

EMPIRICAL ARTICLE

# Dominance and transitivity

Michael H. Birnbaum 

Department of Psychology, California State University, Fullerton, Fullerton, CA, USA

**Corresponding author:** Michael H. Birnbaum; Email: [mbirnbaum@fullerton.edu](mailto:mbirnbaum@fullerton.edu)

**Received:** 29 May 2023; **Revised:** 14 August 2023; **Accepted:** 11 October 2023

**Keywords:** risky decision making; transparent dominance; transitivity of preference

## Abstract

This article describes a study of transitivity of preference and of transparent dominance with 220 participants who judged each choice problem 4 times. It shows how a true-and-error model with 2 error terms per choice problem can be applied to replicated data, to ask if violations of dominance or of transitivity are ‘real’ rather than due to random response errors. These models allow one to estimate the incidence of systematic violations and of error rates. The new data showed about 3% violations of transitivity, corrected for error. This incidence might be statistically significant, but a skeptic might dismiss it as too small to build a theory upon. Tests of dominance found violations with overall rates from 4% to 18%. As in previous research, violations of transparent dominance appeared almost exclusively among people who systematically prefer ‘safe’ gambles (with low ranges of outcomes) over ‘riskier’ gambles with higher expected values when the ‘safe’ gamble was dominated by the higher-ranged gamble. For those participants and choice problems, rates of violation were 28%–45%, corrected for error. It was theorized that these violations of dominance may be due to a subgroup of risk-averse participants using a strategy in this experiment to find ‘safe’ alternatives, without comparing outcomes between gambles.

## 1. Introduction

There are 2 principles of decision making that are regarded by many, but not all, theoreticians as rational: dominance and transitivity. *Dominance* is the principle that if option A yields consequences that are at least as good and sometimes better than the corresponding consequences for option B for every state of the world, then one should prefer option A over B. *Transitivity* holds that if one prefers X to Y and Y to Z, then one should prefer X to Z.

Transitivity and dominance are not only widely regarded as rational, they are also implied by certain descriptive theories of decision making such as expected utility (EU) theory and cumulative prospect theory (Tversky and Kahneman, 1992). However, other descriptive theories of risky decision making satisfy transitivity but can violate dominance, such as the transfer of attention exchange (TAX) and rank-affected multiplicative weights models (Birnbaum, 2008). Another group of theories can violate transitivity and must satisfy dominance, such as the most probable winner (MPW) model (Blavatsky, 2006; Butler and Blavatsky, 2020; Butler and Pogrebná, 2018) and regret theory (Loomes and Sugden, 1982). Therefore, testing dominance and transitivity empirically allow psychologists to learn which models can be ruled out or retained as possible descriptive theories of how people actually make decisions.

### 1.1. Background and purposes of the present study

Butler and Pogrebna (2018) constructed a set of choice problems in which systematic violations of transitivity were observed. Their design used sets of 3 gambles ('triples') with 3 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 2 equal chances to win 15 pounds and 1 equal chance out of 3 to win 3 pounds. The choice problems were designed such that every choice compared a 'safer' option (with lower range of outcomes) against a 'riskier' alternative (higher range) that had a higher expected value.

They were also devised such that  $X$  has better outcomes than  $Y$  for 2 of 3 branches,  $Y$  beats  $Z$  on 2 branches, and  $Z$  beats  $X$  on 2 branches. If people choose gambles that are most likely to give a better outcome, called the MPW strategy, they would choose  $X$  over  $Y$ ,  $Y$  over  $Z$ , and  $Z$  over  $X$ , violating transitivity. Although their experiment was designed to investigate this MPW model, and some violations of this type were observed, Butler and Pogrebna (2018) reported that their results showed more violations of the opposite type from those implied by MPW. Birnbaum (2020) reanalyzed those data via the group true-and-error (TE) model and concluded that 4 of the 11 triples showed evidence of small but significant violations of transitivity (Butler, 2020).

These 4 triples were later tested in a new study with 22 individuals who responded 60 times to each choice problem (Birnbaum, 2023a), which allowed analysis via the individual TE model. Birnbaum (2023a) concluded that although most individuals satisfied transitivity, 1 person showed violations consistent with the MPW model, and 6 others showed evidence of intransitive behaviors in at least one of the triples at least part of the time.

Birnbaum (2023a) also tested dominance using test trials interspersed among the transitivity trials and found surprisingly high rates of violation by some individuals. There were 5 participants who violated transparent dominance more than half the time, including 2 who violated dominance 60 times out of 60 tests despite being almost perfectly self-consistent and transitive in their responses on 720 other trials with non-dominated choices.

In Birnbaum's (2023a) dominance tests, a low-variance alternative was always dominated by a higher-variance alternative. All 5 of the participants who violated dominance systematically preferred 'safe' (low-variance) alternatives on other choice problems not involving dominance. It was conjectured that the experimental design testing transitivity may have induced a strategy in some individuals to look for the 'safe' alternative, which left them vulnerable to violating dominance in tests where the low-variance, 'safe' alternative was dominated. It was hypothesized that if choice problems were used in which the 'safe' alternative was dominant, these people would likely satisfy dominance (Birnbaum, 2023a).

This article builds on the designs of Butler and Pogrebna (2018) and Birnbaum (2023a) to test this new hypothesis about dominance violations and to reassess evidence of intransitive preferences, as measured by a more general TE model than has been employed in previous research on transitivity. The more general model allows 2 error rates per choice problem instead of 1. This study investigates (1) whether violations of transitivity observed by Butler and Pogrebna (2018) and Birnbaum (2023a) can be observed with a new, larger sample in order to check the estimated incidence with the more complex TE model; (2) whether the rate of violation of transparent dominance will be significant and higher among persons who most often prefer low-variance gambles (as previously reported); and (3) whether the rate of dominance violations would be lower in new tests of dominance in which the lower-variance alternative dominates the higher-variance alternative (compared against the case in which the dominant gamble has the higher variance).

### 1.2. Dominance violations

Research on dominance has discovered situations in which a majority of college undergraduates systematically violate first-order stochastic dominance (Birnbaum, 1999; Birnbaum and Navarrete, 1998). These violations were implied by configural weight models as fit to previous data (Birnbaum

and Chavez, 1997), and the predicted violations were published before research had been done on the problems (Birnbbaum, 1997). For example, about 70% of college students choose  $G = (\$96, 0.85; \$90, 0.05; \$12, 0.10)$  over  $F = (\$96, 0.90; \$14, 0.05; \$12, 0.05)$  even though  $F$  dominates  $G$  by first-order stochastic dominance. When the gambles  $F$  and  $G$  are presented in canonical split form (in which the number of branches of the gambles are equal and the probabilities of corresponding, ranked branches are the same), it is found that the vast majority prefer the split form of the dominant gamble; for example,  $F' = (\$96, 0.85; \$96, 0.05; \$14, 0.05; \$12, 0.05)$  is preferred to  $G' = (\$96, 0.85; \$90, 0.05; \$12, 0.05; \$12, 0.05)$ , satisfying dominance. Since  $F'$  and  $G'$  are equivalent to  $F$  and  $G$ , respectively, except for splitting, this reversal of the majority preference is also a violation of coalescing (a 'splitting effect').

Another line of research has found systematic violations of monotonicity, finding that when the lowest consequence of a gamble is increased from 0 to a small positive amount, the judgment of value of a gamble can actually decrease. Such violations have been found in judgment- and choice-based certainty equivalents, but not in direct choice (Birnbbaum, 1992, 1997; Birnbbaum, Coffey, Mellers, and Weiss, 1992; Birnbbaum and Sutton, 1992; Birnbbaum and Thompson, 1996; Mellers et al., 1992b). For example, people offer a higher price to buy (and demand a higher price to sell)  $M = (\$96, 0.9; \$0)$  than  $N = (\$96, 0.9; \$12)$ , when these gambles are presented on separate trials intermixed among other gambles to evaluate. These violations have been theorized to result from lower configural weight applied to the outcome 0 than to small positive consequences (Birnbbaum, 1997).

The theoretical explanations of these 2 cases described in this section (in terms of configural weighting), however, are not applicable to the violations of transparent dominance observed in Birnbbaum (2023a). Further, the magnitude of the effects reported by Birnbbaum (2023a), when assessed over all participants, is much smaller than the other examples of dominance violation cited above. Therefore, some additional statistical tools are required for their analysis.

### 1.3. *The problem of response variability*

When testing formal properties such as dominance or transitivity with fallible data, properties might appear to be violated by random errors in responding. When confronted with the same choice problem on different occasions, people often make different choice responses. How do we distinguish whether violations of a principle are due to systematic behavior or instead to random errors in responding?

