Working Paper, December 30, 2020, pp. 1–13

Testing Transitivity of Preference in Individuals

Michael H. Birnbaum*

Abstract

This study presents a new experiment testing transitivity of preferences in individuals using the stimulus design of Butler and Pogrebna (2018). Each individual responded to each choice problem 60 times, replicated twice in each of 30 sessions. The individual true and error model was used to estimate the incidence of transitive and intransitive preference patterns and error rates for the choice problems for each person. Although the data of most participants were consistent with transitivity, a few participants showed convincing evidence of intransitive preferences patterns at least part of the time, and several participants showed clear evidence of changing true preferences over time. This study also tested and found violations of the assumption that choice responses are independently and identically distributed over repetitions, an assumption used previously in certain random utility models and statistical analyses. Violations of iid are compatible with Markov True and Error Models in which parameters drift gradually over time.

Keywords: Error Theory, Risky Decision Making, Transitivity, Regret theory

1 Introduction

If preferences are *transitive*, then if X ≻ Y and Y ≻ Z, then X ≻ Z, where ≻ denotes "is truly preferred to". When a formal property like transitivity is tested empirically, however, it might be that individual responses violate the property because responses might contain random error. Further, different people might have different true preferences, and the same person might change true preferences from session to session. So there has been an issue of how to decide whether observed violations might be due to random error, to changing preferences, or instead are "real."

When devising a test of transitivity, researchers begin with a rival model that is not transitive and choose X, Y, and Z such that this rival model implies an intransitive cycle of preferences. A number of papers explored violations of transitivity predicted by a lexicographic semiorder model (Tversky, 1969; Budescu & Weiss, 1987; Birnbaum, 2010; Birnbaum & Gutierrez, 2007; Birnbaum & Bahra, 2012; Cavagnaro & Davis-Stober, 2014; Raynard, Montgomery, Konstantinidis, & Taylor, 2020; Regenwetter, Dana, & Davis-Stober, 2011).

Regret theory (Loomes & Sugden, 1982) is a model that can violate transitivity, and a separate branch of literature developed searching for violations of transitivity implied by regret theory(Birnbaum & Schmidt, 2008), the rival similarity theory (Leland, 1998), or by related integrative contrast models (Birnbaum & Diecidue, 2015; González-Vallejo, 2002).

Editing mechanisms and contextual assimilation or contrast effects on components might also produce intransitive preferences (Birnbaum & Gutierrez, 2007; Birnbaum, Navarro-Martinez, Ungemach, Stewart, & Quispe-Torreblanca, 2016; Müller-Trede, Sher, & McKenzie, 2015).

Some reviews concluded that violations of transitivity of preference reported in this literature are not that impressive and might be due to error (e.g., Luce, 2000; Rieskamp, Busemeyer, & Mellers, 2006; Cavagnaro & Davis-Stober, 2014).

However, Butler and Pogrebna (2018) devised a set of gambles based on an intransitive, most probable winner (MPW) theory (Butler & Blavatskyy, 2019) that appeared to produce systematic violations of transitivity. Their design used 11 sets of three gambles ("triples"), each of which provided exactly three equally likely cash prizes. For example: X = (15, 15, 3), Y = (10, 10, 10), and Z = (27, 5, 5), where X = (15, 15, 3) represents a gamble with two equal chances to win 15 pounds and one chance out of three to win 3 pounds.

If the gambles are independent, the probability that X gives a higher prize than Y is 2/3; the probability that Y gives a higher outcome than Z is 2/3; and the probability that Z gives a higher prize than X is 5/9. So, if a person chose the MPW (the gamble most likely to give a higher outcome), her or his choices would be intransitive with this triple.

A reanalysis of their data using true and error theory found that there was modest, but convincing evidence of systematic violations of transitivity (Birnbaum, 2020), although their data showed more violations of the opposite type from those predicted by MPW theory and contained other results that allow one to reject the MPW theory as descriptive. It

I thank Rebecka Hahnel for assistance and Bonny Quan for comments on an earlier draft.

*Department of Psychology, California State University, Fullerton. Email: mbirnbaum@fullerton.edu.

was estimated that about 18% of the sample had intransitive preferences of the opposite type from MPW, but consistent instead with regret theory (Birnbaum, 2020).

One purpose of this research is to test whether consistent violations of transitivity can be observed using these stimuli devised by by Butler and Pogrebna (2018), while improving upon the methods and analyses.

1.1 Methodological issues in Transitivity

In an attempt to deal with the issue that individual responses might contain error, some researchers re-defined transitivity in terms of binary choice probabilities. But this approach is vulnerable to wrong conclusions. For example, Weak Stochastic Transitivity (WST) is defined as $P(XY) > 1/2$ and $P(YZ) > 1/2 \implies P(ZX) < 1/2$, where $P(XY)$ is the probability that A is preferred to B. However, if one third of the sample has the preference order $X \succ Y \succ Z$, one third has the preference order $Y \succ Z \succ X$ and one third has the order $Z \succ X \succ Y$, then WST is violated even though all preference patterns were perfectly transitive.

Similarly, the Triangle Inequality (TI) is defined as:
$1 \leq P(XY) + P(YZ) + P(ZX) \leq 2$.
However, it is possible that both WST and TI can be satisfied even when most of the preference patterns are intransitive. For example, suppose one-third have the intransitive preferences, $X \succ Y, Y \succ Z$, and $Z \succ X$; one-third have the intransitive pattern, $Y \succ X$, $Z \succ Y$, and $X \succ Z$, and one-third have the transitive pattern, $X \succ Y, Y \succ Z$, and $X \succ Z$. In this case both TI and WST are satisfied and yet two-thirds of the preference patterns are intransitive. See Birnbaum and Gutierrez (2007) and Birnbaum and Wan (2020) for other examples in which analyses based on binary choice proportions are unable to distinguish transitive from intransitive datasets.

Because analyses based on binary choices might be misleading, some investigators examined response patterns; the frequency of one type of intransitive response pattern was compared with the opposite intransitive pattern, and if one type was significantly more frequent than the opposite, it was taken as evidence of intransitive preferences. However, such asymmetry can easily occur as a result of error if error rates of different choice problems are not equal (Sopher & Gigliotti, 1993). Therefore, inequality of response patterns is also not a diagnostic test of transitivity (Birnbaum & Schmidt, 2008). In order to properly address the substantive question of transitivity, one must have a method for estimating error that does not assume transitivity.

1.2 True and Error (TE) Model

If one obtains replications of the same choice problems within person and within session, one can estimate error rates (Birnbaum, 2004, Appendix). The key assumption in true and error models is that within a brief session, reversals of preference by the same person to the same choice problem are due to random error. If there are three choice problems that are presented twice in each session, there are 64 possible response patterns ($2^6 = 64$); these provide the degrees of freedom to estimate error rates and the mixture of true preference patterns. The 3 error rates, e_1, e_2, and e_3, represent the probabilities that the participant's response in the three choice problems (XY, YZ, and ZX) would not match the person's true preferences. The probabilities of the 8 possible true preference patterns, p_{111}, p_{112}, p_{121}, p_{122}, p_{211}, p_{212}, p_{221}, and p_{222} represent the relative frequencies of the true preference patterns in a mixture.

