Birnbaum, Michael H.; Schmidt, Ulrich

Working Paper

Empirical Tests of Intransitivity Predicted by Models of Risky Choice

Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2006,10

Provided in cooperation with:
Christian-Albrechts-Universität Kiel (CAU)

Empirical Tests of Intransitivity Predicted by Models of Risky Choice

by Michael H. Birnbaum and Ulrich Schmidt

C | A | U

Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2006-10

Empirical Tests of Intransitivity Predicted by Models of Risky Choice

Michael H. Birnbaum
Department of Psychology
California State University, Fullerton

Ulrich Schmidt
Department of Economics
University of Kiel

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

August 2006

Empirical Tests of Intransitivity Predicted by Models of Risky Choice

Abstract

Recently proposed models of risky choice imply systematic violations of transitivity of preference. Five studies explored whether people show patterns of intransitivity predicted by four descriptive 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 need not be transitive. Error rate for a choice is estimated from preference reversals between repeated presentations of the same choice. Results of five studies 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, gambling effect, reference points, regret, transitivity

JEL classification: C91, D81

1 Introduction

The most popular theories of decision making under risk and uncertainty assume that people compute values (or “utilities”) for the alternatives and choose (or at least, tend 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), 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; 2004a; 2004b; 2005a; 2005b), 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), 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 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 modified to allow such effects. Models that can violate transitivity provide a basis for designing experiments to test transitivity. These studies will explore violations predicted by additive difference models such as regret theory and by the models of Bleichrodt & Schmidt (2002, 2005).

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 intransitive models, 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 the variability of repeated choices. For the lotteries in Studies 1-3, CDG and CRU models make the same pattern of predicted violation given parameters chosen to describe well-known phenomena in risky decision making. In Studies 4 and 5, we investigate intransitivity predicted by additive difference models, including regret theory and majority choice, which make opposite predictions.

The rest of this paper is organized as follows. The next section describes CDG and CRU models and shows their predicted pattern of violation of transitivity; section 3 describes the error model; section 4 describes Studies 1, 2, and 3; the fifth section presents results of three studies, which show that transitive models can be retained for these data. The sixth section presents predictions of regret theory and majority rule, the seventh and eighth sections present methods and results of Studies 4 and 5, and the ninth concludes that despite powerful tests, we do not find evidence of systematic violation where predicted by the four descriptive models tested. Whereas the models and choices of the first three studies involve risky gambles formatted without states of the world, the models and choices of the last two studies involve gambles defined on states of the world.

2 Theoretical Predictions of CDG and CRU

The experimental design of Studies 1-3 involves variations of three lotteries shown in Table 1, where *p* and *q* are probabilities, and $a > b > c$ are monetary consequences. In the terminology of the preference reversal literature, *A* is the "$-bet," *B* is the "p-bet," and *C* is cash.

Insert Table 1 about here.

It will be shown below that CDG and CRU models both imply the intransitive pattern,

and Bordley's (1992, p. 135) intransitive variant of PRT implies the opposite intransitive pattern, $B \succ A$, $A \succ C$, and $C \succ B$.

Before presenting the models, let us introduce some notation. Lotteries are denoted by capital letters *A*, *B*, *C*; X is the set of all 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*.

(i) The Context-Dependent Model of the Gambling Effect(CDG)

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. CDG postulates two distinct utility functions, *u* and *v*, such that

$$A \text{ f } B \Leftrightarrow \begin{cases} \sum_{\iota=1}^{\nu} \pi(\alpha_\iota)\upsilon(\alpha_\iota) > \sum_{\iota=1}^{\nu} \pi(\beta_\iota)\upsilon(\beta_\iota) & A \wedge B \notin \Xi \\ \sum_{\iota=1}^{\nu} \pi(\alpha_\iota)\varpi(\alpha_\iota) > \sum_{\iota=1}^{\nu} \pi(\beta_\iota)\varpi(\beta_\iota) & A \vee B \in \Xi \end{cases} \tag{1}$$

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, Schmidt & Wakker (2003), CDG does not imply violations of first-order stochastic dominance but it does allow violations of transitivity.

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* *A* according to the utility function *u*. Assuming *v* is more concave than *u*, we can not have *A* *C* and *C* *B* under *v*. This means the cycle implied by regret theory and Bordley's (1992) model is ruled out under CDG (See Section 6). Now suppose instead that *A* *B* according to utility function *u*. Again, *v* is more concave than *u* so we can have *B* *C* and *C* *A* which shows that the opposite intransitive cycle is admissible under CDG.

(ii) Context- and Reference-Dependent Utility (CRU)

CRU is based upon Sugden's (2003) model of subjective expected utility with state-dependent reference point. Reference points had been discussed by Markowitz (1952), Edwards (1954) and in Kahneman & Tversky (1979), in order 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

$$V(A) = \sum_{\iota=1}^{\nu} \pi_\iota \upsilon(\alpha_\iota, \rho_\iota)\,, \tag{2}$$

where *V(A)* represents the context- and reference- dependent utility of gamble *A*. The reference point in prospect theory was taken as 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.