A family of models known as 'TE' models have been developed in a series of papers to address the problem of distinguishing true violations from those that might be due to error (Birnbbaum, 2004, 2008, 2010, 2013, 2023a, 2023b; Birnbbaum et al., 2016; Birnbbaum and Bahra, 2012a, 2012b; Birnbbaum and Gutierrez, 2007; Birnbbaum and Quispe-Torreblanca, 2018; Birnbbaum and Schmidt, 2008; Birnbbaum and Wan, 2020). Research using replicated designs analyzed by TE models concluded that most people satisfy transitivity of preference, but a minority of participants can be found who exhibit systematic deviations from transitivity in specially constructed designs (Birnbbaum, 2023a; Birnbbaum and Bahra, 2012b; Birnbbaum and Diecidue, 2015; Birnbbaum and Gutierrez, 2007).

The key idea in TE models is the use of replications to estimate error. When the same person is asked to respond to the same choice problem on 2 occasions in the same brief session, suitably separated by filler trials, it is assumed that true preferences remained the same and any differences in response are due to error. How these assumptions can be used to measure error in order to test dominance is described in the next section.

### 1.4. *A test of dominance*

Consider the choice problem displayed in Figure 1. There is an urn containing exactly 33 Red marbles, 33 White marbles, and 33 Blue marbles. A marble will be drawn blindly and randomly from the urn, and its color will determine the participant's prize depending on which option the participant chose. For

Which do you choose?

Choose	Colors / Prizes \$		
	Red	White	Blue
First Gamble	\$10	\$9	\$8
Second Gamble	\$8	\$8	\$8

Figure 1. An example choice problem, illustrating a test of transparent dominance.

example, if the participant chose the First Gamble and a Red marble is drawn, the prize would be \$10, but if the participant chose the Second Gamble, the prize for the Red marble would be only \$8. First Gamble dominates the Second Gamble because for any color of marble drawn, the prize is at least as high and for 2 colors (Red or White), the prize is higher.

Suppose that this choice problem was presented among many other choice problems and it were found that 25% of the expressed preferences violated dominance. Are such results evidence of systematic bias or simply due to error? Of course, it is an intellectual ‘mistake’ or ‘error’ for a participant to choose the Second Gamble, but the question is whether or not the 25% violations are due to a flawed, but systematic decision rule as opposed to random errors, such as misreading the problem, forgetting the information or decision, or accidentally pushing the wrong button. If a person were to use a decision rule such as ‘choose a sure thing when one is available’: such a person might choose the Second Gamble systematically without contrasting consequences for each event against those of the First Gamble, which would violate dominance.

Figure 2 illustrates TE components in a choice problem. Suppose A and B represent 2 gambles in a choice problem (such as the First and Second in Figure 1), and let > denote ‘truly chosen’, so that  $A > B$  means A is systematically preferred to B. People who truly chose  $A > B$  might respond B by random error with probability  $e$ , and those who prefer  $B > A$  might respond A with probability  $f$ .<sup>1</sup>

Let  $p_A$  denote the probability that  $A > B$  in a certain population. The observed choice proportion,  $P(A)$ , might be used to make inferences about a population choice probability,  $p(A)$ , but the question is, can we estimate the true preference probability,  $p_A$ , from the observed choice proportion or inferred choice probability? No, because the choice probability is given as follows:

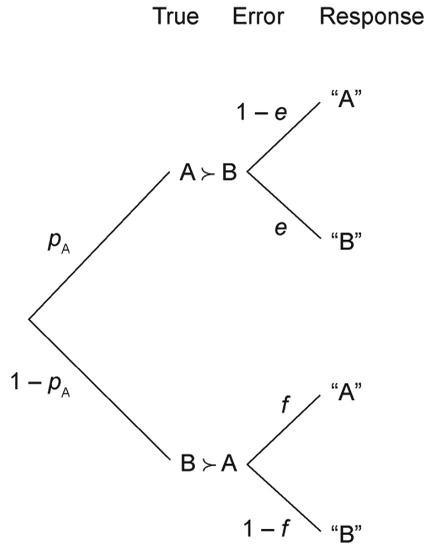
$$p(A) = p_A(1 - e) + (1 - p_A)f, \tag{1}$$

so from  $p(A)$  we cannot estimate  $p_A$ , unless we know or can estimate  $e$  and  $f$ .<sup>2</sup>

If someone were to assume that  $p_A = 1$  (e.g., that dominance is always satisfied), then the observation of 25% violations would indicate that  $e = 0.25$ ; alternatively, another person who assumed that  $e = f = 0$  might conclude that there are 25% true violations. However, these arbitrary conclusions might be refuted by empirical evidence, if a better experiment (with replications) had been done.

<sup>1</sup>The preference relation, >, representing ‘true’ preference should not be given too much excess meaning; it is entirely possible that the so-called ‘true’ preference relation may depend on the experimental context including experimental design and instructions. Indeed, if someone pointed out the mistake, it is not assumed the person would persist in choosing A. Thus, there may exist different ‘true’ relations,  $>_k$ , in different contexts or with different instructional conditions,  $k$ . The point here is to distinguish what is systematically driving choices in a particular experimental paradigm from what produces varied responses in the same situation. Thus, the word ‘true’ should not be imbued with excess meaning, for example, that it is also ‘correct’, or that it would be invariant with respect to context.

<sup>2</sup>The population probability,  $p(A)$ , is inferred from the observed choice proportion,  $P(A)$ . This inference should not be confused with the relation of the true preference probability,  $p_A$ , to the expressed preference probability,  $p(A)$ . In group TE theory, where each participant serves in a brief session,  $p_A$  represents the probability that a person holds the true preference,  $A > B$ , and different people may hold different preferences. In individual TE theory, the same person may have different true preferences in different sessions separated over time, so this parameter represents the probability that the individual has this preference state in a session.



**Figure 2.** True-and-error model of choice between A and B;  $p_A$  is the probability to truly prefer A over B;  $e$  and  $f$  are the probabilities to erroneously respond ‘B’ when A is truly preferred and to respond ‘A’ when B is truly preferred, respectively.

**Table 1.** Hypothetical data for a replicated choice problem.

Response in Replicate 1	Response in Replicate 2		Row totals
	A'	B'	
A	65	10	75
B	10	15	25
Column totals	75	25	100

*Note:* There are 75% responses for A in either replicate and 20% reversals between replicates.

**1.5. TE model with replications**

Suppose that the same choice problem is presented twice to the same participants in the same brief session, with both presentations suitably separated and embedded among a number of other filler trials. There are 4 possible response patterns: a person might choose A both times (AA), B both times (BB), switch from A to B (AB), or switch from B to A (BA).

Table 1 contains hypothetical data for 100 participants who responded twice to the same choice problem. The marginals show that 75% chose A in either replicate, but 20% switched responses between replications. Since there are 100 participants, these values can be interpreted as percentages.

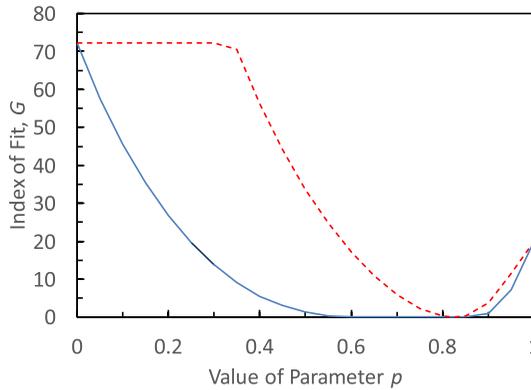
Table 2 shows the theoretical probabilities of these 4 response patterns,  $p(AA)$ ,  $p(AB)$ ,  $p(BA)$ , and  $p(BB)$  according to the TE model of Figure 2, assuming that true preferences are the same and errors on 2 replications have equal probability and are independent of each other. Table 2 shows that the probabilities of the 2 types of reversal are implied to be equal to each other,  $p(AB) = p(BA)$ , which can be used to test these assumptions.

There are 3 parameters in the model,  $p_A$ ,  $e$ , and  $f$ , and there are 3 degrees of freedom in the data of Table 1. The data of this design constrain the parameters, but the constraints do not impose a unique solution. To illustrate these partial constraints, Figure 3 shows the fit of the TE model (Figure 2) to

**Table 2.** TE4 analysis of replication of a single choice problem.

Response in Replicate 1	Response in Replicate 2	
	A	B
A	$p_A(1 - e)(1 - e) + (1 - p_A)(f)(f)$	$p_A(1 - e)(e) + (1 - p_A)(1 - f)(f)$
B	$p_A(1 - e)(e) + (1 - p_A)(1 - f)(f)$	$p_A(e)(e) + (1 - p_A)(1 - f)(1 - f)$

Note:  $p(A) = p_A(1 - e) + (1 - p_A)(f)$ .



**Figure 3.** Fit of true-and-error models as a function of the parameter,  $p_A$ , the probability that  $A > B$ . Solid line shows the fit of the 3-parameter model of Figure 2 (TE4), and the dashed line shows the fit of the 2-parameter model with  $e = f$  (TE2).

