transitivity that are predicted by lexicographic semiorders and reported by Tversky (1969). This approach uses preference reversals with repeated presentations of the same choices. Assume that each person has a "true" preference for each choice, and that each choice can have a different "error" rate. Suppose we present the choice between *A* and *B* twice. The probability that a person will choose *A* the first time and *B* the second time is given as follows:

$$P(AB) = pe_1(1 - e_1) + (1 - p)(1 - e_1)e_1 = e_1(1 - e_1) \quad (6)$$

Where p is the "true" probability of preferring *B* and e_1 is the error rate for the *AB* choice. Similarly, the probability of choosing *B* the first time and choosing *A* on the second replication is also $e_1(1 - e_1)$. There are four frequencies that can be fit by two parameters (*AA*, *AB*, *BA*, *BB*), leaving one degree of freedom to test the model. The error rates for the *BC* and *CA* choices can be estimated in the same way. Use of replications thus provides neutral way to estimate error terms.

In addition to allowing an independent standard for evaluating transitivity, the use of replications also places greater constraint on the estimation of the “true” probabilities of the sequences. In addition to counting the number of times that each person shows a given pattern on one presentation of the three choices, we also count the number of times that each person repeats the same pattern on both replications. Whereas there are just 8 response patterns in an experiment with three choices, for example, there are 64 response patterns when three choices are replicated twice. In sum, use of replications increases the degrees of freedom in the data without adding any new parameters.

4 Experimental Design

Study 1 was conducted with 314 undergraduates enrolled in lower division psychology at California State University at Fullerton (USA). Gambles were described in terms of a container holding 100 tickets numbered from #1 to #100, from which one would be chosen at random to determine the prize. The relationship between tickets and prizes was displayed in matrix format, as shown in Figure 1, which defines payoffs for states of the world in three of the choices used (#4, #5, and #6). Participants viewed choices via the Internet and clicked the button beside the gamble they would rather play. They were informed that at the end of the study, 10 people would be chosen randomly who would receive the prize of one of their chosen gambles.

Insert Figure 1 about here.

There were 15 choices in the study. The first three assessed risk aversion in two-branch gambles and served as a warm-up. There were two replications of three choices each. Position of the two gambles in each pair was counterbalanced between the two replications (Table 1). The difference between Choices 11 and 13, for example, is first or second position of the gambles in the choice. These six choices were alternated with six filler trials. In Series I, *A* = ($10, 0.4; $3, 0.3; $3, 0.3), *B* = ($7.5, .4; $7.5, .3; $1, .3), *C* = ($5, 0.4; $5, 0.3; $5, 0.3)

Subjects were tested either in labs containing Internet-connected computers or via the Internet at times and using computers of participants' own choosing. They participated as one option toward an assignment in lower division psychology. Of these, 60.5% were female; 92% were 21 years or younger, only 1% were older than 26.

Complete materials can be examined at the following URL: http://psych.fullerton.edu/mbirnbaum/decisions/Loomes_table.htm

Study 2 was conducted at the University of Hannover (Germany) with 103 undergraduate economics and management students. Format of gambles and choices were the same as in Study 1, except the materials were printed on paper and presented in a classroom setting. Participants marked their choices in pencil and received a flat payment of 5 Euro. There were 15 choices, including the six choices used in Study 1 (Table 1), three filler choices, plus a second series of six choices to test transitivity. In Series II, A = ($18, 0.3; $0, 0.3; $0, 0.4), B = ($8, 0.3; $8, 0.3; $0, 0.4), and C = ($4, 0.3; $4, 0.3; $4, 0.4). Starmer and Sugden (1998) reported that Series II led to frequent violations of transitivity.

5 Results

The percentages choosing the second gamble in each choice are displayed in Table 1. The observed frequencies for response patterns of Series I of both studies are presented in the left and center portion of Table 3. Results from Starmer and Sugden (1998) are included for comparison. Whereas Starmer & Sugden (1998) reported 20% showing the intransitive pattern 111, we found only 8% and 7% who showed this pattern on Choices #11, 5, and 13 and on #9, 15, and 7 of Study 1, respectively; only 0.6% showed this pattern on both repetitions.