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 choices. More formally, we have

$$A \succ B \Leftrightarrow \sum_{i=1}^{v} \pi_i v[\alpha_i, \rho_i(A, B)] > \sum_{i=1}^{v} \pi_i v[\beta_i, \rho_i(A, B)] \qquad (3)$$

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 & Schmidt (2005) and pursued in this paper is that when the initial endowment is zero, the reference point is the maximum of the lowest consequences of the two gambles in a choice. Assuming that utility depends on the difference between consequence and reference point, we have

Bleichrodt & Schmidt (2005) 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* *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*; therefore, *B* *C* $\Leftrightarrow (p + q)u(b - c) + (1 - p - q)u(-c) > 0$ and *C* *A* $\Leftrightarrow 0 > pu(a - c) + (1 - p)u(-c)$.

(iii) Numerical Predictions

In our first series, *A* = (\$100, 0.5; \$0), *B* = (\$50, 0.9; \$0), and *C* = \$37 for sure. The cycle *A* *B*, *B* *C*, and *C* *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 < β < 0.6, which are compatible with previously published results. For instance, in the standard common ratio effect (Kahneman & Tversky, 1979), it has been found that ($3000, 1) ($4000, 0.8; $0) and ($4000, 0.2; $0) ($3000, 0.25; $0); this result is implied by CDG when $\alpha >$ 0.78 and β < 0.78.

For CRU we approximate the utility function as follows:

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

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

3 Transitivity and Error Models

Testing transitivity with fallible data has been a controversial topic in psychology (Tversky, 1969; Iverson & Falmagne, 1985; Iverson, Myung & Karabatsos, 2005; Brandstaetter, et al., 2006; Regenwetter & Stober, 2006). 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 instead to noise or “error” (Humphrey, 2001; Loomes, Starmer, & Sugden, 1991; Sopher & Gigliotti, 1993; Starmer & Sugden, 1998; Starmer, 1999). Different models of error have been discussed by Birnbaum (2004b; 2006), Carbone and Hey (2000), Harless and Camerer (1994), Hey (2005), Hey and Orme (1994), Luce (1959; 1994), Sopher and Gigliotti (1993), Thurstone (1927), and others.

Because we plan to test transitivity, we think it best to use a choice model that is neutral with respect to transitivity. Models of Thurstone (1927), Luce, (1959), Busemeyer & Townsend (1993), Hey and Orme (1994) implicitly assume or imply transitivity in the absence of error. Our approach is described by Birnbaum (2006), and is similar to that of Harless & Camerer (1994) and Sopher & Gigliotti (1993). Whereas Harless & Camerer (1994) assumed that matched choices have the same error probability, Sopher & Gigliotti allowed the error rates in different choices to be unequal. Birnbaum's (2006) improvement over those papers is to use repeated presentations of the same choice in order to estimate error rates for different choices. The advantage of this approach is that the error terms are not constructed to conveniently "explain away" violations of transitivity, but are estimated by a method that is neutral with respect to the issue under investigation.

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 probability of switching from *A* to *B* is given as follows:

$$P(AB) = \pi(1-\varepsilon)\varepsilon + (1-\pi)(1-\varepsilon)\varepsilon = \varepsilon(1-\varepsilon) \qquad (5)$$

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 that 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, as listed in Table 2.

Insert Table 2 about here.

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

$$
\begin{aligned}
P(000) = {} & \pi_{000}(1-\varepsilon_1)(1-\varepsilon_2)(1-\varepsilon_3) + \pi_{001}(1-\varepsilon_1)(1-\varepsilon_2)\varepsilon_3 + \\
& \pi_{010}(1-\varepsilon_1)\varepsilon_2(1-\varepsilon_3) + \pi_{011}(1-\varepsilon_1)\varepsilon_2\varepsilon_3 + \\
& \pi_{100}\varepsilon_1(1-\varepsilon_2)(1-\varepsilon_3) + \pi_{101}\varepsilon_1(1-\varepsilon_2)\varepsilon_3 + \\
& \pi_{110}\varepsilon_1\varepsilon_2(1-\varepsilon_3) + \pi_{111}\varepsilon_1\varepsilon_2\varepsilon_3
\end{aligned} \tag{6}
$$

where *P(000)* is the probability of 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, and e_3 are the probabilities of making an "error" in expressing preference on the three respective choices, which are assumed to fall between zero and one-half. There are seven other equations like the above for the other seven observed patterns.

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

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., . We can test transitivity by comparing the fit of the transitive model to a more general model in which all parameters are free.

To illustrate how this model produces asymmetry of intransitive patterns, consider the following hypothetical case in which there are 100 identical participants, each of whom has the "true" preference pattern 001 (i.e., $A \succ B \succ C$), and $e_1 = \varepsilon_2 = \varepsilon_3 = 0.2$. If each participant made each choice once, this model implies that 12.8%, 51.2%, 3.2%, 12.8%, 3.2%, 12.8%, 0.8%, and 3.2% would exhibit response patterns 000, 001, 010, 011, 100, 101, 110, and 111, respectively. Note that there are four times as many intransitive patterns of type 000 than of type 111 (12.8% versus 3.2%). If the error terms were reduced to 0.1, the overall rate of intransitivity is reduced, but the asymmetry ratio is increased to 9:1. Intuitively, the reason this model produces asymmetry is that the 000 pattern requires only one "error", whereas the 111 pattern requires two "errors" from the "true" pattern of 001.

