

# Testing transitivity of preferences using linked designs

Michael H. Birnbaum\*

Jeffrey P. Bahra†

## Abstract

Three experiments tested if individuals show violations of transitivity in choices between risky gambles in linked designs. The binary gambles varied in the probability to win the higher (better) prize, the value of the higher prize, and the value of the lower prize. Each design varied two factors, with the third fixed. Designs are linked by using the same values in different designs. Linked designs allow one to determine if a lexicographic semiorder model can describe violations of transitivity in more than one design using the same parameters. In addition, two experiments tested interactive independence, a critical property implied by all lexicographic semiorder models. Very few people showed systematic violations of transitivity; only one person out of 136 showed violations of transitivity in two designs that could be linked by a lexicographic semiorder. However, that person violated interactive independence, as did the majority of other participants. Most individuals showed systematic violations of the assumptions of stochastic independence and stationarity of choice responses. That means that investigators should evaluate models with respect to response patterns (response combinations) rather than focusing entirely on choice proportions.

Keywords:

## 1 Introduction

Descriptive theories of risky decision making can be divided into two groups: those that satisfy transitivity of preference and those that do not. Transitivity of preference is the assumption that, if a person prefers  $A$  to  $B$  and prefers  $B$  to  $C$ , then that person should prefer  $A$  to  $C$ , apart from random error. We use the symbol,  $\succ$ , to denote preference, so the property can be denoted as follows:  $A \succ B$  and  $B \succ C \Rightarrow A \succ C$ .

Theories that represent each gamble by a single number automatically imply transitivity. These theories assume that  $A \succ B \Leftrightarrow U(A) > U(B)$ , where  $U(A)$  and  $U(B)$  are the numerical *values* or *utilities* of the two gambles. Expected utility theory (EU), cumulative prospect theory (CPT), and the transfer of attention exchange model (TAX), as well as many other theories, fall in this class of theories that satisfy transitivity (Birnbaum, 2008b; Tversky & Kahneman, 1992; Luce, 2000; Wakker, 2011).

Theories that represent choice in terms of contrasts between the components of the alternatives, however, need not satisfy transitivity of preference. Theories that violate transitivity include the family of lexicographic semiorder (LS) models, the priority heuristic, regret theory (RT),

the stochastic difference model (SDM) and others (Birnbaum, 2010; Birnbaum & Gutierrez, 2007; Birnbaum & Schmidt, 2008; Brandstätter, Hertwig, & Gigerenzer, 2006; González-Vallejo, 2002; Loomes, Starmer, & Sugden, 1991; Luce, 1956; 2000; Myung, Karabatsos, & Iverson, 2005; Regenwetter, Dana, and Davis-Stober, 2010, 2011; Rieskamp, Busemeyer, & Mellers, 2006; Tversky, 1969).

An example of a lexicographic semiorder (LS) is presented next to illustrate how such a model can account for intransitive preferences.

### 1.1 Lexicographic semiorders

Let  $G = (x, p; y)$  represent a two-branch gamble in which prize  $x$  is received with probability  $p$  and otherwise  $y$  is received, where  $x > y \geq 0$ . In such two-branch gambles, there are three variables that can be manipulated experimentally:  $y =$  Lowest (L) consequence;  $x =$  Highest (H) consequence; and  $p =$  Probability (P) to win the higher prize.

We use the notation LPH LS to refer to the lexicographic semiorder (LS) model in which the person is assumed to compare the attributes in the order L, then P, then H. The three attributes might be examined by a participant in any of five other possible orders: LHP, HPL, HLP, PLH, and PHL.

In the LPH LS model, a person is assumed to compare two such gambles,  $G = (x, p; y)$  and  $F = (x', q; y')$  by contrasting attributes with thresholds ( $\Delta_L$  and  $\Delta_P$ ) as follows:

We thank William Batchelder, Daniel Cavagnaro, Geoffrey Iverson, R. Duncan Luce, Michel Regenwetter, and Clinton Davis-Stober, for helpful discussions of issues in this project. This work was supported in part by a grant from the National Science Foundation, SES DRMS-0721126. Experiment 1 is based on a Master's thesis by the second author under supervision of the first.

\*Dept. of Psychology, CSUF H-830M, Box 6846, Fullerton, CA 92834-6846, USA. Email: mbirnbaum@fullerton.edu.