Table 1 for fixed values of  $p$ . For each value of  $p$ , error rates are selected to minimize  $G$ , which is defined as follows:

$$G = -2 \sum O_i \ln(O_i/E_i), \tag{2}$$

where  $O_i$  and  $E_i$  are the observed and ‘expected’ (fitted) frequencies of the 4 response patterns. In this case, the  $O_i$  are the entries of Table 1 and the  $E_i$  are the corresponding frequencies implied by Table 2; e.g., the ‘expected’ frequency of AA responses is  $np(AA)$  from Table 2, where  $n = 100$ . Minimizing  $G$  (aka  $G^2$ ) is equivalent to finding a maximum likelihood solution.

The solid curve in Figure 3 plots the best-fit value of  $G$  as a function of  $p_A$ . The model fits Table 1 perfectly ( $G = 0$ ) for a range of values, about  $0.65 < p_A < 0.85$ , where  $e$  and  $f$  trade off from  $e = 0.03$  and  $f = 0.35$  at  $p_A = 0.65$  to  $e = 0.10$  and  $f = 0.16$  at  $p_A = 0.80$  to achieve the perfect fit.

Although we cannot identify a unique best-fitting  $p_A$ ,  $e$ , and  $f$ , we can reject the model if we assumed that  $p_A = 1$ , as shown by the large  $G$  value in Figure 3. (The critical values of  $\chi^2(1)$  and  $\chi^2(2)$  are 6.34 and 9.21 with  $\alpha = 0.01$ .) That is, even though the 3-parameter model exhausted the df in the data, it still imposes constraints that allow one to test specific hypotheses about certain specific values of  $p_A$ .<sup>3</sup> The assumption that  $p_A = 1$  corresponds to the null hypothesis that dominance is always satisfied in a choice problem such as illustrated in Figure 1.

Furthermore, if we can assume that  $e = f$ , then the model is fully identified and testable; that is, we can estimate both  $p_A$  and  $e$  from the data, and we can test the model without specifying  $p_A$ .

<sup>3</sup>By assuming a fixed value of  $p_A$ , we have clawed back 1 df, rendering the special case testable. Note also that when  $p_A$  is fixed to 1,  $f$  drops out of the equations, leaving only 1 parameter to estimate. In sum, the TE model with  $p_A = 1$  might deserve rejection, as in the example of Figure 3; it might fit acceptably by statistical standards; or it might even fit perfectly. For example, if the 4 entries of Table 1 had been 64, 16, 16, and 4 (instead of 65, 10, 10, 4, respectively), then these data would be perfectly consistent with TE with  $p_A = 1$  and  $e = 0.2$ .

This 2-parameter model, in which each choice problem is allowed to have a different true probability and a different error rate, is termed the TE2 model in Birnbaum and Quispe-Torrealblanca (2018), and the more complex model of Figure 2, with 2 error terms for each choice problem, is known as TE4.

With  $e = f$ , it follows from the equations (Table 2) that the sum of response reversals between replicates is a quadratic function of  $e$ , as follows:

$$p(AB) + p(BA) = 2(1 - e)e. \tag{3}$$

Because there are 20% reversals between replications in Table 1,  $e = 0.113$ .<sup>4</sup> Once  $e$  is determined,  $p_A$  can be computed from the estimated binary choice proportion, substituted for  $p(A)$  in the expression,  $p(A) = p_A(1 - e) + (1 - p_A)e$ . For Table 1,  $0.75 = p_A(1 - 0.113) + (1 - p_A)(0.113)$ , which implies  $p_A = 0.823$ . In practice, one solves for  $p_A$  and  $e$  to fit all the data (e.g., all values in Table 1) minimizing  $G$ . Because this TE2 model uses only 2 parameters, there remains 1 df to test the model.

The dashed line in Figure 3 shows the index of fit of TE2 to Table 1 as a function of  $p_A$ ; unlike TE4 which has multiple solutions, TE2 has a unique minimum at  $p_A = 0.82$ . According to TE2, values of  $p_A$  in the range from about 0.75 to 0.90 are considered statistically acceptable, but values outside that range are not, including  $p_A = 1$ .<sup>5</sup> Although both TE2 and TE4 achieve a perfect fit to this example, that need not be the case. Both TE2 and TE4 would be rejected (binomial  $p < 0.01$ ) if the entries of Table 1 were instead 65, 5, 20, and 10, for example, because both models imply that  $p_{AB} = p_{BA}$ .

In sum, if we replicate a single choice problem, we can use either TE4 or TE2 to test whether violations of a property like dominance are ‘real’ (systematic), or instead might be due to random error. Furthermore, if we can assume that  $e = f$ , we can uniquely estimate the incidence of true violations.

### 1.6. True-and-error model of transitivity

The issue of transitivity is important because descriptive theories disagree about whether transitivity of preference can be systematically violated. Some important papers reviewing transitivity in terms of empirical results, theory, or error analysis include Bhatia and Loomes (2017), Birnbaum (2023a), Budescu and Weiss (1987), Cavagnaro and Davis-Stober (2014), Fishburn (1991), Gonzalez-Vallejo (2002), Iverson and Falmagne (1985), Leland (1998), Loomes and Sugden (1982), Luce (2000), Morrison (1963), Müller-Trede et al. (2015), Ranyard et al. (2020), Regenwetter et al. (2011), Rieskamp et al. (2006), Sopher and Gigliotti (1993), and Tversky (1969).

Much of the empirical research on the descriptive accuracy of transitivity has been highly controversial because that research has not used appropriate experimental designs coupled with appropriate quantitative analyses that can actually diagnose whether transitivity violations are real or due to error (Birnbaum, 2023a; Birnbaum and Wan, 2020). Birnbaum (2023a) and Birnbaum and Wan (2020) noted that quantitative methods used in the past can easily conclude that data generated by an intransitive process are ‘transitive’ or vice versa. Fortunately, TE models can be applied to separate the issue of systematic violation of transitivity from that of error, but one needs to do a proper experiment with replications to apply these methods.

Methods for analyzing transitivity via TE models have been presented in Birnbaum (2020, 2023a), Birnbaum et al. (2016), Birnbaum and Bahra (2012a), Birnbaum and Diecidue (2015), Birnbaum and Gutierrez (2007), Birnbaum and Schmidt (2008), and Birnbaum and Wan (2020). In these earlier papers, however, it had been assumed that each choice problem has a different error rate, but it was assumed  $e_i = f_i$ .

<sup>4</sup>The expression can be rewritten as  $2e^2 - 2e + p(AB) + p(BA) = 0$ ; from Table 1, we estimate  $p(AB) + p(BA) = 0.2$ , so  $e^2 - e + 0.1 = 0$ . The quadratic solutions to the equation,  $ax^2 + bx + c = 0$ , are given by the formula:  $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ . In this case,  $a = 1$ ,  $b = -1$ , and  $c = 0.1$ , so the 2 solutions are  $x = 0.113$  and  $x = 0.887$ . But  $0 \leq e < 1/2$ , so  $e = x = 0.113$ . If  $e > 1/2$ , it would mean that participants are liars; that is, they are more likely to report the opposite of their true preferences than to report true preferences.

<sup>5</sup>The horizontal part of the dashed curve for values of  $p_A$  below 0.4 occurs because  $e$  is restricted to be less than 1/2.

In this article, the more general model of Figure 2 where  $e$  need not equal  $f$  will be employed. It has not yet been ruled out that this more complex error specification might allow transitive models to be accepted that would have been rejected under the simpler models, despite simulations in Birnbaum and Quan (2020) showing that conclusions regarding transitivity appear to be robust with respect to the error specifications in TE models.

Consider a test of transitivity with three choice problems, XY, YZ, and ZX. If we code 1 to represent choice of the first listed option and 2 for choice of the second, there are 8 possible response patterns: 111, 112, 121, . . . , 222. For example, 111 is the intransitive pattern of choosing X over Y, Y over Z, and Z over X; 222 is the opposite intransitive pattern, and the other 6 patterns are transitive. However, if we present each choice problem twice (replicate), there are 64 possible response patterns for the 6 choice problems.<sup>6</sup>

The observed frequencies of the 64 response patterns can be fit to the TE model of Figure 2. In this case, there are 8 parameters representing the 8 possible true preference patterns,  $p_{111}, p_{112}, p_{121}, p_{122}, p_{211}, p_{212}, p_{221}$ , and  $p_{222}$ , and 6 error rates,  $e_1, e_2, e_3, f_1, f_2$ , and  $f_3$  for the 3 choice problems. Since the 8 true preference pattern probabilities sum to 1, they consume 7 df, so the parameters use  $7 + 6 = 13$  df. The 64 cell frequencies have 63 df because their probabilities also sum to 1, so there are  $63 - 13 = 50$  df remaining to test the model.