Insert Table 3 about here.

In Study 2, this intransitive pattern was observed with relative frequencies of 6% (#11, 5, and 13), 6% (# 9, 15, and 7), and 3% (both repetitions). In the choices of Series II (Study 2), shown in the right-most portion of Table 3, the intransitive pattern 111 is observed even less often, i.e. 3% (#6, 10, 14), 3% (#4, 12, 8), and 2% (both repetitions), compared to 11% in Starmer and Sugden (1998).

Tables 4 and 5 show how error rates in each choice are estimated from preference reversals between repetitions of the same choices. The three rows in Table 4 show responses to choices between *A* and *B*, *B* and *C*, and *C* and *A*, respectively. The Chi-Square test of independence assesses whether the probability of choice combinations can be represented by the product of probabilities for individual choices. These tests are all significant, clearly violating independence. The Chi-Square tests of the true and error model (also with 1 df), however, are all nonsignificant, indicating that the true and error model can be retained for these data.

Insert Tables 4 and 5 about here.

The error rates are estimated from preference reversals between repeated presentations of the same choice with position of the gambles reversed. In Study 1, these were estimated to be 0.13, 0.16, and 0.14 for the three choices (*AB*, *BC*, *CA*), respectively. In Study 2, the corresponding values were 0.08, 0.14, and 0.13 for the first set of gambles and 0.02, 0.04, and 0.05 for the second set (Table 2). Recall that these estimates of error rates assume nothing about transitivity.

There are 64 possible data patterns in each test of transitivity. But many of these have small frequencies; therefore, the data are partitioned into the number who show each of the eight patterns on both replicates and the average frequency of showing each pattern on either the first or second replicate but not both. The sum of these frequencies adds to the number of participants, leaving 15 degrees of freedom in the data. Tables 6, 7, and 8 show the fit of the “true and error” model to the observed frequencies, using error rates estimated from replications. The purely transitive model gave a good approximation to the data of Study 1; deviations of fit are not significant, $\chi^2(7)$= 14.3. When all parameters were free, the improvement in fit was not significant, $\chi^2(2)$ = 2.85, and the estimated rate of intransitivity of both types was $\hat{p}_{000} + \hat{p}_{111}$ = 3%. Therefore, we can retain the hypothesis that everyone was transitive in Study 1.

Insert Tables 6, 7, and 8 about here.

In Study 2, the fit of the true and error model to the data in Table 7 yielded $\chi^2(9) = 4.53$. However, the fit of the purely transitive model was worse, $\chi^2(11) = 23.72$, suggesting that the 7% estimated incidence of intransitivity is significantly greater than zero. Fitting the true and error model to the second set, in Table 8, the value of $\chi^2(9) = 2.19$, again indicating acceptable fit. With the probabilities of both intransitive patterns set to zero, $\chi^2(11) = 36.0$, which is again significant. This analysis indicates that the estimated rate of 2% intransitivity is "significant," relative to the small error rates of Table 5. With paper and pencil method, people can easily check for consistency between repetitions of the same choice, so these error rates might be lower than they would have been had other procedures been used. Although significant, these rates of violation are very small.

9 Summary and Conclusions

Loomes, Starmer, & Sugden (1989, 1991) and Starmer & Sugden (1998) found that the pattern of intransitivity predicted by regret theory was more frequent than the opposite pattern. As noted above, their asymmetric intransitivites might have resulted from response errors, so inequality of two types of violations is not a test of transitivity. In addition, some of the studies confounded tests of transitivity with event-splitting effects, or other complications, rather than from "real" intransitivity (Humphrey, 2001; Sopher & Gigliotti, 1993; Starmer & Sugden, 1998). Our attempts to replicate these studies yielded data that did not show systematic intransitivity predicted by regret theory; in fact, neither our German nor American samples showed the asymmetry previously reported.

Blavatskyy (2003) reported a substantial incidence of violations of transitivity. He postulated a heuristic of relative probability comparison (see also Blavatskyy, 2006). 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 format of presentation may well be crucial to the effect he reported. We found few people who repeated the pattern predicted by most probable winner.