According to the TE model used here, the "expected" (i.e., "fitted" or "predicted") frequency that a person would show the response pattern 111, for example, on both replications of three choice problems (denoted 111,111) is given as follows:

$$
\begin{aligned}
E_{111,111} = n[&p_{111}(1-e_1)^2(1-e_2)^2(e_3)^2 \\
&+p_{112}(1-e_1)^2(1-e_2)^2(1-e_3)^2 \\
&+p_{121}(1-e_1)^2(e_2)^2(e_3)^2 \\
&+p_{122}(1-e_1)^2(e_2)^2(1-e_3)^2 \\
&+p_{211}(e_1)^2(1-e_2)^2(e_3)^2 \\
&+p_{212}(e_1)^2(1-e_2)^2(1-e_3)^2 \\
&+p_{221}(e_1)^2(e_2)^2(e_3)^2 \\
&+p_{222}(e_1)^2(e_2)^2(1-e_3)^2]
\end{aligned}
$$

where $E_{111,111}$ is the "expected" frequency that a person shows the 111 response pattern both times. Note that if a person has the true preference pattern of 111, then she or he would have to push the right buttons on randomly ordered trials in order to make no errors on six choice problems to exhibit this response pattern. If the true pattern were 112, then he or she made an error on the third choice problem twice. There are 64 equations (including this one) for the predicted frequencies of the 64 possible response patterns for the six responses. Each "expected" frequency is simply n times the theoretical probability, calculated using parameter estimates best-fit to the data, where n is the number of sessions.

To fit the model to the 64 observed frequencies of response patterns to three choice problems replicated in each session, one can use a computer program that minimizes the index G (sometimes denoted G^2) is defined as follows:

$$G = 2 \sum \sum O_{ij} \ln (O_{ij}/E_{ij}) \quad (1)$$

where the summation is over the 64 cells, O_{ij} is the observed frequency (count) in the cell, E_{ij} is the "expected" frequency. The indices, i and j, represent the 8 response patterns for the first and second replications, respectively; i.e., i = 1, 2, 3, ..., 8 correspond to 111, 112, 121, ..., 222, respectively;

i.e., E_{11} corresponds to $E_{111,111}$. Minimizing G is equivalent to a maximum likelihood solution.[1]

Transitivity is the assumption that no one has true preferences that are intransitive; i.e., $p_{111} = p_{222} = 0$. It is a special case of the TE model with all parameters free. The difference in G between the general model and the transitive special case is theoretically Chi-Square distributed with 2 df.

The TE models can thus be viewed as a quantitative data analytic device, like Analysis of Variance, factor analysis, or Signal Detection Theory, and like those models, a TE model is also a testable descriptive model.

1.3 Group versus Individual Studies

In some studies, group data are obtained when a number of participants is presented with choice problems, and each person responds to each choice problem twice. The relevant theory in this case is known as *group* True and Error Theory (*g*TET), and it allows that each person might have a different set of true preferences (which may or may not be transitive), and that preference reversals by the same person to the same choice problems within the same session are due to error.

In the other case, each individual serves in many sessions, and within each session, the person responds at least twice to each choice problem. Data for each individual in this case are analyzed separately via *individual* True and Error Theory (*i*TET), which allows that an individual may change true preferences from session to session, but it makes the same assumption that preference reversals within the same session are due to random error.

In either form of TE, when a person is asked to respond to a choice problem, the person might make an "error", which might be caused by any of a number of factors such as misreading the problem, failing to remember or properly aggregate the information, failing to remember the decision, or failing to push the appropriate response button.

The study by Butler and Pogrebna (2018) was a *group* study in which 100 individuals judged each choice problem twice.

When a small percentage of data in the analysis of a group of participants (as in the gTET analysis of Butler and Pogrebna, 2018) show a particular phenomenon (in this case, violate transitivity), there are two possibilities: perhaps each person might exhibit the property some fraction of the time, or perhaps a few people show the effect consistently.

A main purpose of the present study, therefore, is to address this question by performing a study in which individual TE model can be applied. This study obtains enough data from each participant so that we can test the issue of transitivity in each individual and estimate a mixture model in which participants might have different preference patterns in different parts of the study.

[1]Two computer programs that can perform these analyses are freely available via the Online supplement to Birnbaum & Wan (2020): http://journal.sjdm.org/vol15.1.html

TABLE 1: Preference Patterns and Compatible Decision Rules)

Preference Pattern	B&P notation	Compatible Decision Rules/Models
111	123	MPW, ADM
112	121	MEDIAN
121	133	MAX; EV, EU, ADM
122	131	Number "sufficing" (> $12) prizes
211	223	MIN; EU, ADM
212	221	ADM, prior TAX
221	233	EU, ADM
222	231	ADM (regret)

Notes: X = (15, 15, 3), Y = (10, 10, 10), Z = (27, 5, 5); 111 denotes preference for X, Y, and Z in choices XY, YZ, and ZX, respectively. Patterns 111 and 222 are intransitive. "B&P pattern" indicates Butler and Pogrebna (2018) notation in which 1, 2, and 3 are used to denote preference for X, Y, or Z, respectively in Choices XY, YZ, and ZX. MIN, MEDIAN, MAX rules choose gamble with best Minimum, Median, or Maximum prize; MPW = Most Probable Winner; EU = Expected utility; EV = Expected value; ADM = Additive difference model

1.4 Theoretical Analysis

Birnbaum (2020) showed how different response patterns in the Butler and Pogrebna (2018) study could have been produced by different decision rules or by different parameters within the same decision model. Table 1 summarizes this analysis, but it uses a different notation system from that used by Butler and Pogrebna and by Birnbaum (2020). Therefore, Table 1 shows the connection between the two notation systems. The triple analyzed is X = (15, 15, 3), Y = (10, 10, 10), and Z = (27, 5, 5).

The intransitive pattern, 111, indicates X ≻ Y, Y ≻ Z, and Z ≻ X in Choices XY, YZ, and ZX, respectively. This pattern was denoted 123 in Butler and Pogrebna (2018), where 1, 2, or 3 were used to indicate preference for X, Y, or Z, respectively, in the same choice problems.

The Most Probable Winner model (MPW) implies this intransitive 111 preference pattern with either dependent or independent gambles.

If a person were to choose the gamble with the better minimum (MIN), median (MEDIAN) or maximum (MAX) prizes, then the preference patterns for these gambles would be 211 (Y ≻ X, Y ≻ Z, and Z ≻ X), 112 (X ≻ Y, Y ≻ Z, and X ≻ Z), and 121 (X ≻ Y, Z ≻ Y, or Z ≻ X), respectively.