This example also illustrates the advantages of using replications. Of the response patterns in a single replication, the model implies that only 51.2% of the observed patterns are the "true" pattern. In a two-replication experiment, however, of those cases where a pattern is repeated, true patterns account for 83.4% of replicated patterns. Intuitively, because "errors" are less likely than correct reports, repetitions of a given response pattern can be treated as more likely "true" than single occurrences of that pattern.

We also analyze the data with respect to a still more general model in which each participant has a different "true" pattern of preference and each participant is also allowed different errors rates on different choices. Because each participant makes each choice four times (twice for each presentation order), we can summarize the data for the *i*th participant as (n_{i1}, n_{i2}, n_{i3}), where, n_{i1}, n_{i2}, and n_{i3} are the number of times (out of four) that this participant chose $A \succ B$, $B \succ C$, and $C \succ A$ on these three choices, respectively. We can express the likelihood of these data, given the hypothesis that the "true" pattern is 000 as follows:

$$L(n_{i1}, n_{i2}, n_{i3} \mid 000) = (1-\varepsilon_{\iota 1})^{4-\nu_{\iota 1}}(\varepsilon_{\iota 1})^{\nu_{\iota 1}}(1-\varepsilon_{\iota 2})^{4-\nu_{\iota 2}}(\varepsilon_{\iota 2})^{\nu_{\iota 2}}(1-\varepsilon_{\iota 3})^{4-\nu_{\iota 3}}(\varepsilon_{\iota 3})^{\nu_{\iota 3}} \qquad (7)$$

where e_{i1}, e_{i2}, and e_{i3} are the estimated error rates for person *i* on these choices, respectively. There are eight such expressions for eight hypotheses ($H_1 = 000$, $H_2 = 001, \ldots$, $H_8 = 111$), and each can be used to compute error rates to maximize the likelihood. These eight maximum likelihood calculations can be made for each participant and the hypothesis chosen that has the highest Bayesian posterior probability given the data. The Bayesian, posterior probability for hypothesis 000 is given as follows:

$$B_i(000 \mid n_{i1}, n_{i2}, n_{i3}) = \frac{\Pi(000)\Lambda(\nu_{\iota 1}, \nu_{\iota 2}, \nu_{\iota 3} \mid 000)}{\sum_{\kappa=1}^{8} \Pi(H_\kappa)\Lambda(\nu_{\iota 1}, \nu_{\iota 2}, \nu_{\iota 3} \mid H_\kappa)} \qquad (8)$$

where $P(H_k)$ are the prior probabilities of the eight hypotheses, and $B_i(H_k \mid n_{i1}, n_{i2}, n_{i3})$ is the posterior probabilities of the *k*th hypothesis for the *i*th participant, given his or her observed data. The prior probabilities for eight hypotheses can be initially set to be uniform, and can then "bootstrapped" from the sum of maximum likelihood for each hypothesis, summed over participants.

Equation 7 shows that even if each participant has a different error rate for each choice, response patterns that are repeated are more likely "true" than sequences that are not. The model of Equation 7 might be criticized (justly, we think, given the fit of the simpler model) as "overfitting," since it uses so many parameters. However, we use this general model here for the purpose of investigating a specific hypothesis concerning its error rates. In particular, we examine if error rates are smaller for choices with greater differences when the "true" order has been inferred for each person. This allows examination of a particular hypothesis about deviations from the simpler model of Equation 6.

Error rates for more distant comparisons should be smaller according to transitive models such as those of Thurstone (1927) and Luce (1959), which can be deduced from the assumption that people choose the alternative with the higher momentary utility, where the utilities of the gambles contain random error. Suppose the momentary utility of gamble *G* is given by

$$U(G) = \upsilon(\Gamma) + \varepsilon_{\Gamma} \tag{9}$$

where *u(G)* is the true utility of *G* and ε_G is a normally distributed "error" component with a mean of zero. Suppose a person chooses the gamble with the higher momentary utility. If so, then the probability of choosing *A* over *B* is the probability that $U(A) > Y(B)$. If the errors are normally distributed, it follows that,

$$P(A,B) = N\left[\frac{\upsilon(A) - \upsilon(B)}{\sigma_{AB}}\right] \tag{10}$$

where σ_{AB} is the standard deviation of the difference ε_A - e_B, and *N* is the cumulative normal distribution function. According to this model, if the standard deviations of differences are equal,

$\sigma_{AB} = \sigma_{BC} = \sigma_{AC}$, then if $P(A,B) = 0.84$ and $P(B,C) = 0.84$, then $P(A,C) = 0.975$. Equation 10 is sometimes called "Thurstone's law" of comparative judgment.

Interpreted with respect to Equations 6 or 7, we have a "true" pattern of $A \succ B \succ C$ where $e_1 = 0.16$, $e_2 = 0.16$, and $e_3 = 0.025$. That is, the rate of "error" on the *AC* choice should be smaller than and predictable from the rates of error on *AB* and *BC*, because the utility difference is the sum of the two differences, $u(A) - u(X) = u(A) - u(B) + [u(B) - u(X)]$. This shows that the random noise error model (Carbone & Hey, 2000: Hey, 2005; Hey & Orme, 1994) can be treated as a (transitive) special case of the most general "true and error" model used here.