†California State University, Fullerton

1. First compare L: If  $|y - y'| \geq \Delta_L$ , choose the gamble with the higher lowest consequence;
2. Else, compare P: if  $|p - q| \geq \Delta_P$ , choose the gamble with the higher probability to win the better prize;
3. Else, check H: if  $|x - x'| > 0$ , chose the gamble with the higher best prize;
4. Else, choose randomly.

The priority heuristic of Brandstätter, et al. (2006) is a variant of this LPH LS in which it is assumed that  $\Delta_L$  equals one tenth of the highest consequence in either gamble, rounded to the nearest prominent number, where prominent numbers are integer powers of 10 plus one-half and twice those values; i.e., 1, 2, 5, 10, 20, 50, 100, etc. If the highest prize always rounds to \$100 (as in the experiments of this article), then  $\Delta_L = \$10$ . Further, the priority heuristic assumes that  $\Delta_P = 0.10$ , presumably, due to the base 10 number system. Therefore, in these studies the priority heuristic is a special case of the LPH LS model. Brandstätter, et al. (2006) showed that with these selected parameters, this model approximates the results of several previously published papers; in addition, they claimed that the priority heuristic is more accurate than other models for these selected studies.

To illustrate how this LPH LS model can violate transitivity, consider the following five gambles:  $K = (\$100, .50; \$0)$ ,  $L = (\$96, .54; \$0)$ ,  $M = (\$92, .58; \$0)$ ,  $N = (\$88, .62; \$0)$ , and  $O = (\$84, .66; \$0)$ . According to the priority heuristic,  $\Delta_L = \$10$  and  $\Delta_P = 0.10$ , so people should prefer  $K \succ L$ ,  $L \succ M$ ,  $M \succ N$ , and  $N \succ O$ , because the differences in probability are only 0.04; these are too small to be decisive (less than  $\Delta_P = 0.10$ ), so preferences are determined by the highest consequences. However,  $O \succ K$ , because the difference in probability is 0.16, which exceeds the threshold of  $\Delta_P = 0.10$ . As long as  $0.16 \geq \Delta_P > 0.04$ , the LPH LS implies:  $K \succ L$ ,  $L \succ M$ ,  $M \succ N$ , and  $N \succ O$ , but  $O \succ K$ , violating transitivity. When  $\Delta_P = 0.10$ , as in the priority heuristic, then two other violations are also predicted,  $O \succ L$  and  $N \succ K$ . If  $\Delta_P \leq 0.04$ , the LPH LS model implies the transitive order  $ONMLK$ , and if  $\Delta_P > 0.16$ , it predicts the transitive order  $KLMNO$ .

Now consider a second design with choices among the following gambles:  $A = (\$84, 0.5; \$24)$ ,  $B = (\$88, 0.5; \$20)$ ,  $C = (\$92, 0.5; \$16)$ ,  $D = (\$96, 0.5; \$12)$ , and  $E = (\$100, 0.5; \$8)$ . According to the priority heuristic,  $E \succ D$ ,  $D \succ C$ ,  $C \succ B$ , and  $B \succ A$ , because in each of these choices, the lowest consequences differ by less than \$10, and probabilities are equal, so these choices are determined by the highest consequences. However, in the choice between  $A$  and  $E$ , the lowest consequences differ by \$16, which exceeds  $\Delta_L = \$10$ , so  $A \succ E$ , violating transitivity. As long as  $\$16 \geq \Delta_L > \$4$ , the LPH LS implies there should be at least one intransitivity in this design, with  $A \succ E$ . If  $\Delta_L = \$10$ , as in the priority heuristic, then two other violations are also predicted,  $A$

$\succ D$  and  $B \succ E$ . If  $\Delta_L \leq \$4$ , the LPH LS predicts the transitive order,  $ABCDE$ , and if  $\Delta_L > \$16$ , it predicts the transitive order,  $EDCBA$ .<sup>1</sup>

## 1.2 Intransitive preferences in linked designs

The LS models, including the priority heuristic, imply that choices in linked designs will be related. This study uses the two designs described above, and also a third design with choices among the following:  $F = (\$100, 0.5; \$24)$ ,  $G = (\$100, 0.54; \$20)$ ,  $H = (\$100, 0.58; \$16)$ ,  $I = (\$100, 0.62; \$12)$ , and  $J = (\$100, 0.66; \$8)$ . Note that the levels of lowest consequence match those in the design with  $A, B, C, D$ , and  $E$  and that probability values match those in the first design, with  $K, L, M, N$ , and  $O$ . Those two designs are linked in turn by the levels of the highest consequence. These designs with linked levels should show predictable patterns of transitivity or intransitivity, if a person used the same LS model in all three designs.