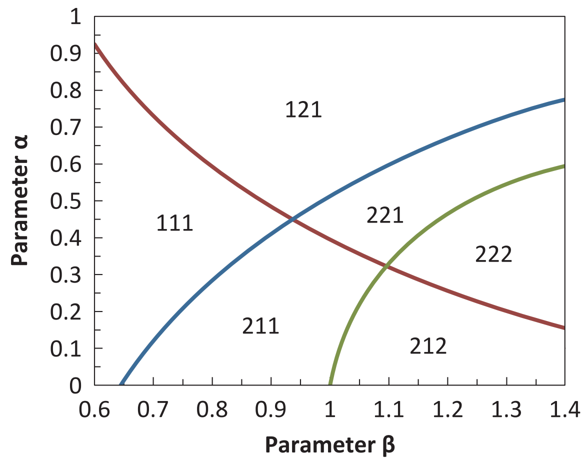

FIGURE 1: Preference patterns in relation to parameters of the additive difference model for dependent gambles. The patterns 111 and 222 are intransitive.

The triples were designed so that expected value (EV) would produce the ordering 121. Expected utility (EU) theory with power function for utility can (with different parameter values) imply three of the transitive orders, 121, 211, and 221 including that of EV (121), which is a special case of EU.

Birnbaum's (2008) TAX model with its "prior" parameters (estimated roughly in 1995) implies the pattern 212, but like EU, which is a special case of TAX, it can also imply other patterns. But TAX and EU are transitive, so they cannot imply true preference patterns of 111 or 222, no matter what parameters they use.

Suppose a prize of 12 is considered "good enough," or "satisficing". Because there are two prizes in X greater than 12, one prize in Z greater than 12 and none greater than 12 in Y, a rule to pick the gamble with more "satisficing" outcomes could have the preference pattern 122, if \$12 is satisfactory.

The additive difference model (ADM), described in the next section, can handle both transitive and intransitive response patterns, depending on the values of its parameters.[2]

1.5 Additive Difference Model (ADM)

For dependent gambles with equally likely branches, the additive difference model (ADM), with power functions (Birnbaum and Diecidue, 2015, Equations 10 and 13), for gambles $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$, can be written:

$$\psi(X, Y) = \sum \sigma(x_i, y_i)|x_i^\alpha - y_i^\alpha|^\beta \quad (2)$$

where X ≻ Y if and only if $\psi(X, Y)$ is positive; α and β are parameters; and $\sigma(x_i, y_i)$ is the augmented sign function $(-1, 0, 1)$ that retains the sign of $x_i - y_i$. This model is fairly general (Birnbaum & Diecidue, 2015) and can be used to represent regret theory (Loomes & Sugden, 1982) as well as advantage-seeking models, like most probable winner, which is an extreme special case.[3]

As shown in Figure 1 (and Table 1), this additive difference model with dependent gambles can imply six preference patterns (111, 121, 221, 211, 212, and 222) when the two parameters vary over the range in Figure 1. The intransitive pattern of 111 is implied, for example, when $\alpha = 0.4, \beta = 0.7$, and the opposite intransitive cycle, 222, is implied for the same α when $\beta = 1.3$; $\beta > 1$ has a "regret" interpretation (Loomes & Sugden, 1982; Birnbaum & Diecidue, 2015). If different people have different parameters, they can show different preferences, and if a person has stochastic parameters that drift in value between sessions, then that person's true preferences could vary from session to session.

1.6 Model of Stochastic Parameters

It seems reasonable to suppose that information (education) can affect the parameters of decision making. But even in a short experiment devoid of systematic new information, it is theorized that "random" factors (spontaneous thoughts and momentary emotions) might cause parameters to drift or fluctuate from session to session (Bhatia & Loomes, 2017; Birnbaum, 2013; Birnbaum & Wan, 2020).

A model to illustrate how parameters in the ADM model might change gradually from session to session has been implemented in a simulation program that is freely available Online at the following URL:
http://psych.fullerton.edu/mbirnbaum/calculators/ADM_sim.htm

In this simulation program, parameters change from Session t to Session $t + 1$ as follows:

$$\alpha(t + 1) = w\alpha(t) + (1 - w)ran(\alpha) \quad (3)$$

$$\beta(t + 1) = w\beta(t) + (1 - w)ran(\beta) \quad (4)$$

where $ran(\alpha$ and $ran(\beta)$ are randomly selected values of the parameters, which in the program are sampled independently from a uniform distribution on a range that the user can specify; $\alpha(t)$ and $\beta(t)$ are the effective value of the parameters in Session t; w is a weight that determines how stable parameters will be over time; when $w = 1$, parameters stay fixed and when $w = 0$, they are chosen randomly and

[2]The models in Table 1 are not exhaustive, because many other decision models have been or might be constructed to make predictions here.

[3]The additive difference model implies the property of restricted branch independence, which has been significantly violated in a number of studies (e.g., Birnbaum, 2008; Birnbaum & Diecidue, 2015). It is sometimes said that "all models are wrong, but some are useful." This model is useful here to illustrate how different preference patterns can be produced by changing parameters, with the caveat that despite its flexibility for fitting the design of Butler and Pogrebna (2018), it might not survive as a viable descriptive theory when properties such as branch independence are tested.

independently in each new session. The larger the value of w, the less parameters change from session to session; i.e., the more "gradual" the random walk.[4]

Birnbaum and Wan (2020) modeled the random walk in terms of preference patterns rather than in terms of model parameters (which determine preference patterns). The "gradual" models they simulated had the property that a preference pattern would likely stay the same between two successive sessions and tend to change to similar patterns. Thus, the model of Equations 2 and 3 provides specific premises from which one might deduce the kinds of gradual MARTER models that were postulated in Birnbaum and Wan (2020).

1.7 Response and sequence independence

Some "random utility" or "random preference" models imply that responses will satisfy the assumption of independence and identical distribution (iid). This assumption has been used in statistical tests of transitivity (e.g., Regenwetter, et al., 2011), but there is strong evidence against iid, even in the Regenwetter, et al. (2011) data (Birnbaum, 2011, 2012, 2013).

For "gradual" random walks, Birnbaum and Wan (2020) showed that tests of sequence independence will be violated in a similar fashion to what has been observed in empirical data (Birnbaum & Bahra, 2012a, 2012b). In particular, Birnbaum's (2012) correlation test should show a positive correlation between the number of preference reversals and the number of intervening sessions, according to gradual MARTER models: People are predicted to be more consistent in their responses when tested closer together in time than when tested farther apart in time. This correlation should be zero according to models that imply iid.

When there is a mixture of true preference patterns, response independence is expected to be violated in TE models, even without sequential effects (Birnbaum, 2013). Birnbaum and Wan (2020) illustrated how Birnbaum's (2012) variance test of iid is violated even when parameters are randomly selected for each new session in simulated MARTER models. In particular, TE models imply that people will be more consistent between replications than required by iid, unless they have only one true preference pattern.