4 Methods of Studies 1-3

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 3 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 used in the first two studies. The difference between Trials 5 and 20 Trials 8 and 17, and Trials 11 and 14 is just the (first or second) positioning 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 *b* = \$53, and *c* = \$33 in Series II; and *b* = \$55 and *c* = \$30 in Series III.

Insert Table 3 about here.

Three studies used different groups of participants. Study 1 used 127 college undergraduates who performed the 20 choices twice, separated by four other intervening tasks that required about 20 minutes. These college students were tested in labs containing Internet-connected computers. They participated as one option toward an assignment in lower division psychology. Of these, 54% were female; 87% were 20 years or younger, and no one was older than 26.

In Study 2, a second group of 162 participants was recruited via the Web; these participated with the understanding that one would be chosen to receive the prize of one of their 20 chosen gambles. Of these 162, 66% were female; 35% were 20 years or less, and 15% were over 40.

Complete materials for Studies 1 and 2 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

For Study 3, a third group of 149 participants was recruited via the Web and tested using different choices for Series II and III. Series I was the same, but Series II used *b* = \$45, and *c* = \$40; Series III used *b* = \$43 and *c* = \$38. This widened our search for intransitive preferences predicted by the CRU or CDG models. Of these 149, 68% were female, 20% were 20 years of age or less, and 19% were 40 years or older. Materials for Study 3 are available at the following URL:

http://psych.fullerton.edu/mbirnbaum/decisions/Schmidt_Ulli2.htm

5 Results of Studies 1-3

The percentages of people who chose the second gamble in each pair are shown in Table 3 for all three series. If most people followed the predictions of CRU or CDG 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, .9; $0) over ($100, .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 is consistent with the model of Bordley (1992). Only one person repeated the predicted pattern 000 of intransitivity in the first replicate and no one repeated this same pattern on both replicates of Series I. One person repeated the opposite intransitive pattern, 111, in both replicates of Series I.

Insert Table 4 about here.

Tables 5 and 6 show response patterns as in Table 4 for Series II and III, respectively. Data from the Web recruits are shown in Table 7 for all 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).

Insert Tables 5, 6, and 7 about here.

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).

Insert Table 8 about here.

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

Insert Table 9 about here.

The Chi-Squares (indices of lack of fit) for unconstrained and transitive models are shown in Table 10. None of these values are significant. The difference in fit between the transitive and free (allowing intransitivity) models is not significant in any of the 9 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, III of Lab or Web data.

Insert Table 10 about here.

Tables 11 and 12 show results for Study 3, which can be compared with Tables 3 and 7, respectively. Study 3 used the same choices for Series I, where the modal preference pattern was 100 in all three studies. The modal choice pattern in Series II and III of Studies 1 and 2 was 101 in all four cases. Study 3 used different choices in Series II and III, which changed the modal choice pattern to 110; however, the rate of intransitivity remained quite low: only 1 person repeated the predicted 000 pattern in Series II and III combined.

Insert Tables 11 and 12 about here.

The error terms (e_1, e_2, e_3) can be estimated directly from the number 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 repetitions in Study 1, the average rate of agreement over all 20 choices was 85.6%. Within replicates, the agreement between the same choices with the first-second position reversed was 88.5%. The correlation between these two estimates of “error” over participants was

0.63, indicating individual differences in the error rate; some people have greater error rates than others. With 87% agreement, there are 13% preference reversals [$2e(1-\varepsilon)$], which corresponds to an average error rate, e = 0.07.

For Replicate 1 of Series I, there were 24 who reversed preferences between Choice 5 and 20, 25 who reversed preferences between Choices 8 and 17, and only 9 who reversed preferences between Choices 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 error rates fixed to these values, χ^2 = 7.09 for the model with probabilities of all sequences free, and χ^2 = 8.71 for the transitive model. The difference, 1.6, is not significant. 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 these 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, while adding 0, 1, 2, 3, 4, or 5 hypothetical cases who repeated the predicted 000 pattern. The χ^2 values are 8.7, 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 3% 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.

With the same data, we can also examine fit as a function of the proportion of assumed intransitive participants. Recall that the best-fit solution indicated that both types of intransitivity represent 1.4% of the participants, with χ^2 = 7.09. Assuming that this theoretical probability of intransitivity [$p_{000} + \pi_{111}$] is 0.05, 0.08, 0.10, 0.15, 0.20, or 0.25, the corresponding χ^2 are 8.9, 11.4, 13.2, 18.3, 24.2, and 31.1. Thus, this analysis shows that the data not only fail to show much evidence for intransitivity, they also allow us to reject the hypothesis that as much as 10% was intransitive.