The stimuli used in these studies with linked levels are listed in Table 1. The designs are named after the variables manipulated: the  $LH$  design varies the lowest consequence (L) and the highest consequence (H), the  $LP$  design varies the lowest consequence and probability (P), and the  $PH$  design varied probability and highest consequence.

For example, suppose that a person conformed to the LPH LS model. If that person showed data consistent with the transitive order  $ONMLK$  in the  $PH$  design, it means that  $\Delta_P \leq 0.04$ ; and suppose that the same person showed intransitive choices in the  $LH$  design consistent with  $\Delta_L = \$10$ . In that case, the model implies intransitive data in the  $LP$  design such that  $J \succ I$ ,  $I \succ H$ ,  $H \succ G$ ,  $G \succ F$  and yet  $F \succ J$ ,  $F \succ I$ , and  $G \succ J$ . So, if results conformed to this prediction, they would represent successful confirmations of new predictions, and if not, the model(s) that predicted them would be disconfirmed. See Appendix A for all possible linked patterns in the LPH LS model.

## 1.3 TAX Model

A transitive model that has been fairly successful in describing violations of EU and CPT is Birnbaum's (1999;

<sup>1</sup>A more general family of LS models can be defined by allowing subjective transformations for prizes and probabilities,  $u(x)$ ,  $u(y)$ , and  $t(p)$ , and assuming that people compare subjective differences against the thresholds, for example as follows:  $|u(y) - u(y')| \geq \Delta_L$ . Because the functions,  $u(x)$  and  $t(p)$  are assumed to be strictly monotonic, this version of LS models also makes the same qualitative predictions concerning the AE choice compared to the set of adjacent choices, but it does not require that  $E \succ D \Leftrightarrow D \succ C \Leftrightarrow C \succ B \Leftrightarrow B \succ A$ , aside from error. The analyses in Tables 3, 4 and 5 and Appendices D, F and G allow for this more general family of LS models.

Table 1: Gambles used in linked tests of transitivity.

<i>LH Design</i>	<i>LP Design</i>	<i>PH Design</i>
A = (\$84, 0.50; \$24)	F = (\$100, 0.50; \$24)	K = (\$100, .50; \$0)
B = (\$88, 0.50; \$20)	G = (\$100, 0.54; \$20)	L = (\$96, .54; \$0)
C = (\$92, 0.50; \$16)	H = (\$100, 0.58; \$16)	M = (\$92, .58; \$0)
D = (\$96, 0.50; \$12)	I = (\$100, 0.62; \$12)	N = (\$88, .62; \$0)
E = (\$100, 0.50; \$8)	J = (\$100, 0.66; \$8)	O = (\$84, .66; \$0)

2008b) special Transfer of Attention Exchange (TAX) model. This model represents the utility of a gamble as a weighted average of the utilities of the consequences, but weight in this model depends on the probabilities of the branch consequences and ranks of the consequences in the gamble. This model can be written for gambles of the form  $G = (x, p; y)$  where  $x > y \geq 0$  as follows:

$$U(G) = \frac{au(x) + bu(y)}{a + b} \tag{1}$$

where  $a = t(p) - \omega t(p)$ ;  $b = t(q) + \omega t(p)$ , where  $q = 1 - p$  and when  $\omega > 0$ . In this case ( $\omega > 0$ ), there is a transfer of attention from the branch leading to the best consequence to the branch leading to the worst consequence. In the case where  $\omega < 0$ , weight is transferred from lower-valued branches to higher ones; in that case,  $a = t(p) - \omega t(q)$ ;  $b = t(q) + \omega t(q)$ . The configural parameter,  $\omega$ , can produce risk aversion ( $\omega > 0$ ) or risk-seeking ( $\omega < 0$ ), even when  $u(x) = x$ . When  $\omega = 0$  and  $t(p) = p$ , this model reduces to expected utility (EU). Expected Value (EV) is a special case of EU in which  $u(x) = x$ .