In this study, we can therefore apply these tests of iid for each participant, to compare family of iid models against the family of models (such as gradual MARTER models) that violate response independence and sequence independence.

[4] Instructions for using the program are included in the Website. The program reports the values of the parameters and the implied preference pattern. The results from the program could be plotted on Figure 1 to illustrate a two-dimensional random walk and to illustrate the corresponding sequence of true preference patterns that would be implied by changing parameters. The JavaScript for the program is entirely contained in the single Web page, so one could easily revise this program to explore other models for stochastic fluctuation of parameters that generate other MARTER models.

2 Method

The participants' task was to choose between pairs of gambles, each of which consisted of three equally likely outcomes. The prize of a gamble would depend on the color of marble drawn blindly from a single urn containing an equal number of red, white, and blue marbles. Each choice problem was displayed as a table with red, white and blue columns, where rows represented the gambles, and entries represented the prizes of that gamble if that color of marble were drawn, as in Birnbaum and Diecidue (2015, Figure 2).

2.1 Instructions and Displays

The instructions, format for display of the choices, and one session of trials can be viewed at the following URL:
http://ati-birnbaum.netfirms.com/Spr_20/MPW_01.htm

The stimulus displays and Web forms were constructed and randomized using a JavaScript program by Birnbaum that is now freely available Online at the following URL:
http://psych.fullerton.edu/mbirnbaum/programs/ChoiceTableColorWiz2.htm

Each choice problem was presented in the format of a table with two rows representing the two choice alternatives and with three columns, colored red, white, and blue, representing the random events. Numerical entries indicated money prizes to be won if a marble drawn randomly from an urn was red, white, or blue, where the urn contained exactly 33 red, 33 white, and 33 blue marbles.

2.2 Design

There were 4 triples of gambles, based on Choice Triplets #3, 4, 7, and 10, as numbered in Butler and Pogrebna (2018). These triples showed the highest incidence of intransitive behavior in Table 1 of their paper. These triples are renumbered 1, 2, 3, and 4 in this paper, respectively. The same numerical values were used as in Butler and Pogrebna, except the numbers were said to represent dollars instead of pounds (the exchange rate was approximately 0.81 pounds/dollar during the study). The amounts are as follows:

Triple 1: $X = (12, 12, 2)$; $Y = (8, 8, 8)$; $Z = (20, 4, 4)$.
Triple 2: $X = (15, 15, 3)$; $Y = (10, 10, 10)$; $Z = (27, 5, 5)$.
Triple 3: $X = (9, 9, 3)$; $Y = (6, 6, 6)$; $Z = (16, 4, 4)$.
Triple 4: $X = (14, 14, 2)$; $Y = (8, 8, 8)$; $Z = (21, 6, 6)$.

Each session consisted of a block of 26 randomly ordered trials (choice problems). There are six choice problems (trials) for each triple as follows: XY, YZ, and ZX; and YX, ZY, and XZ, where XY and YX denote the same choice problem, except X is displayed in the first or second position. With four triples and six choice problems per triple, there are 24 experimental choice problems. Two additional "check"

TABLE 2: Crosstabulation. Frequencies of response patterns in first (rows) and second (columns) repetitions for participant S20

Rep 1	111	112	121	122	211	212	221	222	Sum
111	35	1	1	0	1	0	0	0	38
112	1	1	0	4	0	5	0	2	13
121	0	0	0	2	0	0	0	0	2
122	1	0	1	3	0	0	2	7	14
211	0	0	0	0	1	2	0	1	4
212	0	6	0	1	2	26	1	1	37
221	0	0	1	0	0	0	0	3	4
222	0	0	1	2	0	1	1	3	8
Sum	37	8	4	12	4	34	4	17	120

Total n = 120 = 4 Triples by 30 sessions, each based on 6 responses (3 choice problems by 2 repetitions) per triple, or 720 binary choices. 111 is the intransitive pattern predicted by most probable winner rule

trials testing transparent dominance were included in each session: $T = (10, 9, 8)$ versus $U = (8, 8, 8)$. and $V = (10, 10, 7)$ versus $W = (12, 12, 8)$. Note that T dominates U and V is dominated by W. The 26 trials were randomly intermixed and re-ordered for each session. There were 30 sessions.

2.3 Procedure

When each session was complete, the participant pushed a button to submit the responses for that session, and then pressed another button to load the materials for the next session. Participants worked at their own paces, and completed 30 sessions within 2 hours.

Students participated via the Internet during the COVID-19 shut down of April, 2020. Instructions stated that three participants would be chosen at random to receive the prize of one of their chosen gambles, so they should choose wisely. Procedures for playing gambles and determining prizes were similar to those in Birnbaum and Diecidue (2015, Experiment 6), except winners were not present when gambles were played out; prizes were sent as cash in the mail.

2.4 Participants

The participants were 24 undergraduates (ages 18 − 22, including 9 males) who received credit as one option toward an assignment in Introductory Psychology.

Because each of the 12 choice problems was presented twice in each session with display position (First or Second) counterbalanced, a person who mindlessly pushed the same button in a session would show zero consistency, and a person who pushed buttons randomly would show 50% agreement. There were 60 tests of dominance per person (2 trials per session by 30 sessions). Two participants were found with mean agreement within session of 51% and 54% and who violated dominance 50% and 52% of the time. Data for these two inconsistent participants are not included in the tables that follow. The remaining 22 participants had median agreement of 90% within sessions and median agreement with transparent dominance of 92%.

3 Results

Table 2 shows frequency (count) of each combination of responses observed in Replicate 1 (rows: XY, YZ, and ZX) and Replicate 2 (columns: YX, ZY, and XZ) for one participant, S20, aggregated over the four triples. Entries on the diagonal represent cases where the person made the same responses on all three choice problems on both replications within sessions. For example, the entry of 35 in Row 111 and Column 111 indicates that this participant chose X over Y, Y over Z, and Z over X on both replicates of these choice problems 35 times, aggregated over 120 trials (30 sessions and 4 triples). The 111 pattern is the intransitive pattern implied by the Most Probable Winner (MPW) rule in all four triples. This participant, S20, also repeated the transitive 212 pattern 26 times.[5]

A table like Table 2 was constructed for each participant. Counts in these tables that are off-diagonal represent response patterns where there was at least one preference reversal between two replications of the three choice problems of a triple in a session.

The TE model was fit to each individual's 8 by 8 crosstabulation matrix, as in Table 2. Birnbaum's (2013) Excel spreadsheet [available as *TE8x8_fit.xlsx*, from the SJDM website

[5]Raw data for Participant S20 are included in the Appendix.

Table 3: Within-session agreement, conformity to transparent dominance, and parameter estimates in the True and Error Model.