The success of transitivity in our data is compatible with findings of Birnbaum and Gutierrez (2007), who tested violations of transitivity predicted by a lexicographic semi-order, as previously studied by Tversky (1969). Brandstaetter, et al. (2006) noted that their priority heuristic model implies that the majority of people should systematically violate transitivity with Tversky's choices. As in the present data, however, Birnbaum and Gutierrez also found very few cases of repeated intransitivity, contrary to the conclusions of Tversky (1969) and Brandstaetter, et al. (2006).[1] With the Tversky gambles, Birnbaum and Gutierrez found that most people satisfied transitivity and agreed with a single order. Interestingly, this consensus among people occurred despite the fact that Tversky's gambles were designed to have nearly identical expected values.

Had either the regret model or the majority rule model been successful in predicting systematic patterns of intransitivity, it would have been a strong point in favor of one of these models. Although they make opposite predictions, some theoreticians find the intuitions of both models appealing. Why not choose the gamble that most often gives the best outcome? Why not choose the gamble that one would least regret? But our data do not confirm these intuitions. Combining these data with those of Birnbaum and Gutierrez for the lexicographic semiorder and Birnbaum and Schmidt (2006) for choice under risk, we think the burden of proof should shift to those who argue that intransitive models are descriptive of more than five percent of the population.

In summary, we searched for violations of transitivity where predicted by two models with

[1] Although the model of Brandstaetter, et al. (2006) model is not always transitive, it does not predict violations of transitivity in any of these studies. In Studies 1-3 it predicts that majority choices should exhibit the transitive order, $C \succ B \succ A$ in all three series, whereas the observed modal choices in Series II and III are $B \succ A \succ C$. In Study 4, this model predicts $C \succ A \succ B$, in agreement with the most frequently repeated pattern. In Study 5 Series II, it predicts $C \succ A \succ B$, which was repeated by only 1 person; instead, the modal pattern was $C \succ B \succ A$.

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 due to “errors,” we find little evidence to refute the hypothesis that nearly everyone had a transitive preference order.

References

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

Bell, D. E. (1982). Regret in decision making under uncertainty. Operations Research, 30, 961-983.

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

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. Journal of Mathematical Psychology, 48(2), 87-106.

Birnbaum, M. H. (2004b). 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). Three new tests of independence that differentiate models of risky decision making. Management Science, 51, 1346-1358.

Birnbaum, M. H. (2006). Testing properties of decision making models with fallible data. Working manuscript, available from M. Birnbaum.

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

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. Journal of Risk and Uncertainty, 17, 49-78.

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.

Birnbaum, M. H., & Schmidt, U. (2006). Testing transitivity in choice under risk. Working paper, available upon request.

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. R. (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. (2005). 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.

Camerer, C. (1989). An experimental test of several generalized utility theories. Journal of Risk and Uncertainty, 2, 61-104.

Camerer, C. (1992). Recent tests of generalizations of expected utility theory. In: W. Edwards (Ed.) Utility: Theories, Measurement, and Applications (pp. 207–251). Norwell: Kluwer Academic Publishers.

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.

Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. Psychological Science, 10, 353-359.

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 (Eds.), Decision making under risk and uncertainty: New models and empirical findings (pp. 3-10). Dordrecht: Kluwer Academic Publishers.

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.

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.

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

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, 805-24.

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

Luce, R. D. (1959). Individual choice behavior. New York: Wiley.

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.

Myung, J. I., Karabatsos, G., & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. Journal of Mathematical Psychology, 49, 205-225.

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.

Schoemaker, P. J., & Hershey, J. C. (1992). Utility measurement: Signal, noise, and bias. Organizational Behavior and Human Decision Processes, 52, 397-424.

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 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.

Table 1. Tests of Transitivity used in Studies 1 and 2 (Series I). Enties show the percentage who chose the second gamble in each choice.