When fitting individual data in a suitable experiment, parameters can be estimated from the data. However, for the purpose of making predictions before conducting new studies, a simple version of the special TAX model has been used (e.g., Birnbaum, 2008b):  $u(x) = x$  for  $0 < x < \$150$ ;  $t(p) = p^{0.7}$ , and  $\omega = 1/3$ , where  $\omega = \delta/(n + 1)$ ,  $\delta = 1$ , and  $n = 2$  is the number of branches. These have been called “prior” parameters, because they have been used in previous studies to design new experiments and predict modal results with similar participants, contexts, and procedures. Although these are not optimal, they have had reasonable success predicting aggregate results of new studies with American undergraduates who choose among gambles with small gains (e.g., Birnbaum, 2004, 2005, 2008b, 2010).

The TAX model with these prior parameters implies the transitive orders, *ABCDE*, *FGHIJ*, and *ONMLK*. Although the TAX model successfully predicted new results that violated CPT, in the tests of transitivity of Table 1, TAX with these parameters makes virtually the same predictions as CPT with the parameters of Tversky and Kahneman (1992). With other parameters, TAX, CPT, and

EU could account for other transitive orders, but these models always imply transitivity. Experiments 2 and 3 also include tests between TAX and CPT.

Transitivity can therefore be considered a critical property of TAX, CPT and EU because these models (with any parameters) cannot account for systematic violations of transitivity. The family of LS models could handle either transitive or intransitive data, so finding transitive preferences would not refute LS models. For example, the LPH LS model with  $\Delta_P \leq 0.04$  and  $\Delta_L \leq \$4$  makes the same transitive predictions for this study as TAX model with its prior parameters. There are critical properties of LS models, however, that can lead to refutation of those models (Birnbaum, 2008a, 2010; Birnbaum & LaCroix, 2008), described next.

### 1.4 Critical properties of LS models

Birnbaum (2010) considered a general family of LS models in which each person might have a different priority order in which to compare the features; each person might have a different monotonic utility function for monetary prizes and a different subjective function for probability; and each person might have different thresholds for determining if a given subjective difference is decisive. Birnbaum (2010) showed that this general family of LS models implies properties of priority dominance, integrative independence, and interactive independence. In Experiments 2 and 3, we test *interactive independence*, which can be written:

$$F = (x', p; y') \succ G = (x, p; y) \Leftrightarrow F' = (x', q; y') \succ G' = (x, q; y).$$

In these two choice problems, note that  $F$  and  $G$  share a common probability to win ( $p$ ), and  $F'$  and  $G'$  also share a common probability ( $q$ ). According to this most general family of LS models, a person should either prefer  $F$  to  $G$  and  $F'$  to  $G'$  or prefer  $G$  to  $F$  and  $G'$  to  $F'$ , or be indifferent in both cases, but a person should not shift from  $F$  to  $G'$  or from  $G$  to  $F'$  as the common probability is changed, except by random error. To test such a property with real data requires a theory to separate random error from systematic violation.

## 1.5 Testing algebraic properties with probabilistic data

Testing properties such as transitivity or interactive independence is complicated by the fact that people are not completely consistent in their responses. Different people can make different responses when asked the same question, so we must allow for individual differences. Furthermore, the same person might make different responses when the same choice problem is presented on a later trial, following other intervening trials. It is possible that the person has changed her or his “true” preferences, that responses contain “error”, or both. Exactly how to analyze data containing variability has been the topic of debate (Loomes & Sugden, 1995; Birnbaum & Bahra, 2012; Regenwetter, et al., 2011).

Morrison (1963) reviewed two properties that are implied by certain stochastic choice models such as Luce’s (1959) choice model: Weak Stochastic Transitivity (WST) and the Triangle Inequality (TI). These properties were also viewed as methods for analyzing transitive models with variable data. The TI can be written:

$$0 \leq p(A, B) + p(B, C) - p(A, C) \leq 1$$

where  $p(A, B)$  is the probability to choose  $A$  over  $B$ . WST can be written:

$$p(A, B) > \frac{1}{2} \text{ and } p(B, C) > \frac{1}{2} \Rightarrow p(A, C) > \frac{1}{2}$$

Morrison (1963) advised that both of these properties should be tested. Tversky (1969) cited Morrison but reported only tests of WST. Tversky’s statistical tests were challenged by Iverson and Falmagne (1985), who noted that Tversky’s tests did not properly allow for individual differences in preference orders (cf., Myung, et al., 2005).