The ‘expected’ (i.e., ‘fitted’ or ‘predicted’) frequency for the response pattern 212, on the first replicate and the pattern 211 on the second (denoted 212, 211), for example, is given as follows:

$$\begin{aligned}
 E_{212,211} = n[ & p_{111}(e_1)^2(1 - e_2)^2(e_3)(1 - e_3) \\
 & + p_{112}(e_1)^2(1 - e_2)^2(1 - f_3)(f_3) \\
 & + p_{121}(e_1)^2(f_2)^2(e_3)(1 - e_3) \\
 & + p_{122}(e_1)^2(f_2)^2(1 - f_3)(f_3) \\
 & + p_{211}(1 - f_1)^2(1 - e_2)^2(e_3)(1 - e_3) \\
 & + p_{212}(1 - f_1)^2(1 - e_2)^2(1 - f_3)(f_3) \\
 & + p_{221}(1 - f_1)^2(f_2)^2(e_3)(1 - e_3) \\
 & + p_{222}(1 - f_1)^2(f_2)^2(1 - f_3)(f_3)],
 \end{aligned}
 \tag{4}$$

where  $E_{212,211}$  is the ‘expected’ (‘fitted’) frequency (count) that people will respond with the 212 and 211 response patterns in the first and second replications of the session. Note that if a person has the true preference pattern of 111, then she or he would have to make error  $e_1$  twice on the first choice problem, make no error on the 2 presentations of the second problem, and make 1 error and 1 correct on the third choice problem. If the true pattern were 212, then this response pattern could occur if the person made only 1 error in the second replicate on the third choice problem. Note that when the true preference is assumed to be 212, the first and third choice problems involve error terms,  $f_1$  and  $f_3$ . Each ‘expected’ frequency is simply  $n$  times the theoretical probability, where  $n$  is the total count of responses. There are 64 equations (including this one) for the 64 possible response patterns.

The TE2 model (in which  $e_i = f_i$ ) is a special case of this TE4 model; TE2 uses 3 fewer parameters, so these models can be compared by computing the difference in fit between the 2 models,  $G(3) = G(53) - G(50)$ . In theory, the difference asymptotically approaches a chi-squared distribution with 3 df as  $n$  increases.

Within either TE2 or TE4, transitivity is a special case in which  $p_{111} = p_{222} = 0$ . The differences in fit testing transitivity each have 2 df.

<sup>6</sup>The same notation can be used to refer to true preference patterns as to expressed (observed) preferences, but even with 64 response patterns in 2 replications, there are only 8 true preference patterns because it is assumed that people maintain the same true preferences in both replications within a brief session. In individual TE theory,  $i$ TET, it is allowed that a person might change true preferences between sessions, but it is still assumed that 2 replications within a session are governed by the same true preferences.

The nesting relationships among these 4 models are the same as in Birnbaum (2019, Figure 2), except where ‘transitivity’ is substituted for ‘EU’. In particular, it is possible that data that would refute transitivity with TE2 might be compatible with transitivity under TE4.

A spreadsheet employing Excel’s Solver to fit either TE2 or TE4 to a replicated test of transitivity is included in the Supplementary Material to this article.<sup>7</sup>

## 2. Method

Participants made choices between pairs of gambles, each of which had 3 equally likely prizes. The prize of a gamble is determined by the color of marble drawn blindly and randomly from a single urn containing an equal number of red, white, and blue marbles.

Choices were displayed as in Figure 1, where rows represented the 2 choice alternatives, and columns (colored red, white, and blue) represented the 3 possible outcomes. Numerical entries indicated money prizes to be won if a marble drawn randomly from an urn was red, white, or blue. The urn contained 33 red, 33 white, and 33 blue marbles. These displays are like those in Birnbaum (2023a) and Birnbaum and Diecidue (2015).

### 2.1. Instructions and materials

Instructions, and one block of trials can be viewed at the following URL:

[https://konstanzworkshop.neocities.org/CSUF22/MPW\\_01a](https://konstanzworkshop.neocities.org/CSUF22/MPW_01a).

Stimulus displays and Web forms were constructed and randomized using a free program, available at the following URL:

<http://psych.fullerton.edu/mbirnbaum/programs/ChoiceTableColorWiz2.htm>.

### 2.2. Transitivity design

There were 4 triples of gambles, which showed the highest incidence of intransitive behavior in Butler and Pogrebná (2018) and used in Birnbaum (2023a). The amounts (in dollars, instead of pounds) 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:  $X = (14, 14, 2)$ ;  $Y = (8, 8, 8)$ ;  $Z = (21, 6, 6)$ .

Note that in all 4 triples, Y is always a ‘sure thing’ with the smallest EV; Z always has the highest EV and greatest range; and X is intermediate in both EV and range. The MPW model favors X over Y, Y over Z, and Z over X in all 4 triples.

There are 6 choice problems 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 4 triples and 6 choice problems per triple, there are 24 experimental choice problems testing transitivity in each block.

### 2.3. Dominance design

Four tests of transparent dominance were included in each block: D1:  $T = (10, 9, 8)$  versus  $U = (8, 8, 8)$ ; D2:  $V = (10, 10, 7)$  versus  $W = (12, 12, 8)$ ; D3:  $A = (4, 4, 2)$  versus  $B = (8, 8, 8)$ ,

<sup>7</sup>Other software to analyze the TE models is available from the supplements to Birnbaum et al. (2016), Birnbaum and Quispe-Torresblanca (2018), and Birnbaum and Wan (2020). Bayesian methods for analyzing TE models have been implemented by Lee (2018) and Schramm (2020); see also Birnbaum (2019).

D4:  $C = (8, 8, 8)$  versus  $D = (5, 5, 2)$ . Note that in Choices D1 and D2, the wider range gamble is dominant, whereas in D3 and D4, the wider range gamble is dominated.

## 2.4. Procedure

Each block consisted of 28 randomly ordered trials (choice problems) intermixed from both subdesigns.

When a block was completed, the participant pushed a button to submit the responses, and then pressed another button to load the materials for the second block. Participants participated via the WWW, and worked at their own paces.

Instructions stated that 3 participants would be selected at random to receive the prize of one of their chosen gambles, so they should choose wisely. Procedures for determining prizes were similar to those in Birnbaum and Diecidue (2015, Experiment 6), except contestants were not present and prizes were sent as cash in the mail.

## 2.5. Participants

The participants were 260 undergraduates who completed 2 blocks (ages 18–38, with  $76\% \leq 20$  years; 24% males, 72% females, and 3% not responding) who participated in partial fulfillment of an assignment in Introductory Psychology.

Because each of the 12 choice problems testing transitivity was replicated twice in each block with display position (First or Second) counterbalanced, the number of consistent responses within a block could range from 0 to 12; a person who mindlessly pushed the same button would show zero consistency, and a person who pushed buttons randomly would be expected to have a score of 6 (50% agreement). There were 40 participants who had low consistency in at least one block whose data were analyzed separately, leaving 220 participants with mean consistency of 84% within-block agreement between replicates of the 12 choices testing transitivity.

## 3. Results

### 3.1. Tests of dominance

The first 4 rows of Table 3 show the numbers of participants (out of 220) who had each response pattern in 2 replications of the 2 tests of dominance. For example, the first row (for choice problem D1) shows that 145 participants (out of 220) satisfied dominance on both replications (AA); 22, 19, and 34 violated it only on the second replication (AB), only on the first (BA), or violated dominance on both replications (BB), respectively. There were thus  $22+19+(2)(34) = 109$  total violations out of 440 choice responses, or 25%, which is shown in the column labeled ‘Resp viol’, representing response violations.

The column in Table 3 labeled ‘ $\chi^2(1)Indep$ ’ is the standard chi-squared test of response independence, which is the assumption that the probability of a conjunction of responses is the product of the marginal probabilities of those responses; that is,  $p(AA) = p(A)p(A)$ . Because the critical value of  $\chi^2(1)$  is 6.34 with  $\alpha = 0.01$ , we can reject the hypothesis that responses are independent in the first 4 rows of Table 3.

Independence is implied to be violated by the TE models, except in special cases, such as when  $p_A = 1$ , which implies  $p(A) = 1 - e$ , so  $p(AA) = (1 - e)(1 - e) = p(A)p(A)$ . Note also that if  $p_A = 1$ ,  $p(BB) = e^2 < p(AB) = p(BA) = e(1 - e)$ ; that is, the probability of a repeated violation should be less than that of a reversal of preference between violation and satisfaction of dominance. The large  $\chi^2$  for violation of independence occur when the frequency of BB exceeds that of AB and BA, indicating  $p_A < 1$ .