Case	Agree	Dom	e_1	e_2	e_3	p_{111}	p_{112}	p_{121}	p_{122}	p_{211}	p_{212}	p_{221}	p_{222}
S16	99	00	01	00	00	00	00	00	00	00	100	00	00
S24	99	00	01	00	01	00	00	00	00	00	100	00	00
S02	84	20	13	04	10	00	04	02	00	02	90	00	02
S11	76	30	15	14	13	00	00	05	02	09	80	02	02
S05	78	42	18	08	12	00	08	02	03	00	83	04	00
S04	96	97	02	02	02	00	27	00	00	00	73	00	00
S10	63	78	24	19	29	00	16	00	05	13	65	01	00
S08	63	62	27	23	23	09	17	05	12	01	51	04	01
S22	60	68	25	21	33	00	00	06	10	01	72	11	00
S18	92	88	04	07	02	00	12	00	03	00	77	00	08
S13	97	100	02	03	00	00	02	00	24	00	56	00	19
S15	96	100	02	03	01	00	01	00	27	00	50	00	22
S20	83	95	13	06	09	34	01	00	15	00	39	01	11
S12	92	100	04	02	08	95	04	00	01	00	00	00	00
S17	95	100	00	04	03	52	00	48	00	00	00	00	00
S21	99	100	00	00	01	00	00	100	00	00	00	00	00
S14	80	78	18	03	15	02	00	96	01	00	00	02	00
S07	86	88	09	06	08	00	01	79	03	13	03	00	00
S23	88	100	05	01	15	00	00	36	05	00	00	27	33
S03	96	100	01	04	01	01	00	07	01	47	00	44	00
S06	99	98	00	00	00	00	00	00	00	100	00	00	00
S01	80	100	03	10	22	00	09	37	54	00	00	00	00
*g*TET			0.08	0.06	0.08	0.09	0.04	0.20	0.07	0.09	0.41	0.05	0.05

Agree = mean percentage agreement within session, Dom = percentage conformance with transparent dominance; Parameters estimated from $TE8x8\_fit.xlsx$. Values are shown as percentages, so 01 indicates 0.01 and 100 indicates 1.00.

supplement to Birnbaum and Wan (2020)] was used to find best-fit parameters to minimize G.[6]

3.1 TE Model Parameter Estimates

Table 3 shows the estimated parameters of the TE model for each participant, along with each person's mean within-session agreement ("Agree") and percentage conformance to transparent dominance ("Dom"). Each row represents a different participant, and the order of rows has been arranged so that participants with similar parameters appear together in the table. To save space in the table, entries are expressed as percentages, so 04 indicates 0.04, and 100 indicates 1.00.

The largest group (13 participants), listed in the first 13 rows in Table 3, had the modal preference pattern of 212. The 212 pattern means preference for the lowest risk alternative, Y over both X and Z and preference for Z over X; e.g., Y = (10, 10, 10) preferred over both X = (15, 15, 2) and Z = (27, 5, 5), and Z = (27, 5, 5) preferred over X = (15, 15, 2). This is the transitive pattern implied by the special TAX model with "prior" parameters (Birnbaum, 2008).

Whereas the first few participants in this group (S16 and S24) appear to have only one preference pattern, some of these first 13 participants (e.g., S13, S15, and S20) also appear to have mixtures that include other preference patterns.

Three participants, S20, S12, and S17, show evidence of the 111 preference pattern implied by the MPW model. Although S20 had a modal response pattern of 212, this participant is estimated to have used the 111 pattern 34%

[6]This program and others are available from the following URL in the supplement to Birnbaum and Wan (2020): http://journal.sjdm.org/vol15.1.html

TABLE 4: Tests of TE, Transitivity, and Independence

Case	G (TE, 53)	G Trans (2)	G Indep (60)
S16	1.44	0.00	1.44
S24	3.27	0.00	3.27
S02	52.35	5.88	84.66
S11	63.43	0.65	103.97
S05	107.60	0.00	133.87
S04	28.58	0.00	130.27
S10	70.68	0.00	79.21
S08	55.66	2.37	76.94
S22	69.62	0.00	93.63
S18	27.59	21.51	110.28
S13	17.22	81.04	273.51
S15	22.91	95.63	284.88
S20	83.83	102.58	317.48
S12	55.77	171.93	67.81
S17	15.59	102.65	117.82
S21	2.79	0.00	2.79
S14	46.83	3.95	53.28
S07	112.95	0.00	250.77
S23	23.51	41.87	156.03
S03	52.23	6.67	219.65
S06	2.80	0.00	2.80
S01	41.91	0.00	69.13

Notes: TE = True and Error Model, Trans = Transitivity, Indep = Response Independence; Critical values of χ^2 with $\alpha = 0.01$, for df = 53, 2, and 60 are 79.84, 9.21, and 88.38, respectively.

TABLE 5: Tests of iid

Case	Mean	Var	p_V	r	p_r
S16	0.19	0.17	1.000	-0.42	0.657
S24	0.32	0.26	1.000	0.18	0.861
S02	5.60	16.37	0.000	0.88	0.000
S11	8.66	21.57	0.000	0.91	0.000
S05	7.31	28.16	0.000	0.79	0.007
S04	1.30	2.71	0.000	0.90	0.000
S10	10.19	12.19	0.000	0.37	0.286
S08	11.77	13.67	0.000	0.72	0.001
S22	11.54	15.80	0.000	0.67	0.004
S18	3.76	4.99	0.000	0.71	0.031
S13	1.28	2.00	0.001	0.90	0.000
S15	1.17	1.92	0.003	0.78	0.085
S20	9.62	31.59	0.000	0.96	0.000
S12	2.40	5.18	0.000	0.53	0.354
S17	1.38	2.49	0.001	0.87	0.002
S21	0.12	0.11	1.000	0.04	0.970
S14	5.25	7.49	0.000	0.90	0.000
S07	7.16	46.85	0.000	0.96	0.000
S23	4.70	5.90	0.000	0.97	0.000
S03	2.29	11.46	0.000	0.89	0.000
S06	0.13	0.12	1.000	0.03	0.976
S01	6.07	7.29	0.000	0.90	0.000

Notes: Mean and Var are the mean and variance of the number of preference reversals between sessions; r is the correlation between the mean number of preference reversals between sessions and the gap between sessions; Estimated $p-$values are based on 10,000 random permutations, using procedure of Birnbaum (2012).

of the time. Examining the raw data (Appendix), one can see that S20 changed from a modal 212 pattern to the 111 pattern after 21 sessions, and was perfectly consistent with this pattern on all trials in the last 8 sessions. S12 was estimated to have used the 111 pattern 95% of the time, and was perfectly consistent with this pattern for the last 12 sessions.

The raw data for S17 reveal that S17 was almost perfectly consistent with the 111 response pattern for Triples 1 and 2 (110 times out of 120), but this person consistently used pattern 121 (consistent with Expected Value) for Triples 3 and 4 (113 out of 120). The MPW model implies the 111 pattern in all four triples, so the data of S17 cannot be reconciled with MPW. However, such behavior would be compatible with the ADM model with parameters in the range of Figure 1; e.g., if $\alpha = 0.65$ and $\beta = 0.65$, then the person should show patterns 111 in Triples 1 and 2 and 121 in Triples 3 and 4.