Regenwetter, et al. (2010, 2011) also criticized Tversky’s failure to test the TI, and proposed statistical tests of these properties based on the assumptions that repeated responses to the same choice problem are independent and identically distributed (iid). They argued that, if each person’s responses can be modeled as an iid sample from a mixture of different transitive preferences (so that choice proportions satisfy the linear order polytope, which includes TI), there would be no reason to argue for LS models.

Birnbaum (2011, 2012) questioned Regenwetter, et al. (2011) for not testing the crucial iid assumptions; when iid assumptions are false, neither WST nor TI (nor any analysis of the linear order polytope defined on choice proportions for individual choice items) can be regarded as unambiguous tests of transitivity.

There are two problems: WST can be violated even when a person has a mixture of strictly transitive orders, and the TI (and the linear order polytope) can be satisfied even when a person has a mixture that includes intransitive patterns (Birnbaum, 2011, 2012). These two properties can be more informative when they agree, but

Birnbaum (2011) argued that we should also examine response patterns in order to ensure that choice proportions also reflect individual behavior that might in fact change systematically during a study.

A debate between Birnbaum (2011, 2012) and Regenwetter, Dana, Davis-Stober, and Guo (2011) has arisen concerning methods for analyzing variability of choice responses. The approach of Regenwetter, et al. (2010, 2011) analyzes only binary choice proportions based on the assumption of iid, whereas the “true and error” (TE) models analyze relative frequencies of response patterns, based on the assumption that errors are independent.

The true and error (TE) model, as applied by Birnbaum and Gutierrez (2007) and Birnbaum and Schmidt (2008) assumes that different people may have a different patterns of true preferences and that different choice problems may have different error rates. Different individuals might also have different levels of “noise” in their data. This type of model has not been the subject of much debate, because these models assume that behaviors of people tested separately are independent (i.e., that people do not influence each other via ESP).

More controversial is the proposal that TE model be applied to individual data with the assumption that a person might have different “true” preferences in different blocks of trials during the course of a long study (Birnbaum, 2011; Birnbaum & Bahra, 2012). This approach uses the variability of response by the same person to the same item within the same block of trials in order to separate variability due to “error” from “true” intention. This model contradicts the iid assumptions of Regenwetter, et al. (2011), which until recently have been assumed but not tested empirically.

This second type of TE model, called individual true and error theory (*i*TET), allows that, in a long experiment, a person might have different “true” preferences at the end of the study from those at the beginning. If a person has only one “true” pattern of preferences in all blocks, then *i*TET model implies that responses will satisfy the assumptions of iid. However, when a person has more than one true pattern (changing systematically during the study), the assumptions of iid will not in general be satisfied in this model (Birnbaum, 2011). The present studies tested these iid properties.

## 1.6 Overview of tests, results and implications

Experiment 1 found overwhelming evidence against iid. The violations of iid suggest that many if not most participants systematically changed their “true” preferences during the course of the study. Violations of iid mean that we cannot properly restrict our analysis to choice proportions, but we should examine response patterns in order

to test properties such as transitivity.

Nevertheless, choice proportions are analyzed for comparison with related theories, such as the priority heuristic, that make predictions at the level of average choice proportions (summarized in appendices). The averaged choice proportions did not agree with this heuristic and not one person had data consistent with it.

If a LS model holds, there can be linked patterns of intransitivity in linked designs. This was the key idea that led to Experiment 1, but as shown below, very few participants showed evidence of intransitive preferences in any of the three experiments and only one person showed intransitivity in two linked designs that might be compatible with a LS model. A further analysis of response patterns in every individual block of data found little evidence that many, if any, people held intransitive patterns as portions of a mixture of strategies.