The next column in Table 3, labeled ‘ $G(1)TE2$ ’, displays the  $G$  tests of the TE2 model ( $e = f$ ), which has 1 df and is asymptotically chi-squared distributed. As can be seen in the table, all of these  $G(1)$  values are much less than corresponding tests of response independence and all show good fits to the

**Table 3.** Response patterns in replicated tests of dominance and TE analysis.

Test (n)	AA	AB	BA	BB	Resp viols	$\chi^2(1)$ Indep	G(1) TE2	Est. $1 - p_A$	Est. $e$	G(2) $p_A \equiv 1$	DIFF
D1 (220)	145	22	19	34	0.25	55.10	0.22	0.18	0.10	50.39	50.17
D2 (220)	160	22	14	24	0.19	49.58	1.79	0.12	0.09	42.34	40.55
D3 (220)	189	10	7	14	0.10	74.26	0.53	0.07	0.04	45.81	45.28
D4 (220)	188	13	11	8	0.09	25.53	0.17	0.04	0.06	16.50	16.33
D1 (89)	36	14	9	30	0.47	20.98	1.10	0.45	0.15	22.50	21.40
D2 (89)	48	13	8	20	0.34	20.66	1.20	0.28	0.14	21.29	20.09
D3 (89)	69	5	5	10	0.17	31.94	0.00	0.12	0.06	25.04	25.04
D4 (89)	70	7	7	5	0.13	9.44	0.00	0.06	0.09	7.18	7.18
D1 (131)	109	8	10	4	0.10	7.10	0.22	0.03	0.07	5.29	5.07
D2 (131)	112	9	6	4	0.09	10.96	0.60	0.03	0.06	7.61	7.01
D3 (131)	120	5	2	4	0.06	35.14	1.33	0.03	0.03	16.59	15.26
D4 (131)	118	6	4	3	0.06	14.97	0.40	0.02	0.04	8.23	7.83

Note: Critical  $\chi^2(1) = 6.64$  and  $\chi^2(2) = 9.21$  for  $\alpha = 0.01$ .

data, even though the number of parameters estimated (and thus the df) is the same as for the tests of independence. The best-fit parameter estimates of the TE2 model are shown in the next columns, labeled ‘ $1 - p_A$ ’ and ‘ $e$ ’, representing the probability of systematic violation of dominance and of random error, respectively.

The next to last column in Table 3 is a  $G(2)$  test of the TE model with the assumption that  $p_A = 1$ ; this test is the same in either TE2 or TE4 because when  $p_A = 1$ , the parameter  $f$  drops out of the equations. The last column in Table 3, labeled ‘DIFF’, shows the differences,  $G(2) - G(1)$ , which test if  $p_A = 1$ . These indicate that the violations of transparent dominance, estimated to range from 4% to 18% in the 4 choice problems of the first 4 rows, are significantly greater than 0, because all 4 values well exceed the critical value of 6.64 with  $\alpha = 0.01$ .

Participants were subdivided according to their modal response patterns in the tests of transitivity. There were 89 participants who most often preferred ‘sure things’ when presented in these non-dominated choice problems, 77 of whom most often preferred the lower range option in any non-dominated choice. That left 131 who did not show one of these modal response patterns.

These two subgroups of data were then analyzed separately, shown by the next 2 groups of 4 rows each. For the group of 89 who prefer ‘sure things’, the estimated true rates of violation were 0.45 and 0.28 for D1 and D2, corrected for error, which are similar to the findings in Birnbaum (2023a) and seem ‘large’ for tests of transparent dominance. The estimated violation rates for the other 131 participants are only around 0.03 for the 4 problems. These data for D1 and D2 therefore are consistent with Birnbaum’s (2023a) findings that relatively large rates of violation of transparent dominance can be observed, and that these high violations are mainly found in people who appear to favor ‘safer’ (i.e., lower-ranged) gambles in non-dominated choice problems.<sup>8</sup>

Among the 89 participants who favor ‘safer’ gambles in non-dominated choices, the rates of violation of dominance for D3 and D4 are lower, 0.12 and 0.06, respectively, compared to 0.45 and 0.28 for D1 and D2. Although D3 and D4 have lower rates, as hypothesized, 12% violations for D3

<sup>8</sup>The  $\chi^2$  and  $G$  tests in the last 6 rows of Table 3 are based on cases where one or more expected cell frequencies are  $\leq 5$ ; these statistics might therefore be considered dubious to compare with the theoretical chi-squared distribution. Nevertheless, Fisher exact tests were also significant in the last 6 rows, with  $p < 0.01$  in all cases except D1(131) where  $p = 0.025$ , so even these small rates of 2%–3% estimated violations might be due to more than chance.

**Table 4.** Crosstabulation. Frequencies of response patterns in first (rows) and second (columns) replications, aggregated over participants, choice triples, and sessions.

Rep 1	Response pattern in Replicate 2								Sum
	111	112	121	122	211	212	221	222	
111	15	23	7	11	6	5	3	1	71
112	15	200	6	33	4	44	1	10	313
121	10	4	248	44	2	2	15	5	330
122	5	38	41	143	1	8	4	14	254
211	8	5	5	1	61	35	10	3	128
212	9	62	0	9	44	352	5	17	498
221	3	1	23	3	6	5	35	6	82
222	2	13	4	7	4	18	10	26	84
Sum	67	346	334	251	128	469	83	82	1,760
Modal	6	40	46	29	12	77	8	2	220

Note: Total  $n = 1,760 = 220$  Participants by 4 Triples by 2 Sessions, each based on 6 responses (3 choices problems by 2 repetitions) per triple. The pattern 111 is the intransitive pattern predicted by most probable winner rule.

still seems surprisingly high for this risk-averse group, since choosing the ‘safer’ alternative in D3 or D4 would have satisfied dominance.<sup>9</sup>

For the group of 131, the rates of violation for D3 and D4 are low in all tests (3% or lower) and the discrepancy (comparing D3 and D4 to D1 and D2) is not as pronounced as for the group of 89.

### 3.2. Tests of transitivity

Table 4 shows a cross-tabulation of response patterns in 2 replicates, aggregated over participants, choice triples, and blocks. Each response pattern represents 6 responses (3 choice problems with 2 replications). For example, the entry of 15 in row 111 and column 111 indicates that there were 15 occasions (out of 1,760) in which a participant chose X over Y, Y over Z, and Z over X on both replicates of a choice triple in a block. This response pattern, 111, is the intransitive cycle implied by the MPW model (for all 4 triples).

The diagonal entries in Table 4 represent frequencies of repeating the same response pattern in both replicates, and the off-diagonal entries are cases where at least 1 of the 3 choice problems produced opposite responses in the 2 replicates. The row and column sums (totals) represent frequencies of response patterns in the first and second replications, respectively.

The last row at the bottom of Table 4, labeled ‘Mode’, shows the number of individuals who showed each modal response pattern, aggregated over triples, replicates, and blocks.

The most frequently observed and repeated pattern is 212, which is the transitive pattern implied by the TAX model with its ‘prior’ parameters. Those parameters were roughly estimated in 1996 (described in Birnbaum and Chavez, 1997) and used to make predictions for new studies for the next quarter century (Birnbaum, 1997, 2008, 2020, 2023a). The 212 pattern was also the most frequently

<sup>9</sup>The estimated error rates are also larger for the group of 89 who preferred ‘safer’ gambles. This result leads one to ask: suppose we selected by within-session consistency; would people with higher consistency have a lower rate of estimated true violations? The answer is no. Participants were selected who achieved at least 71%, 79%, or 88% consistency to define 3 subgroups. The TE2 model was then fit in each subgroup, as in Table 3. It was found that the estimated average rate of violations for D1 and D2 increased from 38% to 42%–51% as self-consistency increased, and violation rates for D3 and D4 decreased slightly (from 10% to 9.5%–9%, respectively). Thus, there is no evidence that more reliable people would be more likely to satisfy dominance in D1 and D2.

**Table 5.** Indices of fit for four models of Table 4.

TE model	Transitivity		Difference $G(2)$
	Trans	Free	
TE2	145.9	85.9	60.0
TE4	131.7	74.7	57.0
Diff $G(3)$	14.2	11.2	

Note: ‘Trans’ indicates transitive; ‘Free’ includes all transitive and intransitive patterns.

observed modal pattern by individuals. This pattern was also the pattern estimated to be most frequent in each Triple when separately analyzed, and was most frequent in previous studies by Butler and Pogrebna (2018) and by Birnbaum (2023a). It is also the response pattern consistent with preference for ‘safer’ (lower range) gambles, despite the lower EV.

Table 4 shows that second most frequently repeated pattern was the transitive pattern, 121, which is the response pattern implied if people chose the gamble with the higher expected value (EV). Other frequently observed patterns include 112, consistent with selecting the gamble with the higher median and 122, which could be compatible with a ‘sufficing’ interpretation (Birnbaum, 2020, 2023a). The least frequent response patterns are the 2 intransitive patterns, 111 and 222. Although 8 of 220 participants showed one of these intransitive patterns as their most frequent response pattern (3.6%), only 3 of these 8 showed one of these patterns more often than 5 times out of 16 possible (4 triples by 2 blocks by 2 replicates).

Table 5 shows the fit of TE models to the data in Table 4. The 64 entries in Table 4 have 63 df. The TE4 model with all parameters free uses 8 estimates of probabilities of the preference patterns ( $p_{111}, p_{112}, \dots, p_{222}$ ) and 6 error terms ( $e_1, e_2, e_3, f_1, f_2, f_3$ ), which consume 13 df, leaving 50 df to test the model. The TE2 model is a special case of TE4 in which  $e_i = f_i$ , so it has 3 additional df. Within each version of the TE model, the transitive special cases assume  $p_{111} = p_{222} = 0$ , and thus leave an additional 2 df. Assuming large  $n$  and assuming that responses across sessions by the same person were independent, the  $G$  should follow a chi-squared distribution with 50–55 df for these 4 models.