The 121 transitive pattern, used by S17 for Triples 3 and 4, was also the modal pattern for S21, S14, S07, and S23. This preference pattern is consistent with Expected Value. S23 responded frequently with the 121 pattern for Triples 3 and 4, but about as often displayed the 222 pattern for Triples 1 and 2 (64 times out of 120 possible). The 222 pattern is intransitive and consistent with a "regret" interpretation in the ADM model.

S15 also showed evidence of the 222 pattern, but only for Triple 4 (54 of 60 possible), as did S13 (47 out of 60) for the same triple.

Participants S03 and S06 had 211 as the modal pattern, which would be consistent with choosing the option that has the best minimum prize, and S02 showed a mixture of 122 and 121 patterns.

3.2 Tests of TE and Transitivity

Each 8 by 8 frequency crosstabulation matrix has 63 degrees of freedom. The TE fitting model has 11 free parameters to approximate each 8 by 8 matrix (3 error rates and 8 probabilities of true patterns). Because the 8 probabilities of true preference patterns sum to 1, they use 7 df; therefore, the model consumes 10 df, leaving $63 - 10 = 53$ df to test the model. The index G has an (asymptotic) Chi-Square distribution with 53 df, according to the TE null hypothesis.

The transitive model is a special case of the TE model in which $p_{111} = p_{222} = 0$. Because this (transitive) model has 2 df fewer than the full TE model, the difference in G is (theoretically) asymptotically Chi-Square distributed with 2 df, under the null hypothesis that transitivity holds.

Table 4 shows G tests of fit of the TE fitting model to the crosstabulations for each individual, listed in the same order as in Table 3. Except for two cases (S05 and S07), violations of the TE model were not significant.[7]

The second column of G tests in Table 4 shows the $G(2)$ difference tests of transitivity; these tests compare the fit of the TE model with all parameters free against the transitive special case in which p_{111} and p_{222} are fixed to 0. The critical value of Chi-Square with 2 df (for $\alpha = 0.01$) is 9.21. Table 4 shows seven individuals with significant violations of transitivity, including S20, S12, and S17, who showed estimated incidence of the 111 pattern ranging from 34% to 95% (Table 3), and also S18, S13, S15, and S23, who showed smaller, but significant incidences of the 222 pattern ranging from 8% to 33% (Tables 3 and 4).

In order to confirm the statistical tests of transitivity by another method, the program $TE8x2\_fit.R$ was applied to each participant's partitioned data.[8] This program used 10,000 bootstrapping samples to generate 95% confidence intervals on the parameters. Exactly the same 7 participants who had significant violations of transitivity in Table 4 had lower limits for either p_{111} or p_{222} that were greater than zero: S20, S12, and S17, had lower limits for the 111 pattern of 86%, 41%, and 31%, respectively, and S18, S13, S15, and S23, had lower limits for the 222 pattern of 11%, 15%, 4%, and 16%, respectively. All other bootstrapped lower limits of intransitive behavior were essentially zero. Thus, the bootstrapping and conventional tests of significance were in agreement. It is worth noting that S18 had an estimated incidence of only 8% with a 95% confidence interval from 4% to 18%, and yet the method was able to detect significant departures from transitivity.[9]

These sensitive tests, in which small violations of transitivity can be detected can be contrasted with older methods, such as testing the Triangle Inequality (TI). According to the TI, $1 \leq P(XY)+P(YZ)+PZX) \leq 2$. Of the seven cases that had significant violations of transitivity according to the TE analysis (S12, S13, S15, S17, S18, S20 and S23), the sums of binary choice proportions were 2.83, 0.83, 0.82, 2.49, 1.03, 1.67, and 1.04, respectively. Therefore, three cases satisfied the TI perfectly (S18, S20, and S23), two cases clearly violated TI (S12 and S17), and two cases (S13 and S15) might be considered close enough that one might retain the TI based on this test. The raw data for S20 are included in the Appendix. It should be clear that this person exhibited intransitive preferences in the last part of the study; however, S20 would be declared to be perfectly consistent with the TI because $P(XY) = 0.53$, $P(YZ) = 0.73$, and $P(ZX) = 0.40$. Therefore, one cannot rely on tests of the TI to find violations of transitivity. Whereas this point has been made previously with hypothetical data (e.g., Birnbaum, 2012; 2013; Birnbaum & Wan, 2020), cases like S20, S18, and S23 show that this possibility is not just hypothetical but occurs in real data.

The last line in Table 3 shows estimated parameters from a gTET analysis of all of the data combined. The estimated coefficients show that the estimated incidences of the preference patterns are in every case within 2% of the mean incidences, averaging over individuals. Averaged over individuals, estimated incidence of the 111 pattern is 0.09 and estimated incidence of 222 pattern is 0.05. Similar group analyses were performed for each of the four triples separately. Although there were some differences in incidences of different patterns, as discussed above, in all four triples, the response pattern most often repeated (and consequently estimated to have the highest probability) was the 212, followed by the 121 pattern.

In summary, Tables 3 and 4 show that although most individuals had modal preference patterns that were compatible with transitivity, seven had significant departures. One person had intransitive preferences consistent with the MPW model. Six others showed evidence of intransitive behavior in some portion of the study.

3.3 An Unexpected Result

Table 3 shows a result that was a surprise: The first five participants as listed violated transparent dominance more than half the time. In fact, S16 and S24, who were in other respects very consistent participants, did so on every trial. Recall the "check" trials were: T = (10, 9, 8) versus U = (8, 8, 8) and V = (10, 10, 7) versus (12, 12, 8), which had been designed to resemble trials in the main study, in order to "force" people to pay attention, but that plan appears to have backfired for these people.

Post hoc, it seems as if these five participants, who consistently preferred (8, 8, 8) to the "riskier" (12, 12, 2), (14, 14, 2), (20, 4, 4) or (21, 6, 6), may have assumed that (8, 8, 8), a

[7]See Birnbaum and Quan (2020) for a discussion of the robustness of TE model estimates to violations of the model.

[8]This open-source program is freely available from the the supplement to Birnbaum and Wan (2020), in the Judgment and Decision Making Website: http://journal.sjdm.org/vol15.1.html.