		Choice		N = 314	N = 103	
Type	No.	*first Gamble*	*second Gamble*	Study 1 Series I	Study 2 Series I	Study 2 Series II
AB	11	40 to win 10 30 to win 3 30 to win 3	40 to win 7.5 30 to win 7.5 30 to win 1	34	36	90
BC	5	40 to win 7.5 30 to win 7.5 30 to win 1	40 to win 5 30 to win 5 30 to win 5	64	52	46
CA	13	40 to win 5 30 to win 5 30 to win 5	40 to win 10 30 to win 3 30 to win 3	47	56	20
BA	9	40 to win 7.5 30 to win 7.5 30 to win 1	40 to win 10 30 to win 3 30 to win 3	66	59	10
CB	15	40 to win 5 30 to win 5 30 to win 5	40 to win 7.5 30 to win 7.5 30 to win 1	41	49	53
AC	7	40 to win 10 30 to win 3 30 to win 3	40 to win 5 30 to win 5 30 to win 5	55	44	79

In Series I, *A* = ($10, 0.4; $3, 0.3; $3, 0.3), *B* = ($7.5, .4; $7.5, .3; $1, .3), *C* = ($5, 0.4; $5, 0.3; $5, 0.3). In Series II, *A* = ($18, 0.3; $0, 0.3; $0, 0.4), *B* = ($8, 0.3; $8, 0.3; $0, 0.4), and *C* = ($4, 0.3; $4, 0.3; $4, 0.4).

Table 2. Fit of three models to frequencies of response patterns in a test of transitivity. To distinguish these models, we need an independent way to estimate the error rates.

		Intransitive, Equal Errors		Transitive, Unequal Errors		Transitive Equal Errors	
	Data	$\hat{f}$	$\hat{p}$	$\hat{f}$	$\hat{p}$	$\hat{f}$	$\hat{p}$
000	0	0.0	0.00	0.8	(0)	4.5	(0)
001	2	2.0	0.00	1.5	0.00	4.3	0.00
010	8	8.0	0.00	7.9	0.00	21.6	0.00
011	16	16.0	0.08	16.0	0.13	15.2	0.10
100	18	18.0	0.06	18.0	0.06	20.9	0.00
101	10	10.0	0.04	9.9	0.02	11.3	0.06
110	100	100.0	0.59	99.9	0.79	99.1	0.84
111	46	46.0	0.23	46.0	(0)	23.0	(0)
Chi-Square		0.0		0.76		753.4	

Note: Intransitive with equal errors, the error rate was estimated to be zero; For the transitive solution with equal errors, $\hat{e} = 0.16$; In the transitive solution with unequal errors, $\hat{e}_1 = 0.01$, $\hat{e}_2 = 0.09$, and $\hat{e}_3 = 0.31$.

Table 3. Response Patterns to choices of Series I in both studies and Series II in Study 2.

		Study 1, Series I			Study 2, Series I				Study 2 Series II		
Pattern	Starmer & Sugden (1998)	#11, 5, 13	#9, 15, 7	Both	#11, 5, 13	#9, 15, 7	Both	Starmer & Sugden (1998)	#6, 10, 14	#4, 12, 8	Both
000	6	20	27	4	7	6	2	3	2	0	0
001	7	45	48	27	22	20	15	5	5	6	5
010	15	90	89	55	16	16	11	4	2	3	1
011	16	53	43	12	21	19	11	2	1	1	1
100	8	23	27	10	11	11	5	9	37	37	31
101	7	25	28	10	9	13	5	12	12	12	9
110	14	32	29	9	11	12	5	45	41	41	37
111	17	25	22	2	6	6	3	10	3	3	2
Total	90	313	313	129	103	103	57	90	103	103	86

Patterns for #9, 15, and 7 have been reflected to correct for the counterbalancing of position, as have Responses to #4, 12, and 8. Regret theory implies the 111 pattern, and majority rule implies the 000 pattern of violations of transitivity.

Table 4. Preference Reversals between Repetitions (Study 1)

	X11	X12	X21	X22	$\chi^2(1)$	TRUE+ERROR estimates		$\chi^2(1)$
XY	XX	XY	YX	YY	CHISQ_Indep	*p*	*e*	CHISQ_TE
AB	171	37	36	69	71.56	0.277	0.135	0.01
BC	79	34	51	150	59.14	0.667	0.164	3.38
CA	132	34	41	107	84.91	0.445	0.139	0.65

Totals do not always sum to the number of participants (314) due to skipped items.