The difference in fit between TE2 and TE4 has 3 df, and the difference between each TE model and its transitive special case has 2 df. However, the  $G$  tests are likely inflated by small frequencies in many cells and the fact that each participant served in 2 blocks and 4 choice triples; therefore, some caution should be exercised when comparing these  $G$  values against the theoretical chi-squared distribution. If we take these tests in Table 5 at face value, caveats notwithstanding, both tests of transitivity far exceed the critical  $\chi^2(2) = 9.2$  for  $\alpha = 0.01$ , indicating that transitivity can be rejected, whereas the test of TE4 against TE2 exceeds the critical  $\chi^2(3) = 11.3$  with  $\alpha = 0.01$  in the transitive case, but it falls just short of the critical value for the TE model with all parameters free. Examining the predictions of TE4 and TE2 against the empirical values in Table 4, the improvement of TE4 over TE2 is not impressive. However, even if TE4 does not provide a substantially better fit to data compared to TE2, it is important to evaluate TE4 for both transitive and intransitive cases, because of its theoretical potential to yield a transitive solution that might be compatible with data. In this case, though, it did not change the conclusions regarding transitivity.

Table 6 shows the best-fit parameters of 4 TE models fit to the data in Table 4. To save space in the table, estimated probabilities are displayed as percentages; for example, 09 indicates 0.09. According to either TE2 or TE4, the sum of probabilities of the 2 intransitive preference patterns is only 3%. Although the incidence of intransitive behavior appears minimal, it is statistically significant; even if we adopted the conservative procedure of dividing each  $G$  difference in Table 5 by 8 (4 triples by

**Table 6.** Parameter estimates for true and error models fit to Table 4.

Model	$e_1$	$e_2$	$e_3$	$f_1$	$f_2$	$f_3$	$p_{111}$	$p_{112}$	$p_{121}$	$p_{122}$	$p_{211}$	$p_{212}$	$p_{221}$	$p_{222}$
TE2 Trans	09	09	10	(09)	(09)	(10)	(0)	20	22	15	06	34	04	(0)
TE2 Free	09	07	10	(09)	(07)	(10)	01	20	22	14	06	33	03	02
TE4 Trans	07	11	13	12	05	09	(0)	20	22	12	07	35	04	(0)
TE4 Free	06	09	12	12	06	09	01	19	22	12	07	34	04	02

Note: Values displayed as percentages, so 09 indicates 0.09. Parameters shown in parentheses are fixed or constrained.

2 blocks per participant), we can see that the  $G$  test would indicate ‘significant’ deviations from the transitive models, since the critical value of  $\chi^2(2) = 5.99$  with  $\alpha = 0.05$ . A further analysis of this statistical issue is presented for each block and each triple in the Appendix.

Even if it is statistically significant, the estimate is only 3% intransitive behavior. For comparison, the data of Butler and Pogrebná (2018) had 18% violations, averaged over 11 triples used in that study (44% violations in the 4 triples selected with the highest levels of violation in that study, which obviously includes capitalization on chance). Birnbaum (2023a) found 14% violations for those same 4 selected triples in a study with 30 sessions. Thus, the finding here of only 3% is substantially lower than observed for the same triples in previous, similar studies.

The same analyses were conducted for each triple of choice problems separately. In all 4 triples, the estimate of intransitive preferences under either TE2 or TE4 was 4% or less. Triple 4 showed the largest evidence of intransitive preferences, with estimated 1% of 111 and 3% of 222 intransitive preference patterns. The  $G(50)$  for the TE4 model was 54.01 with all parameters free, and it was 80.4, when it was assumed that  $p_{111} = p_{222} = 0$ , yielding  $G(2) = 26.4$ , more than double the critical value required for statistical significance ( $p < 0.01$ ). The TE2 model fit nearly as well as TE4,  $G(53) = 55.3$ , which is not significantly worse than the fit of TE4. The transitive special case of TE2 had  $G(55) = 82.62$ .

Additional details of the separate analyses for each triple are included in the Appendix and Supplementary Material; overall, the results of separate analysis of the triples appear consistent with the conclusions that the transitive 212 pattern was most frequent for all triples, that intransitive behavior was probably ‘real’ but minimal, that the TE2 model fits almost as well as TE4. Consistent with simulations by Birnbaum and Quan (2020), the choice of TE model for these real data did not appear to affect conclusions regarding transitivity.

The data for the 40 unreliable participants were analyzed separately. Aggregating over triples, as in Table 4, there were only 6 instances (out of 320 possible occasions) in which a participant repeated an intransitive response pattern in a block. The TE4 and TE2 models estimated the probability of intransitive response patterns to be 0.003 and 0.0, respectively, with 0.97 or more of the probability concentrated in 3 transitive response patterns, 121, 122, and 212 in both solutions. The estimated error rates ranged from 0.25 to 0.50 for TE4 and from 0.37 to 0.39 for TE2. Thus, although the error rates were higher for these participants, there was no evidence of systematic violation of transitivity in this group.

#### 4. Discussion

The tests of transparent dominance corroborate 2 findings observed in Birnbaum (2023a) and reveal a new result: first, violations of transparent dominance, while infrequent, are statistically significant when analyzed for all participants; second, violations of dominance are mainly found among people who systematically preferred the ‘safer’ gambles in other choice problems and are very rare for other participants; third, the new tests show that violations are much less frequent when the ‘safe’ gamble is dominant rather than when the ‘safe’ gamble is dominated by the ‘risky’ gamble.

#### 4.1. Why violate transparent dominance?

I suspect that these violations of transparent dominance may be the result of the experimental design in which almost every trial paired a ‘safe’ gamble against a ‘riskier’ one with higher expected value. Brunswik (1956) theorized that humans adapt their behavior to the environment; in the environment of an experiment, stimulus designs will affect the behavior that the experiment is designed to explore. Brunswik thought that between-subjects designs totally isolate the participant from the variation and co-variation of the important variables. He found systematic factorial designs an improvement but thought that these must induce zero correlations among factors, and so he favored hybrid design, in which a factorial design is modified to remove unrepresentative combinations of variables. Ultimately, he argued against systematic design in favor of representative design, in which stimuli and situations are sampled from the environment to which generalization was desired.

Birnbaum (1975) reviewed several studies using *systextual design*, in which the context is systematically manipulated; these studies show that the direction of a main effect of a variable can even be reversed by systematically changing the correlation between that variable and the stimulus to be judged. Mellers et al. (1992a) showed that by systematically changing the selection of stimuli presented in a study, one could change the rank order of judgments of a set of common stimuli (preference order), consistent with the theory that the experimental context caused a change in the model or process by which the information was combined.

Thus, it is not implausible that the Butler and Pogrebna (2018) design, in which every choice fit the same recipe, may induce a decision strategy in some risk-averse participants to seek the low-range option without contrasting outcomes between gambles, a strategy that would lead them to violate dominance on trials in which the ‘safe’ gamble was dominated by the alternative.

Choosing gambles with lower range, without comparing consequences between gambles, would perhaps be an easy way to find ‘safe’ gambles in the Butler and Pogrebna design, but can induce people to violate dominance, if a choice problem is included in which the ‘safe’ gamble is dominated by the ‘risky’ gamble. This strategy for this experimental design would produce the 212 response pattern in tests of transitivity.

How would a person detect and satisfy dominance in a choice problem such as in Figure 1? If one compared the consequences on each branch (for each color of marble) of the gamble, one could detect dominance by counting the number of branches that are equal or favor one gamble over the other. Then choose the alternative with the greater number of winning branches. Ironically, such a strategy that can detect and satisfy dominance could also lead to violations of transitivity. If one chooses the gamble with the greater number of contrasts that favor one gamble over the other, one would violate transitivity in the design testing transitivity here with the response pattern, 111. With equally likely branches, this strategy is equivalent to the MPW model (aka, ‘majority rule’) that was used to design the studies of Birnbaum and Diecidue (2015) and of Butler and Pogrebna (2018).

#### 4.2. Few violations of transitivity

Although Birnbaum (2023a) and Butler and Pogrebna (2018) found evidence of 18% and 14% systematic violations of transitivity, the present data contained only very small incidence of intransitive behavior, about 3%, including about 1% of the MPW type and 2% of the opposite (‘regret’) type. These estimates were similar whether the TE2 model (used in previous analyses) or the TE4 model was used to provide these estimates while correcting for error.

Although the TE4 model fits the data only slightly better than TE2, the main reason to use this analysis is to find out whether or not the conclusions regarding violations of transitivity might be altered when 2 error rates are allowed for each choice problem instead of 1. In this case, the estimates of intransitive behavior are virtually the same for both TE2 and TE4, so one would draw the same conclusions about transitivity from either analysis.