[9]Schramm (2020) has criticized $TE8x2\_fit.R$, and recommended Bayesian methods that he argues would be even more sensitive.

low variance gamble, would be preferable to any alternative with which it might be paired, so they always chose (8, 8, 8), perhaps without realizing that on some trials, (8, 8, 8) was actually dominated by the alternative. Similarly, if (10, 10, 7) were misread as (10, 10, 10) or if (12, 12, 8) were misread as (12, 12, 2), as in trials of in the main design, then one can understand how people might make systematic mistakes on these (less frequent) check trials because of the resemblance to the main design. This result occurred in 5 of the 13 who had 212 as their modal preference pattern.[10]

3.4 Tests of Independence

Some "random preference" or "random utility" models assume that people have a mixture of true preference patterns and independently sample from them on each trial. The probability of choosing X over Y in these models is assumed to be the sum of the probabilities of patterns in which X is preferred to Y. Models of this type imply that responses are independently and identically distributed (iid). In contrast, TE models allow systematic violations of iid, and MARTER models with certain assumptions imply specific patterns of violation of response and sequence independence (Birnbaum, 2012, 2013; Birnbaum & Wan, 2020).

The third column in Table 4 (G Indep) are tests of response independence, which is one aspect of iid. These G values indicate how well or how poorly the entries in each crosstabulation table (as in Table 2) can be reproduced from products of the binary response proportions. Response independence implies, for example, that the expected frequency of repeating the 111 pattern in both replicates can be calculated by the product as follows:

$$E_{111,111} = n[1 - P_1]^2[1 - P_2]^2[1 - P_3]^2 \qquad (5)$$

where P_1, P_2, and P_3 are the proportions of choosing the second gamble (Y, Z, or X) in the XY, YZ, and ZX choices, respectively.

Response independence is not implied by the TE models (Section 1.4), except in special cases, such as when the participant has only a single true preference pattern. In the more general case, where there is a mixture of true preference patterns, the TE model implies that the probability of repeating a true pattern will be greater than expected by the assumption of independence.

Table 4 shows that 13 of 22 individuals have significant violations of response independence by this G test.

Birnbaum (2012) provided two other tests of iid that can be applied with small samples. As illustrated in Birnbaum and Wan (2020), these tests allow one to distinguish different classes of stochastic models that violate iid in particular ways. Different MARTER models can violate one or both of these properties.

Table 5 presents the results of Birnbaum's (2012) *iid_test.R* analysis.[11] The data analyzed are a 30 (Sessions) by 26 (Choice problems) matrix for each person. The column in Table 5 labeled "Mean" shows the mean number of preference reversals between sessions (averaged over all pairs of sessions) for each participant, column "Var" shows the variance of these preference reversals, and column "r" shows the correlation coefficient between the average number of preference reversals between two sessions and the gap between those sessions.

According to the iid random utility models, we expect the variances of preference reversals to be small and the correlations to be zero, but according to gradual MARTER models in which parameters drift gradually (which can produce different but similar true preference patterns in successive sessions), we expect variances to be large and correlations to be positive, unless each person has only one true preference pattern. The entries p_V and p_r are simulated probability values, computed by randomly and independently permuting the columns of the raw data 10,000 times and re-calculating the test statistics. These numbers (p_V and p_r) represent the proportion of randomly permuted samples in which the simulated test statistic exceeds or equals the value observed in the actual data, so they are estimates of the probability of observing the data if the null hypothesis of iid held.

Table 5 shows that iid can be rejected for the Variance test for all cases except those four participants who were inferred from the TE analysis to have a single "true" preference pattern (S16, S24, S21, and S06). Of the 18 remaining participants, all 18 correlation coefficients were positive, and 15 of these 18 are significantly different from zero. In sum, Tables 4 and 5 contain overwhelming evidence that iid can be rejected as a description of these data, except for those four who apparently have only one true preference pattern. As shown in Birnbaum and Wan (2020), gradual MARTER models can produce results like these.

4 Discussion

The majority of participants (20 of 22) had transitive modal preference patterns, including 13 with the pattern 212 (Y > X > Z), implied by the TAX model with its prior parameters, 4 with the pattern 121, implied by Expected Value, 2 with the pattern 211 (best minimum prize), and 1 with 122 (consistent with a sufficing model). The TE analysis estimated that intransitive cycles were significant, but infrequent, accounting

[10]It seems doubtful that these statistically significant violations of dominance would generalize to a study using other stimuli in the main design; however, future experiments might investigate how the context of the experimental design could be used to induce such systematic violations that would occur if a participant failed to read or process carefully each choice problem.

[11]This open-source, free program is available from the Online supplements to either Birnbaum (2012) or Birnbaum and Wan (2020) at URL: http://journal.sjdm.org/vol15.1.html

for about 14% of the sessions, averaged over participants. Thus, one can say that most of the participants behaved according to transitivity most of the time.

However, TE analysis indicated that 7 individuals had significant violations of transitivity, at least part of the time. Three people had significant incidence of the intransitive pattern 111 (implied by MPW), one of whom appears to have used this pattern throughout the study, one who displayed it only in two of the four triplets, and one who switched to this response pattern only in the last part of the study. Four others showed smaller but statistically significant violations of transitivity via the 222 (regret) preference pattern.

A number of conclusions can be reached:

1. The hypothesis that everyone had the same true preference pattern, including the hypothesis that the MPW model is descriptive, can be rejected. Only one participant had data compatible with the MPW model, which implies the 111 preference pattern for all triplets. None of the other theories that imply only a single preference pattern (e.g., MIN, MEDIAN, MAX in Table 1) for each triplet can be retained for all participants.
2. The hypothesis that each individual had a transitive preference pattern or a mixture of transitive preference patterns can be rejected, ruling out transitive theories like EU, TAX, and others, no matter what parameters they use, as descriptive of all participants.
3. The hypothesis that each person has a fixed set of true preferences that might or might not be transitive, including the hypothesis that they are governed by different models with different parameters that are fixed for each individual, can be rejected. The TE analyses combined with tests of independence showed that many people had data that could be described as mixtures of preference patterns.
4. The hypothesis that all persons are governed by the same model with different parameters, where the parameters change over sessions cannot yet be rejected.
5. The hypothesis that individuals use different models or processes (as in Table 1), and can change among models from session to session cannot be rejected.

From these findings and the perspective of MARTER models, one should not expect perfect correlations in studies that compare estimated parameters from the same person obtained on two occasions that are separated by a considerable gap of time.

Instead of assuming that all persons are governed by the ADM model (as in Figure 1), and a person might change true preferences from the 212 pattern to the 111 pattern because she changed parameters within that model, one might theorize instead that each person uses a different strategy for decision making (such as those listed in Table 1), and in this case she might switch from using the transitive TAX model (212) to a different choice process, choosing by the MPW model (111).

By principles of simplicity, one would prefer to retain a single model with variable parameters rather than to theorize that people change models. Nevertheless, criteria for arguing on the basis of data that a model has changed might include scale convergence and other principles of converging operations (Birnbaum, 1982; Mellers, Ordóñez, & Birnbaum, 1992). But a larger burden of evidence should apply to those who argue that different people use different processes.

An even more complex representation has been suggested, which is that people not only have different models and change from time to time, but they might use different decision rules for different choice problems. This approach is sometimes called the "adaptive toolbox", and seems to involve extra decision stages in which a person first decides what decision tools to use and then uses that tool to make the decision. So far, I have not been able to ascertain how one might test the implications of this "theory".