Table 5. Preference Reversals between Repetitions (Study 2). Last three rows show Series II.

	X11	X12	X21	X22	$\chi^2(1)$	TRUE+ERROR estimates		$\chi^2(1)$
XY	XX	XY	YX	YY	CHISQ_Indep	*p*	*e*	CHISQ_TE
AB	56	10	5	32	49.96	0.36	0.08	1.63
BC	37	12	13	41	27.21	0.53	0.14	0.04
AC	33	12	12	46	28.54	0.59	0.13	0.00
A'B'	8	2	2	91	62.42	0.92	0.02	0.00
B'C'	52	4	3	44	76.79	0.46	0.04	0.14
A'C'	77	5	4	17	55.77	0.18	0.05	0.11

Table 6. Fit of Purely Transitive Model to Observed Frequencies (Study 1).

				Observed Data		Predicted		Est.
Pattern	#11, 5, 13	#9, 15, 7	Both	Both	OR-Both	Both	OR-Both	$\hat{p}$
000	20	27	4	4	19.5	3.4	23.8	0.00
001	45	48	27	27	19.5	25.0	24.0	0.20
010	90	89	55	55	34.5	52.0	38.2	0.42
011	53	43	12	12	36	15.6	28.6	0.11
100	23	27	10	10	15	9.8	14.2	0.07
101	25	28	10	10	16.5	11.4	16.5	0.09
110	32	29	9	9	21.5	10.1	21.4	0.07
111	25	22	2	2	21.5	5.0	13.9	0.03
Total	313	313	129	129	184	132.4	180.6	

The error terms were estimated from replications only. The best-fit values are 0.13, 0.16, and 0.14 for Choices #11 and 9, #5 and 15, and #13 and 7, respectively. For the transitive model, the value of $\chi^2(7)$= 14.3, which is not significant. Allowing intransitivity, the estimated proportion of intransitive participants was .03; the fit was not significantly improved.

Table 7. Fit of True and Error Model to Observed Frequencies (Study 2, Series I)

	N = 103			Observed Data		Predicted		Est.
Pattern	#6, 10, 14	#4, 12, 8	Both	Both	OR-Both	Both	OR-Both	$\hat{p}$
000	7	6	2	2	4.5	1.7	6.1	0.02
001	22	20	15	15	6	12.6	8.9	0.26
010	16	16	11	11	5	9.3	7.5	0.19
011	21	19	11	11	9	9.3	9.1	0.18
100	11	11	5	5	6	5.4	5.0	0.11
101	9	13	5	5	6	4.7	5.6	0.09
110	11	12	5	5	6.5	5.1	5.4	0.10
111	6	6	3	3	3	2.7	4.8	0.05
Total	103	103	57	57	46	50.7	52.3	

The estimated “true” rate of intransitivity is 7%.

Table 8. Fit of True and Error Model to Observed Frequencies (Study 2, Series II).

	n = 103			Observed Data		Predicted		Est.
Pattern	#11, 5, 13	#9, 15, 7	Both	Both	OR-Both	Both	OR-Both	$\hat{p}$
000	2	0	0	0	1	0.0	1.0	0.00
001	5	6	5	5	0.5	4.7	0.8	0.06
010	2	3	1	1	1.5	1.0	1.0	0.01
011	1	1	1	1	0	0.9	0.4	0.01
100	37	37	31	31	6	31.0	5.2	0.37
101	12	12	9	9	3	8.8	2.8	0.10
110	41	41	37	37	4	35.7	5.2	0.43
111	3	3	2	2	1	1.9	2.5	0.02
Total	103	103	86	86	17	84.1	18.9	

Figure 1. Appearance of three choices in the browser (Study 1). All gambles were presented as three branch gambles using states of the world, matrix format. In Study 2, the materials were printed on paper and participants marked their preferred choices in pencil.