The more general error model of Expressions 7 and 8 was applied to the data of Study 1 Series I. The maximum likelihood solution for each hypothesis was calculated for each participant, and these were used to boostrap the most probable Bayesian solution for each person. It was found that hypotheses 000, **001**, 010, 011, **100**, **101**, **110**, and 111 were most probably correct for 0, **33**, 6, 3, **45**, **21**, **18**, and 1 participants, respectively. The four most probable hypotheses coincide with those four patterns most often repeated as well as those four that were most probable in the group analysis. Only 1 person was deemed probably intransitive by this analysis. For the remaining 126 participants in Study 1, we can renumber the gambles for each person so that we can estimate error rates for the choice between the best two gambles, the worst two, and the choice between the best and worst for each person, respectively. With this rearrangement, the average error rates are 0.14, 0.11, and 0.04, respectively. Thus, average individual error rates are consistent with the hypothesis that error rates are smaller for gambles that are more distant, as predicted by the transitive models of Thurstone and Luce (Thurstone, 1927; Luce, 1959, 1994).

This analysis was applied to the 306 Web data of Series I of Studies 2 and 3 who completed all choices in Series I. In this case, there were 0, **67**, 14, 7, **96**, **49**, **63**, and 10 participants had highest Bayesian posterior probabilities for 000, **001**, 010, 011, **100**, **101**, **110** and 111, respectively. Rearranging the data for the 296 who appeared transitive, the three average error rates were 0.12, 0.04, and 0.02 for the errors choosing between the two best, worst two, and between best and worst. Again, the average error rate is least for the choice that is most discrepant. Despite the fact that the simpler error model provides an adequate fit to the data, these analyses suggest that deviations from that model are systematic and consistent with the class of Thurstone models.

However, whichever of these error models is assumed, the data provide virtually no evidence that people are systematically intransitive for the choices we tested. We found no significant violations of transitivity of the type predicted by CRU and CDG, nor did we find substantial violations of the opposite type (Bordley, 1992). Satisfaction of transitivity does not rule out these models, of course, because these models could allow transitive preferences for this experiment with different parameters. In other words, our results do not rule out intransitivity for choices not yet tested. Nevertheless, one author of this paper was also an author of the CRU and CDG models, and we did our best to find violations, if they were to be found.

The CRU and CDG models had no historical record of claims of violations of transitivity. However, the additive difference models, including regret theory, have been at the center of a controversy concerning intransitivity. In Studies 4 and 5, we apply the same experimental and analytical methods to the study of intransitivity predicted by regret theory, using choices that had been previously tested.

6 Additive Difference Models: Regret and Majority Rule

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

$$A \succ B \Leftrightarrow \sum_{\iota=1}^{v} \phi_{\iota}[\alpha_{\iota}, \beta_{\iota}] > 0 \qquad (9)$$

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 *i*, and the function ϕ_i maps a contrast on this state of the world (or dimension) into a preference between the gambles.

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 the regret. For example, suppose that regrets are given by the following expression:

$$\phi_i[a_i, b_i] = p_i(x_i - y_i)^3 \qquad (10)$$

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 particularly large regrets, 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 13. These were used in by Loomes, Starmer, and Sugden (1991) and produced the greatest percentage of predicted 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 cycle that differed from it by only one choice. According to Equations 9 and 10, $\sum \phi_i[b_i, c_i] = –8.3$, so $C \succ B$; $\sum \phi_i[a_i, b_i] = –18.7$, so $B \succ A$; however, $\sum \phi_i[a_i, c_i] = 45.2$, so $A \succ C$, violating transitivity.

Insert Table 13 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:

$$\phi_i[a_i, b_i] = \begin{cases} p_i, a_i \succ b_i \\ -p_i, a_i \prec b_i \end{cases} \qquad (11)$$

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, $\sum \phi_i[a_i, b_i] = 0.4$, $\sum \phi_i[b_i, c_i] = 0.4$, yet $\sum \phi_i[a_i, c_i] = -0.2$; therefore, $A \succ B$, $B \succ C$, but $C \succ A$.]

A problem in previous empirical tests of regret theory is that a number of confounds were present in those studies (Humphrey, 2001; Starmer & Sugden, 1993; 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, $A = (\$10, 0.4; \$3, 0.3; \$3, 0.3)$, $B = (\$7.5, 0.4; \$7.5, 0.3; \$1, 0.3)$). However, the so-called choice between *B* and *C* was presented in a form in which the two upper branches of B' and C' were coalesced ($B' = (\$7.5, 0.7; \$1, 0.3)$, $C' = (\$5, 0.7; \$5, 0.3)$). The so-called choice between $C = (\$5, 0.4; \$5, 0.3; \$5, 0.3)$ and *A* was presented with the two lower branches of C'' and A'' coalesced ($C'' = (\$5, 0.4; \$5, 0.6)$, $A'' = (\$10, 0.4; \$6, 0.3)$). According to the transitive, TAX model (e.g., Birnbaum & Navarrete, 1998), with parameters estimated from previous data, the splitting and coalescing of branches could account for apparent violations of transitivity. According to TAX, $U(A)$ = 4.33, $U(B)$ = 5.33; $U(C)$ = 5.00; $U(B')$ = 3.79; $U(X')$ = 5.00, $U(A'')$ = 5.01; $U(X''')$ = 5.00. Thus, this TAX model implies that $A \prec B$, $B' \prec C'$, and $A'' \succ C''$.