References

Bhatia, S. & Loomes, G. (2017). Noisy Preferences in Risky Choice: A Cautionary Note. *Psychological Review, 124*(5), 678–687.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp 401-485). Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

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

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comments on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review, 118*, 675-683.

Birnbaum, M. H. (2012). A statistical test of the assumption that repeated choices are independently and identically distributed. *Judgment and Decision Making, 7*, 97-109.

Birnbaum, M. H. (2013). True-and-error models violate independence and yet they are testable. *Judgment and Decision Making, 8*, 717-737.

Birnbaum, M. H. (2020). Reanalysis of Butler and Pogrebna (2018) using true and error mode. *Judgment and Decision Making, 15*(6), 1044-1051.

Birnbaum, M. H., & Bahra, J. P. (2012a). Separating response variability from structural inconsistency to test models of risky decision making, *Judgment and Decision Making, 7*, 402-426.

Birnbaum, M. H., & Bahra, J. P. (2012b). Testing transitivity of preferences in individuals using linked designs. *Judgment and Decision Making, 7*, 524-567.

Birnbaum, M. H., & Diecidue, E. (2015). Testing a class of models that includes majority rule and regret theories: Transitivity, recycling, and restricted branch independence. *Decision, 2*, 145-190.

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

Birnbaum, M. H., Navarro-Martinez, D., Ungemach, C., Stewart, N. & Quispe-Torreblanca, E. G. (2016). Risky decision making: Testing for violations of transitivity predicted by an editing mechanism. *Judgment and Decision Making, 11*, 75-91.

Birnbaum, M. H., & Quan, B. (2020). Note on Birnbaum and Wan (2020): True and error model analysis is robust with respect to certain violations of the MARTER model. *Judgment and Decision Making, 15*(5), 861-862.

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

Birnbaum, M. H., & Wan, L. (2020). MARTER: Markov true and error model of drifting parameters. *Judgment and Decision Making, 15*, 47-73.

Budescu, D. V., & Weiss, W. (1987). Reflection of transitive and intransitive preferences: a test of prospect theory.*Organizational Behavior and Human Decision Processes, 39*, 184–202.

Butler, D. J., & Blavatskyy, P. (2019). The voting paradox... with a single voter? Implications for transitivity in choice under risk. *Economics & Philosophy*, 2019, 1-19.

Butler, D. J., & Pogrebna, G. (2018). Predictably intransitive preferences. *Judgment and Decision Making, 13*, 217-236.

Cavagnaro, D.R., & Davis-Stober, C. P. (2014). Transitive in our preferences, but transitive in different ways: An analysis of choice variability. *Decision, 1*, 102-122.

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

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review, 109*(1), 137-155.

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., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92,* 805–824.http://dx.doi.org/10.2307/2232669

Luce, R. D. (2000). *Utility of gains and losses: measurement-theoretical and experimental approaches*. Mahwah,NJ:Lawrence Erlbaum Associates.

Mellers, B. A., Ordóñez, L., & Birnbaum, M. H. (1992). A change-of-process theory for contextual effects and preference reversals in risky decision making. *Organizational Behavior and Human Decision Processes, 52*(3), 331-369.

Müller-Trede, J., Sher, S., & McKenzie, C. R. M. (2015). Transitivity in context: A rational analysis of intransitive choice and context-sensitive preference. *Decision, 2*, 280-305.

Ranyard, R., Montgomery, H., Konstantinidis, E., & Taylor, A. L. (2020). Intransitivity and transitivity of preferences: Dimensional processing in decision making. *Decision, 7(4)*, 287–313. https://doi.org/10.1037/dec0000139

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of Preferences. *Psychological Review, 118*, 42–56.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. (2006). Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice. *Journal of Economic Literature, 44 (3)*, 631-661.

Schramm, P. (2019). The individual true and error model: Getting the most out of limited data. *Judgment and Decision Making, 15(5)*, 851-860.

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.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*, 31-48.

Appendix: Raw Data for S20

Table 6 shows the responses by one participant (S20) to the 24 trials testing transitivity. Each row represents a different session, and each column represents a set of three responses to choice problems XY, YZ, and ZX. R1 and R2 refer to the two replications, which were intermixed in the session, but used different positions of the gambles within the display. For example, response pattern in the first row and first column (T1 R1) is 212, which indicates that the person chose Y over X, Y over Z, and X over Z on Triple 1 in the first replicate (R1) of the first session. The column labeled T1 R2 shows the responses to the same three choice problems, except in this replication the positions of the gambles were reversed in the displays. The response pattern 112 in the first row and second column indicates that this participant reversed preferences on the XY choice, choosing X over Y on this replication (R2) in the first session. The column labeled "agree" shows that in the first session, this participant had two preference reversals between the two replications in the first session. The mean of this column divided by 12 (the number of distinct choice problems in the main design) is the consistency index for this participant, .83, shown as 83% in Table 4 for S20. This participant ranged from 7 to 11 agreements for the first 21 sessions, but S20 became perfectly consistent in the last 8 sessions with the intransitive 111 pattern.

Table 6: Response patterns and within-session agreement for participant S20.

Session	T1 R1	T1 R2	T2 R1	T2 R2	T3 R1	T3 R2	T4 R1	T4 R2	agree
1	212	112	212	212	212	212	122	222	10
2	212	212	212	212	212	212	122	222	11
3	212	212	212	211	111	211	221	222	9
4	212	112	111	112	121	122	122	222	8
5	212	112	112	212	122	221	112	122	7
6	212	112	212	112	222	222	122	222	9
7	212	212	212	212	122	122	221	222	11
8	112	212	212	212	212	222	122	111	8
9	112	112	212	212	112	222	222	122	9
10	112	212	212	212	112	122	222	122	9
11	112	212	212	112	112	122	122	122	9
12	212	212	112	212	222	222	122	222	10
13	212	212	212	212	222	222	221	222	11
14	212	212	212	212	222	221	222	121	9
15	211	212	212	212	122	122	112	122	10
16	212	212	212	211	112	222	121	122	8
17	212	212	212	212	122	222	122	221	9
18	212	212	212	212	122	121	221	121	10
19	212	212	212	212	122	222	222	212	10
20	212	212	212	212	212	221	211	222	8
21	211	212	211	211	112	111	212	122	8
22	111	111	111	111	111	121	111	111	11
23	111	111	111	111	111	111	111	111	12
24	111	111	111	111	111	111	111	111	12
25	111	111	111	111	111	111	111	111	12
26	111	111	111	111	111	111	111	111	12
27	111	111	111	111	111	111	111	111	12
28	111	111	111	111	111	111	111	111	12
29	111	111	111	111	111	111	111	111	12
30	111	111	111	111	111	111	111	111	12

Note: 111 is the intransitive pattern predicted by most probable winner (MPW) rule