The property of interactive independence should be satisfied according to all LS models; however, most participants in Experiments 2 and 3 violated this property systematically (as predicted by interactive models such as EU, CPT, and TAX) including even the one participant who showed linked violations of transitivity. The results led to the surprisingly strong conclusion that LS models can be rejected for nearly every participant.

## 2 Method

Each participant made choices between gambles, knowing that 10 participants would play one of their chosen gambles for real cash prizes. Each gamble was described as an urn containing 100 otherwise identical tickets, which differed only in the prize values printed on them. A ticket would be drawn randomly from the chosen urn to determine the cash prize. Participants were told that any of the choice problems might be selected for play, so they should choose carefully. At the conclusion of the study, randomly selected participants were awarded prizes, as promised.

### 2.1 Stimuli and designs

Each choice was displayed as in the following example:

**First Gamble:**  
50 tickets to win \$100  
50 tickets to win \$0  
OR  
**Second Gamble:**  
50 tickets to win \$35  
50 tickets to win \$25

Participants viewed the choices via computer and indicated their decisions by clicking one of two buttons to identify the gamble they would rather play in each choice.

Three linked sub-designs were used to test transitivity (Table 1). The *LH design* used 5 binary gambles in which probability was 0.5 and in which the Lowest (L) and Highest (H) consequences were varied. In the *LP design*, the highest prize was fixed to \$100 and both probability (P) and lowest consequence (L) were varied. In the *PH design*, the lowest consequence was fixed to \$0 and both probability (P) and highest consequence (H) were varied.

The five gambles within each of the *LH*, *LP*, or *PH* designs could appear as either First or Second gamble, making  $5 \times 5 = 25$  possible choice trials; however, a gamble was not presented with itself, leaving 20 trials in each of these three sub-design. Note that each of 10 distinct choice problems was presented in each of two counter-balanced arrangements in each block.

There were 5 other “filler” sub-designs containing 6 to 48 choices each. These other sub-designs included trials in which a person was asked to choose between gambles with up to five branches (including choices listed in Table 11 of Birnbaum, 2008b), or to choose between gambles and cash prizes to be received for certain. For the purpose of this article, trials in these other designs served as “fillers” that separated blocks of trials. Complete instructions and materials, including the filler tasks, can be viewed at the following URL: [http://psych.fullerton.edu/mbirnbaum/Birnbaum\\_Bahra\\_archive.htm](http://psych.fullerton.edu/mbirnbaum/Birnbaum_Bahra_archive.htm).

### 2.2 Procedures and participants of Experiment 1

Trials in the three main subdesigns (*LH*, *LP*, and *PH*) were blocked in sets of 25 to 26 choices each. Each block included all 20 trials from one sub-design, intermixed with 5 or 6 fillers, and put in restricted random order. This means that each of the 10 choices was presented twice within each block of trials, with position (first or second gamble) counterbalanced. A block of trials including any of the *LH*, *LP*, or *PH* designs was not presented again until at least 98 intervening trials and at most 175 intervening trials with choices from other designs had been presented.

Participants of Experiment 1 were 51 undergraduates enrolled in lower division psychology at California State University, Fullerton. Participants were tested in a lab via computers. Each participant served in two sessions of 1.5 hours each, separated by one week.

Each person worked alone, viewing instructions and materials via computer, and worked at his or her own pace for the time allotted. Therefore, some participants completed more repetitions than others. In Experiment 1, the limit was 20 blocks of trials, meaning each of the choice problems testing transitivity was judged up to 40 times by each person. For additional detail, see Bahra (2012).