In this paper, we keep all gambles in the same three-branch form to remove this confound with event splitting. Starmer & Sugden (1998) and Humphrey (2001) recognized this confound and controlled for it by presenting the choices in fully split forms or by using a different format for display ("strip") in which gambles could be 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 noted in Section 3, such asymmetry is entirely compatible with an error model in which people make occasional "errors" in reporting their preferences, even if everyone is truly transitive.

7 Methods of Studies 4 and 5

Study 4 was conducted with 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 the matrix format shown in Figure 1, which illustrates the payoffs for states of the world in three of the choices used. Participants viewed 15 choices via the Internet and indicated their choices by clicking 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 13). The difference between Choices 11 and 13, for example, is first or second positioning 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) Insert Table 13 about here.

The 314 college undergraduates 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, and 1% were older than 26.

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

Study 5 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 4, 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 4 (Table 13), 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). Series II were previously used by Starmer & Sugden (1998), leading to relatively high rates of asymmetry.

8 Results of Studies 4 and 5

The percentages choosing the second gamble in each choice are displayed in Table 13. The observed frequencies for response patterns of Series I of Studies 4 and 5 are presented in Table 14, with results from Starmer and Sugden (1998) included for comparison. However, whereas Starmer & Sugden (1998) observed 20% showing the intransitive pattern 111, we found in Study 1 only 8% and 7%, who showed this pattern on Choices #11, 5, and 13, on # 9, 15, and 7, respectively; only 0.6% showed this pattern on both repetitions.

In Study 5, 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 Table 15 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). Insert Tables 14 and 15 about here.

Tables 16 and 17 show how error rates in each choice are estimated from preference reversals between repetitions of the same choices. The three rows in Table 16 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, 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 16 and 17 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 4, these were estimated to be 0.13, 0.16, and 0.14 for the three choices (*AB*, *BC*, *CA*), respectively. In Study 5, 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 14).

Recall that these estimates of error rates assume nothing about transitivity.

Tables 18, 19, and 20 show the fit of the "true and error" model to the observed frequencies, using error rates estimated from replications. This model was fit with the probabilities of intransitive patterns freely estimated or with these parameters set to zero. The purely transitive model gave a good approximation to the data of Study 4, and deviations of fit are not significant, $\chi^2(12) = 14.3$. When all parameters were free, the fit was only slightly improved, and the estimated rate of intransitivity of both types was 3%. This improvement in fit was not significant, $\chi^2(2) = 2.85$. Therefore, we can retain the hypothesis that everyone was transitive in Study 4.

Insert Tables 18, 19, and 20 about here.

In Study 5, the fit of the true and error model to the data in Table 19 yielded $\chi^2(10) = 4.53$. However, the fit of the purely transitive model was worse, $\chi^2(12) = 23.72$, which suggests that the 7% estimated incidence of intransitivity can be considered significantly greater than zero. Fitting the true and error model to the second set, in Table 20, the value of $\chi^2(12) = 2.19$, again indicating acceptable fit. With the probabilities of both intransitive patterns set to zero, $\chi^2(10) = 36.0$, which is again significant. This analysis indicates that the estimated rate of 2% intransitivity is "significant," relative to the small error rates in Table 17. With the 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 memory been required to reproduce identical choices.

9 Summary and Conclusions

The success of transitivity in our data is compatible with findings of Birnbaum and Gutierrez (2006), who tested violations of transitivity predicted by a lexicographic semi-order and 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 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, TAX, and other transitive models. Interestingly, this consensus among people occurred despite the fact that Tversky's gambles were designed to have nearly identical expected values.

These data show greater individual differences among people for their transitive orders than found by Birnbaum and Gutierrez. There are four popular transitive patterns in our Studies 1-3: 001, 100, 101, and 110. In those studies there is a sure thing, a highly probable prize and a medium probability prize. The Tversky gambles, in contrast, included no sure thing; perhaps these individual differences in Studies 1-3 arise because of this feature of our study. Although people disagreed with each other, 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 82% of the time between replicates. Perhaps our higher internal consistency was facilitated by many repetitions of the same or similar choices Study 1-2.

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, 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; in fact, neither our German nor American samples showed the

[1] Although the model of Brandstaetter, et al. (in press) 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, in all three series, whereas the observed modal choices in Series II and III are . In Study 4, this model predicts C f A f B, in agreement with the most frequently repeated pattern. In Study 5 Series II, it predicts C f A f B, which was repeated by only 1 person; instead, the modal pattern was C f B f A.

asymmetry previously reported.

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). 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. We found very few people who repeated the pattern predicted by that model in Studies 4 and 5.

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; 2004a; 2004b; 2005a; 2005b). 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, in order to account for intransitive preference.

Similarly, had either 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 the data do not confirm these intuitions: we found few cases where participants repeated an intransitive pattern. Combining these data with those of Birnbaum and Gutierrez for the lexicographic semiorder, we think the burden of proof should shift to those who would argue that intransitive models are descriptive of more than two or three percent of the population.

In summary, we tested for violations of transitivity where predicted by four 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 due to "errors," we find little evidence to refute the hypothesis that everyone had a transitive preference order.

References

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. (submitted). Testing for intransitivity of preferences predicted by lexicographic semiorder. http://psych.fullerton.edu/mbirnbaum/birnbaum.htm, Web document.