Because the rates of violation are lower in this study than in 2 previous studies with similar design analyzed via TE models, it seems worthwhile to review some possible reasons that may be responsible for the difference in results.

Differences between the methods of Birnbaum (2023a) and Butler and Pogrebnina (2018) were discussed in Birnbaum (2023a). Butler and Pogrebnina (2018) used choices between independent gambles, whereas Birnbaum (2023a) and this study used fully dependent gambles. If anything, use of dependent gambles should enhance intransitive behavior, because it is easier to determine the MPW with dependent than with independent gambles, where one would have to work out the probabilities of the combinations of outcomes possible under independent plays of the 2 gambles. Similarly, it is easier to conceive of regrets when the outcomes are clearly dependent on a common event.

The instructions used here and in Birnbaum (2023a) called attention to contrasts in value between consequences for events, which was also intended to promote any tendencies to use MPW or a regret strategy, both of which can lead to intransitive cycles. So, one would have reasoned before doing the study that such instructions would increase tendencies to use strategies that can be intransitive. Post hoc, one might speculate that instructions may have done something in the opposite direction, but it is hard to construct a coherent theoretical argument.

In this study, participants with low self-consistency were separated from the main analysis. When they are analyzed separately, the unreliable participants showed no evidence of intransitive behavior, and when they are combined with the main group, the estimated incidence of intransitive behavior is decreased by a trivial amount (still rounds to 3%), so no conclusions are altered by inclusion of the unreliable participants (see the Supplementary Material).

The stimulus materials, instructions, and participant recruitment were the same here as used in Birnbaum (2023a), so the lower incidence found here compared to that study is unlikely due to these experimental features.

An important difference between this study and that of Birnbaum (2023a) is that Birnbaum (2023a) had many blocks of trials, whereas this study had only 2. The intransitive behavior in Birnbaum (2023a) emerged in a number of participants only after many sessions, so the shorter experiment here may not have allowed as much intransitive behavior to develop as was found in that study. However, this study is comparable in length to that of Butler and Pogrebnina (2018), which found the highest rate of violation of the 3 studies.

A second difference with Birnbaum (2023a) is in the experimental designs: the additional tests of dominance included in this study (not present in Birnbaum, 2023a) may have reduced the incidence of intransitive choices; indeed, Choices D3 and D4 depart from the confound of EV and range used in both previous studies that had higher reported incidence.

A third difference is that Birnbaum (2023a) used fewer participants: perhaps the incidences of 3% violations observed in a sample of 220 is closer to the population value than estimate of 14% in Birnbaum (2023a) with only 22 participants. Whatever the reason for the smaller incidence of intransitive behavior compared to two previous studies, it is worth keeping in mind that very few studies using other designs have claimed to find more than modest incidences of violation of transparent dominance or transitivity.

This study had small financial incentives. It is traditional to theorize that increasing or decreasing financial incentives might increase or decrease either error rates or true violations of 'rational' principles like transitivity or dominance. Because systematic intransitive behavior would allow a person to become a 'money pump', it is often argued that a rational person should become more transitive when larger prizes are available. In contrast, one can also argue that a person might find the strategy to select the MPW to seem the 'right thing' to do when large prizes are involved (Butler and Blavatsky, 2020; Butler and Pogrebnina, 2018).

Large prizes might induce people to be careful and thus reduce error rates, or they might induce excitement, which might increase error rates. It could be argued that the effect of financial incentives might be non-monotonic in increasing or decreasing error rates or violations of 'rational' principles. As far as I am aware, no proper study has yet been done in which error rates have been appropriately

separated from true violation rates to systematically examine the effects of financial incentives on both true preferences and on error rates.

### 5. Concluding comments

In summary, we can draw the following ‘big picture’ view of the results: small incidences of violations of transitivity and dominance can be detected and estimated using TE models. They may be real, but perhaps strongly influenced by the context provided by the stimuli presented in the experiment. Violations of transitivity or dominance observed in specially designed studies may be real, but second-order effects, analogous to friction in physics labs. Systematic violations of basic equations in science are often observed, and basic equations often fit better in a vacuum, in low gravity, with low temperatures, or with other special conditions. Thus, these small incidence violations may be real but not indicative of the principles by which people generally make decisions. In contrast, other paradoxes of choice, such as the Allais paradoxes and violations of stochastic dominance have proven to occur with large incidence, and to be robust to variation of methods and resistant to attempts to eliminate them by changing features of the experiments (Birnbaum, 1999, 2004, 2005, 2008) or even by training designed to help people satisfy first-order stochastic dominance (Quispe-Torreblanca et al., 2022). Such phenomena seem to be ones that a theoretician must be able to account for in any viable descriptive theory of behavior.

It might be comforting to some that violations of principles like dominance or transitivity are relatively rare by college undergraduates (4%–18% violations of dominance and 3%–4% violations of transitivity). But these violations, even if rare, will remain disconcerting to theoreticians who would hope to use the same model to describe all people. If one theorizes models in which a person computes a utility for a gamble and compares utilities, violations might mean that at least some of the participants in our studies are using models or strategies that are not compatible with such a process, so that a complete description of behavior may have to concede that different people are using different models in the same study or even that the same person may use different strategies or models in different experimental contexts.

### Appendix: A note on statistical significance

The TE models indicate that the amount of intransitive behavior, averaged over triples, blocks, and participants, is about 3%. This is a very small figure, and it seems reasonable for a skeptic to ask: what exactly in the data indicates that there is more intransitive behavior than ‘chance’, aside from the finding that the statistical index of the transitive model is ‘significantly’ worse than the fit of the TE model that allows intransitive preference patterns?

One way to address this question is to compute how often we expect to see intransitive response patterns repeated and compare that expectation against the observed frequency. How often do we expect to see someone show the 111 or 222 pattern on 2 replications of the same choice triple? In order to repeat any response pattern in this study, the participant must use opposite buttons on randomly ordered trials. The experimental counterbalancing insures that a repeated response pattern cannot be achieved by mindlessly pushing the same button, which would produce opposite patterns between 2 replications.

A simple calculation illustrates how to calculate predictions under a null hypothesis. Suppose there were only one true preference pattern, 212, for a particular triple ( $p_{212} = 1$ ); that is,  $Y = (10, 10, 10) > X = (15, 15, 3)$ ,  $Y = (10, 10, 10) > Z = (27, 5, 5)$ , and  $X = (15, 15, 3) > Z = (27, 5, 5)$ . Suppose also that the error terms are  $e_i = f_i = 0.10$  for all  $i$ . It follows that the probability to show the 222 pattern is as follows:

$$p(222, 222) = p_{212}(1 - e)^2(e)^2(1 - e)^2 = (1)(.9)^2(.1)^2(.9)^2 = 0.00656, \tag{5}$$

**Table 7.** Predicted probabilities of repeated violations of transitivity and frequencies of obtained violations.

Triple	Block	Model		Pattern		Observed
		TE2	TE4	111	222	Sum
1	1	0.0059	0.0059	0	5	5*
1	2			4	0	4*
2	1	0.0046	0.0046	2	2	4*
2	2			2	1	3
3	1	0.0062	0.0065	1	5	6*
3	2			2	2	4†
4	1	0.0060	0.0065	2	4	6*
4	2			2	7	9*

Note: Total  $n = 220$  per Block.

\*Indicates binomial  $p < 0.05$ .

†Triple 3 Rep 2 is significant for TE2 only.

where  $p(222, 222)$  is the probability to repeat the 222 pattern in both replications, given that 212 is the only true preference pattern,  $p_{212} = 1$ . Similarly, the probability to repeat the 111 pattern is  $e^2(1 - e)^2$   $e^2 = 0.00008$ . The probability of the union of both intransitive patterns is then 0.00664. In a single block with 2 replications with 220 independent participants, we expect to see only 1.46 repeated patterns, on average. The binomial probability that 5 or more people (out of 220) would have a repeated intransitive pattern is 0.016. So, if we observed 5 or more, we could reject the null hypothesis with  $p < 0.05$ . Five out of 220 is less than 3% of the sample. As will be shown, even when we fit the full TE models to the data, the numbers will work out close to this example case.

When either TE2 or TE4 is fit separately to a Triple's data with the assumption that  $p_{111} = p_{222} = 0$ , we can use the parameter estimates to calculate  $p(111) + p(222)$ , the predicted frequency of repeated violations of transitivity. These can be compared to the actual number of violations by means of the binomial distribution for each Triple and block.

Table 7 shows the probabilities calculated under the null hypothesis for  $p(111) + p(222)$  under the TE2 and TE4 models. The observed number of replicated violations of 111 and 222 patterns is shown, along with their sum. The sum has been compared to the binomial distribution, with  $n = 220$ . In every triple and session, the observed number exceeds the expectation, and in all except 2 of 8 cases (Triple 2 Rep 2 and Triple 3 Rep 2), the probability to observe the observed number or greater is less than 0.05. (In the case of Triple 3 Rep 2, 4 is significant according to TE2 but not according to TE4.)