### 2.3 LS design in Experiments 2 and 3

Experiments 2 and 3 included *LH*, *LP*, and *PH* designs plus additional trials that tested interactive independence, the priority heuristic, and CPT. The *LS design* consisted of 16 choices. Five choices testing interactive independence were of the form,  $R = (\$95, p; \$5, 1 - p)$  versus  $S = (\$55, p; \$20, 1 - p)$ , where  $p = 0.95, 0.9, 0.5, 0.1$ , or  $0.05$ . Six others were formed by presenting each of three choices:  $S = (\$99, p; \$1, 1 - p)$  versus  $R = (\$40, p; \$35, 1 - p)$ , where  $p = 0.9, 0.5$ , or  $0.1$ , with either  $S$  or  $R$  presented first. There were five additional trials, as follows:  $R = (\$90, 0.05; \$88, 0.05; \$2, 0.9)$  versus  $S = (\$45, 0.2; \$4, 0.2; \$2, 0.6)$ ,  $R_+ = (\$90, 0.1; \$3, 0.7; \$2, 0.2)$  versus  $S_- = (\$45, 0.1; \$44, 0.1; \$2, 0.8)$ ,  $S2 = (\$40, 0.4; \$5, 0.1; \$4, 0.5)$  versus  $R2 = (\$80, 0.1; \$78, 0.1; \$3, 0.8)$ ,  $S3_- = (\$40, 0.2; \$39, 0.2; \$3, 0.5)$  versus  $R3_+ = (\$80, 0.2; \$4, 0.7; \$3, 0.1)$ , and  $G4 = (\$99, 0.30; \$15, 0.65; \$14, 0.05)$  versus  $F4 = (\$88, 0.12; \$86, 0.70; \$3, 0.18)$ . These five trials test implications of the priority heuristic and CPT (see Birnbaum, 2008c).

### 2.4 Procedures and participants in Experiment 2

Experiment 2 was conducted as a replication of Experiment 1 with new participants, except with different filler designs between blocks and the addition of the *LS design*, which allowed us to test if those people showing signs of intransitive preferences also satisfied a critical property of LS models. In Experiment 2, blocks containing *LH*, *LP*, and *PH* designs (each with 5 or 6 intermixed trials from the *LS design*) were separated by at least 76 intervening trials, which included different intervening choices from those used in Experiment 1. There were 43 different undergraduates from the same “participant pool” tested with this procedure. Further description of the “filler” tasks, which tested restricted branch independence and stochastic dominance, are described in Birnbaum and Bahra (2012, Study 1).

### 2.5 Procedures and participants in Experiment 3

Experiment 3 was conducted to investigate two conjectures: First, it was conjectured that if all three transitivity designs were intermixed to make larger and more heterogeneous blocks, it might be more “confusing” to subjects, and this might induce more intransitivity. In the first two experiments, where a block of trials contained 25 trials, it was argued, people might remember preferring  $A$  to  $B$  and  $B$  to  $C$  when they were comparing  $A$  versus  $C$ , so they might obey transitivity because of an experimental demand for consistency. The idea was that by intermixing

trials and spreading them out over larger blocks, memory would be overburdened, so intransitive data might be observed. Therefore, trials of all three transitivity designs were intermixed in Experiment 3.

Second, it was argued (Regenwetter, et al., 2011) that if trials from one design are separated by 3 intervening “filler” trials, responses might satisfy iid, presumably also due to the burdens of memory. Therefore, order was constrained in Experiment 3 such that any two trials from the *LH*, *LP*, or *PH* designs were separated by at least 3 trials from other designs. Perhaps these procedures would “help” iid to be satisfied.

Each block of Experiment 3 consisted of 107 trials (including 20 trials of the *LH design*, 20 trials of the *PH design*, 20 trials of the *LP design*, 16 trials of the *LS design*, and 31 other trials consisting of choices problems like those in Birnbaum, 2008b). Following a warmup of four “filler” trials, each pair of trials from any of the *LH*, *LP*, or *PH* designs was separated by at least 3 intervening trials from other designs. Each block of 107 mixed trials was separated by a separate “filler” task with 57 trials involving choices between three-branch gambles, with equally likely consequences. Experiment 3 used 42 undergraduates from the same pool, who served in two, 1.5-hr sessions, one week apart, and who completed at least 10 blocks. Materials from Experiment 3, including, the filler task, can be viewed from the following URL: [http://ati-birnbaum-2009.netfirms.com/Spr\\_2010/thanks3.htm](http://ati-birnbaum-2009.netfirms.com/Spr_2010/thanks3.htm).

## 3 Results

### 3.1 Data reliability and consistency

Each of the ten basic choices in each of the *LH*, *LP*, and *PH* designs was presented twice in each block of trials, with positions counterbalanced. We define *within-block consistency* as the number of consistent choices out of 10 in each block, which required the participant to push opposite buttons for the two versions of each choice. If a person “went to sleep” and clicked the same button throughout a block, within-block consistency would be 0; if a person clicked buttons randomly, expected consistency would be 5, and if a person made perfectly consistent preferences (appropriately clicking opposite buttons), this index would be 10 (100%).