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.

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. (in press). Axiomatization of a preference for most probable winner. Theory and Decision, in press.

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.

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. (1991). Observing violations of transitivity by experimental methods. Econometrica, 59, 425-440.

Loomes, G., Starmer, C., & Sugden, R. (1989). Preference reversal: information-processing effect or rational non-transitive choice? Economic Journal, 99, 140-151.

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. Design of Lotteries used to test transitivity in Studies 1-3.

Lottery	p	q	1 – p – q
A	*a*	*0*	*0*
B	*b*	*b*	*0*
C	*c*	*c*	*c*

Note: $a > \beta > \gamma > 0$.

Table 2. Patterns of Choice.

Notation	Preference Pattern	Preference Order
000	A f B; B f C; C f A	Intransitive
001	A f B; B f C; C p A	A f B f C
010	A f B; B p C; C f A	C f A f B
011	A f B; B p C; C p A	A f C f B
100	A p B; B f C; C f A	B f C f A
101	A p B; B f C; C p A	B f C f A
110	A p B; B p C; C f A	C f B f A
111	A p B; B p C; C p A	Intransitive

The pattern predicted by CRU and CDU is 000.

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

		Choice		% Second Gamble		
Code	Trial	First 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. Trial numbers apply to Series I only. Predicted pattern of intransitivity is A f B f C f A for CRU and CDU.

Table 4. Response Patterns for Series I.

	First Rep			Second Rep			First and Second
Pattern	#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 1 person showed the predicted pattern of intransitivity on both Choices 5, 8, and 11, and 20, 17, and 14 in the First replicate. However, 25 repeated the transitive pattern, $B \succ C \succ A$.

Table 5. Response patterns for Series II. Each entry is the number of participants who showed each choice combination.

	First Rep			Second Rep			First and Second
Pattern	#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

Table 6. Response patterns in Series III.

	First Rep			Second Rep			First and Second
Pattern	#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 of Study 2.

	Series I			Series II			Series III		
Pattern	#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).

	First Rep		Second Rep	
Pattern	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 df in the data.

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

	First Rep #5,8,11 and #20, 17, 14		Second Rep # 5, 8, 11 and 20, 17, 14.	
Pattern	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 .09, .08, and .10 in the first repetition, and .10, .08, and 0 in the second repetition. Although one or two people may have been systematically intransitive in Series I, a good fit is still obtained when we assume there were no intransitive participants.

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

Data Set	Model (df)		
	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 *df* are 14.1, 11.1, and 6.0 for $\alpha = 0.05$, respectively. None of the values are significant, indicating that the transitive model provides a satisfactory description of the data.

Table 11. Tests of transitivity in Study 3 (n = 149). Series II and III used different values from those in Study1.

		Choice		% Second Gamble		
Code	Trial	First Gamble	Second Gamble	Series I b = 50 c = 37	Series II b = 45 c = 40	Series III b = 43 c = 38
AB	5	*A*: 50 to win $100 50 to win $0	*B*: 90 to win $*b* 10 to win $0	70	62	57
BC	8	*B*: 90 to win $*b* 10 to win $0	*C*: $*c* for sure	37	65	61
CA	11	C: $c for sure	*A*: 50 to win $100 50 to win $0	46	32	42
BA	20	*B*: 90 to win $*b* 10 to win $0	*A*: 50 to win $100 50 to win $0	33	41	44
CB	17	*C*: $*c* for sure	*B*: 90 to win $*b* 10 to win $0	60	34	38
AC	14	*A*: 50 to win $100 50 to win $0	*C*: $*c* for sure	54	66	60

Note that Trials 5, 8, and 11 are the same as 20, 17, and 14, respectively, but counterbalanced for position. Trial numbers apply to Series I only.

Table 12. Response patterns in Web participants of Study 3.

	Series I			Series II			Series III		
Pattern	#5, 8, 11	#20, 17, 14	Both	#10, 4, 7	#19, 15,13	Both	#6, 3, 12	#18, 9,15	Both
000	2	2	0	3	3	1	6	1	0
001	26	26	20	18	18	13	22	23	15
010	14	11	8	16	18	12	14	17	9
011	3	9	3	19	22	14	21	24	14
100	32	37	26	25	26	15	23	24	14
101	31	22	16	5	3	1	7	8	4
110	32	31	23	57	51	45	42	45	31
111	6	8	3	5	7	1	12	5	2
Total	146	146	99	148	148	102	147	147	89

Notes: Each choice was repeated only with position counterbalanced. Totals do not sum to number of participants (n = 149) due to occasional skipped items.

Table 13. Tests of Transitivity used in Studies 4 and 5 (Series I).

		Choice		N = 314	N = 103	
Type	No.	*first Gamble*	*second Gamble*	Study 4 Series I	Study 5 Series I	Study 5 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

Enties show the percentage choosing the second gamble in each choice. 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 14. Response Patterns to choices of Series I in both studies.