These are very small incidences, but they are significant. There is an argument that could be made to discount statistical significance: if the estimates of the error terms are too small, then the significance level is also too small. Perhaps these error estimates could be underestimated because unreliable participants were excluded.

I think that this analysis illustrates issues inherent in research that presents only a statistical test of significance without estimating incidence or magnitude of an effect. An effect can be small and also statistically significant.

In my opinion, a 3% estimate of violations of transitivity may be statistically 'significant' but not represent an effect that a general theory of decision making should be built upon.

**Supplementary material.** An Excel program to fit TE4 to a test of transitivity is included in the Supplementary Material. The Supplementary Material for this article can be found at <https://doi.org/10.1017/jdm.2023.43>.

**Data availability statement.** The data are available in the Supplementary Material to this article, in the same Excel file that contains a README sheet describing the data, supplementary analyses, and worksheets containing programs implementing the Solver to fit TE models.

**Acknowledgments.** Thanks are due to Julien Rouvere for assistance and to anonymous reviewers for helpful suggestions.

## References

- Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review*, *124*(5), 678–687. <https://doi.org/10.1037/rev0000073>
- Birnbaum, M. H. (1975). Expectancy and judgment. In F. Restle, R. Shiffrin, N. J. Castellan, H. Lindman, & D. Pisoni (Eds.), *Cognitive theory* (Vol. 1, pp. 107–118). Hillsdale, NJ: Erlbaum. <https://doi.org/10.4324/9780203781548-7>
- Birnbaum, M. H. (1992). Violations of monotonicity and contextual effects in choice-based certainty equivalents. *Psychological Science*, *3*(5), 310–314. <https://doi.org/10.1111/j.1467-9280.1992.tb00679.x>
- Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ: Erlbaum. <https://doi.org/10.4324/9781315789408-6>
- Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, *10*(5), 399–407. <https://doi.org/10.1111/1467-9280.00176>
- Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes*, *95*(1), 40–65. <https://doi.org/10.1016/j.obhdp.2004.05.004>
- Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, *31*(3), 263–287. <https://doi.org/10.1007/s11166-005-5103-9>
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501. <https://doi.org/10.1037/0033-295x.115.2.463>
- 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. <https://doi.org/10.1016/j.jmp.2010.03.002>
- Birnbaum, M. H. (2013). True-and-error models violate independence and yet they are testable. *Judgment and Decision Making*, *8*, 717–737. <http://journal.sjdm.org/13/13422c/jdm13422c.pdf>
- Birnbaum, M. H. (2019). Bayesian and frequentist analysis of true and error models. *Judgment and Decision Making*, *14*(5), 608–616. <http://www.sjdm.org/journal/19/190822/jdm190822.pdf>
- Birnbaum, M. H. (2020). Reanalysis of Butler and Pogrebna (2018) using true and error model. *Judgment and Decision Making*, *15*(6), 1044–1051. <http://journal.sjdm.org/20/200216/jdm200216.pdf>
- Birnbaum, M. H. (2023a). Testing transitivity of preference in individuals. *Decision*, *10*(2), 153–180. <https://doi.org/10.1037/dec0000185>
- Birnbaum, M. H. (2023b). True and error analysis instead of test of correlated proportions: Can we save lexicographic semiorder models with error theory? *Psychological Methods*, *28*(5), 1087–1099. <https://doi.org/10.1037/met0000557>
- 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. <http://journal.sjdm.org/12/12315/jdm12315.pdf>
- Birnbaum, M. H., & Bahra, J. P. (2012b). Testing transitivity of preferences in individuals using linked designs. *Judgment and Decision Making*, *7*, 524–567. <http://journal.sjdm.org/11/111122/jdm111122.pdf>
- Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, *71*(2), 161–194. <https://doi.org/10.1006/obhd.1997.2721>
- Birnbaum, M. H., Coffey, G., Mellers, B. A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 331–346. <https://doi.org/10.1037/0096-1523.18.2.331>
- 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. <https://doi.org/10.1037/dec0000031>
- 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. <https://doi.org/10.1016/j.obhdp.2007.02.001>
- Birnbaum, M. H., & Navarrete, J. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty*, *17*(1), 49–78. <https://doi.org/10.1023/a:1007739200913>
- 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. <http://journal.sjdm.org/15/15615a/jdm15615a.pdf>
- 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. <https://sjdm.org/journal/20/200413b/supp.pdf>
- Birnbaum, M. H., & Quispe-Torreblanca, E. G. (2018). TEMAP2.R: True and error model analysis program in R. *Judgment and Decision Making*, *13*(5), 428–440. <http://www.sjdm.org/journal/18/18507/jdm18507.pdf>
- 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. <https://doi.org/10.1007/s11166-008-9043-z>
- Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes*, *52*(2), 183–215. [https://doi.org/10.1016/0749-5978\(92\)90035-6](https://doi.org/10.1016/0749-5978(92)90035-6)
- Birnbaum, M. H., & Thompson, L. A. (1996). Violations of monotonicity in choices between gambles and certain cash. *American Journal of Psychology*, *109*(4), 501–523. <https://doi.org/10.2307/1423391>

- Birnbaum, M. H., & Wan, L. (2020). MARTER: Markov true and error model of drifting parameters. *Judgment and Decision Making*, 15, 47–73. <http://journal.sjdm.org/19/190727/jdm190727.pdf>
- Blavatsky, P. R. (2006). Axiomatization of a preference for most probable winner. *Theory and Decision*, 60, 17–33. <https://doi.org/10.1007/s11238-005-4753-z>
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press. <https://doi.org/10.1525/9780520350519>
- 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. [https://doi.org/10.1016/0749-5978\(87\)90037-9](https://doi.org/10.1016/0749-5978(87)90037-9)
- Butler, D. (2020). Intransitive preferences or choice errors? A reply to Birnbaum. *Judgment and Decision Making*, 15(6), 1052–1053. <http://journal.sjdm.org/20/200216r/jdm200216r.pdf>
- Butler, D. J., & Blavatsky, P. (2020). The voting paradox... with a single voter? Implications for transitivity in choice under risk. *Economics & Philosophy*, 36(1), 61–79. <https://doi.org/10.1017/s026626711900004x>
- Butler, D. J., & Pogrebna, G. (2018). Predictably intransitive preferences. *Judgment and Decision Making*, 13(3), 217–236. <https://sjdm.org/journal/17/17912b/jdm17912b.pdf>
- 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. <https://www.apa.org/pubs/journals/features/dec-0000011.pdf>
- Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4, 113–134. <https://doi.org/10.1007/bf00056121>
- Gonzalez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, 109(1), 137–155. <https://doi.org/10.1037/0033-295x.109.1.137>
- Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences*, 10, 131–153. [https://doi.org/10.1016/0165-4896\(85\)90031-9](https://doi.org/10.1016/0165-4896(85)90031-9)
- Lee, M. D. (2018). Bayesian methods for analyzing true-and-error models. *Judgment and Decision Making*, 13(6), 622–635. <https://sjdm.org/journal/18/18507c/jdm18507c.pdf>
- Leland, J. W. (1998). Similarity judgments in choice under uncertainty: A re-interpretation of the predictions of regret theory. *Management Science*, 44, 659–672. <https://doi.org/10.1287/mnsc.44.5.659>
- Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92, 805–824. <https://doi.org/10.2307/2232669>
- Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410602831>
- Mellers, B. A., Ordóñez, L., & Birnbaum, M. H. (1992a). 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. [https://doi.org/10.1016/0749-5978\(92\)90025-3](https://doi.org/10.1016/0749-5978(92)90025-3)
- Mellers, B. A., Weiss, R., & Birnbaum, M. H. (1992b). Violations of dominance in pricing judgments. *Journal of Risk and Uncertainty*, 5(1), 73–90. <https://doi.org/10.1007/bf00208788>
- Morrison, H. W. (1963). Testable conditions for triads of paired comparison choices. *Psychometrika*, 28, 369–390. <https://doi.org/10.1007/bf02289558>
- 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. <https://doi.org/10.1037/dec0000037>
- Quispe-Torreblanca, E. G., Stewart, N., & Birnbaum, M. H. (2022). Surprisingly robust violations of stochastic dominance despite coalescing training: A quasi-adversarial collaboration. Submitted for publication. <https://psyarxiv.com/c327y/>
- 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. <https://doi.org/10.1037/a0021150>
- 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. <https://doi.org/10.1257/jel.44.3.631>
- Schramm, P. (2020). The individual true and error model: Getting the most out of limited data. *Judgment and Decision Making*, 15(5), 851–860. <https://sjdm.org/journal/19/190516/jdm190516.pdf>
- 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. <https://doi.org/10.1007/bf01075203>
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48. <https://doi.org/10.1037/h0026750>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323. <https://doi.org/10.1007/BF00122574>