Pattern	Starmer & Sugden (1998)	Study 4, *n* = 314			Study 5, *n* = 103		
		#11, 5, 13	#9, 15, 7	Both	#11, 5, 13	#9, 15, 7	Both
000	6	20	27	4	7	6	2
001	7	45	48	27	22	20	15
010	15	90	89	55	16	16	11
011	16	53	43	12	21	19	11
100	8	23	27	10	11	11	5
101	7	25	28	10	9	13	5
110	14	32	29	9	11	12	5
111	17	25	22	2	6	6	3
Total	90	313	313	129	103	103	57

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

Table 15. Response Patterns in the Study 5 Series II.

Pattern	Starmer & Sugden (1998)	#6, 10, 14	#4, 12, 8	Both	
000	3	2	0	0	
001	5	5	6	5	
010	4	2	3	1	
011	2	1	1	1	
100	9	37	37	31	
101	12	12	12	9	
110	45	41	41	37	
111	10	3	3	2	
Total	90	103	103	86	

Responses to #4, 12, and 8 have been reflected for comparability.

Table 16. Preference Reversals between Repetitions (Study 4)

	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 a skipped item.

Table 17. Preference Reversals between Repetitions (Study 5)

	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 18. Fit of Purely Transitive Model to Observed Frequencies (Study 4).

				Observed Data		Predicted		Est.
Pattern	#11, 5, 13	#9, 15, 7	Both	Both	OR-Both	Both	OR-Both	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 $\chi^2(10) = 14.3$, which is not significant. Allowing intransitivity, the estimated proportion of intransitive participants was .03; the fit was slightly improved, $\chi^2(12) = 11.4$, a nonsignificant improvement, $\chi^2(2) = 2.85$.

Table 19. Fit of True and Error Model to Observed Frequencies (Study 5, first series)

	N = 103			Observed Data		Predicted		Est.
Pattern	#6, 10, 14	#4, 12, 8	Both	Both	OR-Both	Both	OR-Both	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	

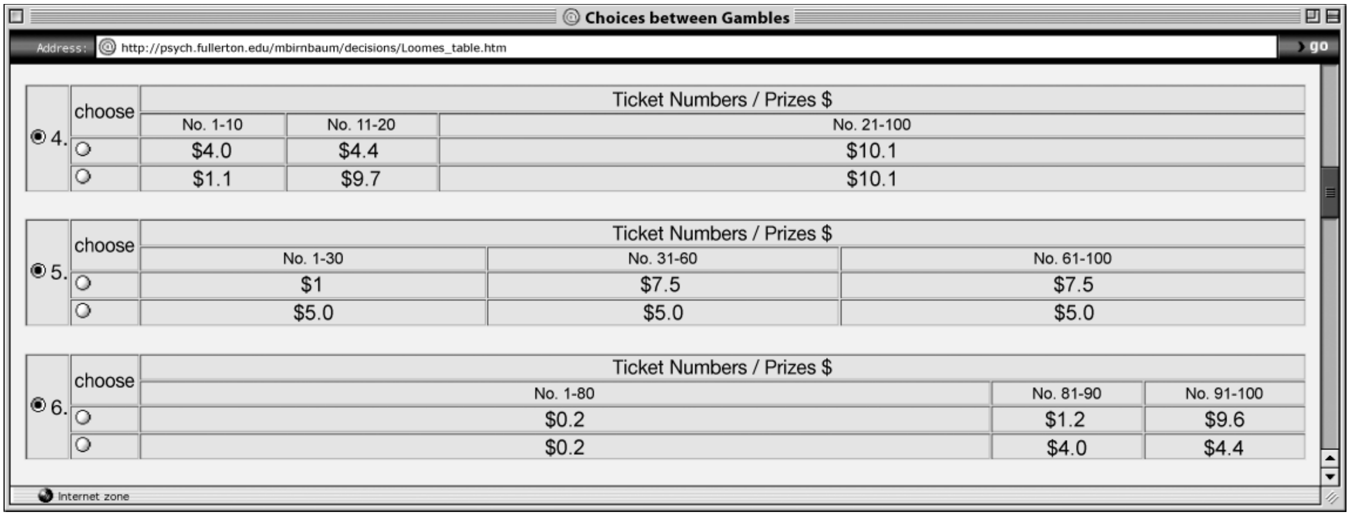
Index of fit is $\chi^2(10) = 4.53$. The estimated "true" rate of intransitivity is 7%.

Table 20. Fit of True and Error Model to Observed Frequencies (Study 5, series II)

	Data as of 5-24-06 n = 103			Observed Data		Predicted		Est.
Pattern	#11, 5, 13	#9, 15, 7	Both	Both	OR-Both	Both	OR-Both	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	

Index of fit is $\chi^2(10) = 2.19$ with estimated "true" intransitivites of 2%.

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.



	choose	Ticket Numbers / Prizes $		
4.		No. 1-10	No. 11-20	No. 21-100
	○	$4.0	$4.4	$10.1
	○	$1.1	$9.7	$10.1

	choose	Ticket Numbers / Prizes $		
5.		No. 1-30	No. 31-60	No. 61-100
	○	$1	$7.5	$7.5
	○	$5.0	$5.0	$5.0

	choose	Ticket Numbers / Prizes $		
6.		No. 1-80	No. 81-90	No. 91-100
	○	$0.2	$1.2	$9.6
	○	$0.2	$4.0	$4.4