Within-block consistency was apparently high in the first two experiments; mean consistencies were 86% and 91% in Exps 1 and 2, respectively). The least consistent individuals in Exps 1 and 2 had self-agreements of 67% and 73%, respectively. Similar figures were found within each subdesign; mean within block consistency was 86%, 85%, and 87% in *LH*, *LP*, and *PH* designs of Exp 1, and 91%, 84%, and 86% in Exp 2, respectively. In Exp 3,

where each block had 107 mixed choices, within-block consistency was 77%.

We defined *between-block consistency* as the mean number of consistent responses out of 20 choices between every pair of trial blocks. If a participant completed 20 blocks of trials, for example, this person judged each of the 10 choices 40 times, and there are 190 pairs of blocks ( $20 \times 19/2$ ) for which the number of agreements can be counted. Comparing two blocks of trials, if a person made the same decisions in all 20 choice problems, the score would be 20; if a person randomly pushed buttons; the expected score would be 10 (50%), and the lowest score possible is 0, if a person made exactly opposite choices on all 20 trials. Note that if a person used the same button on all 20 responses in two blocks (which would produce within-block consistency of 0), between block consistency would be 20 (100%). If a person had a response bias, for example, clicking button 2 when unsure or indifferent, such a bias would increase between-block consistency and decrease within-block consistency.

We found that mean between-block consistency was significantly lower than within-block consistency: 80%, 84%, and 75% in Exps 1, 2, and 3 respectively,  $t(50) = 6.19$ ,  $t(42) = 7.13$ , and  $t(41) = 4.26$ .

### 3.2 Violations of stochastic independence and stationarity

Examining individual data of Experiment 1, we found a surprising result: Some individuals had exactly opposite responses on 20 out of 20 choices between two trial blocks and had perfect within-block consistency on both blocks. The probability of observing this result, assuming iid is as probable as predicting the exact sequence of 20 tosses of a fair coin: less than 1 in a million! Yet 10 people out of 51 (#101, 106, 109, 113, 124, 130, 134, 141, 145, 149) showed such patterns in Exp 1, and most of these showed multiple instances of perfect reversals in their data. In addition, three other people showed reversals of 19 out of 20 choices, which has a probability less than 1 in 50,000. Such results mean that the assumptions of iid are seriously and systematically violated.

Table 2 shows raw data for Participant #134, who showed complete reversals in all three designs. Responses to the 20 choice problems are listed in the *LH* design in the order: *AB, AC, AD, AE, BA, BC, BD, BE, CA, CB, CD, CE, DA, DB, DC, DE, EA, EB, EC, ED*. The integers 1 and 2 indicate responses indicating preference for first or second gamble, respectively. Entries under the columns labeled “order” indicate cases where all 20 responses in a block are perfectly consistent with a transitive order. This person started the experiment with three trial blocks showing inconsistency, but on the seventh block, this person finished the first day with all 60

responses perfectly consistent with the transitive orders, *ABCDE, JIHGF*, and *ONMLK*.

Participant #134 began the second session (day 8) with the same behavior in the *LH* and *PH* designs, but the *LP* data were different; by the 10<sup>th</sup> block of trials, all 60 responses were now perfectly consistent with the orders, *ABCDE, FGHIJ*, and *ONMLK*. During the 11<sup>th</sup> block, data were not perfectly consistent in any of the designs, but on the 12<sup>th</sup> and 15<sup>th</sup> blocks, data were now perfectly consistent with *EDCBA, FGHIJ*, and *KLMNO*, exactly the opposite of that shown at the end of the first day. The odds of a single such perfect reversal, assuming iid, of 60 trials is less than one in a million cubed!

Such dramatic and surprising results from Exp 1 led us to conduct Exp 2 as a replication with different “filler” tasks between blocks and with new participants. In Exp 2 there were 7 of 43 participants who showed at least one such complete reversal of 20 responses (#201, 221, 235, 212, 222, 230, and 232); an additional 4 showed 19 reversals between at least two blocks.

In Experiment 3, where all three designs were intermixed in blocks of 107 trials, one person (#334) showed a complete reversal in the *LP* design. Summed over the three studies, we observed 410 instances of perfect reversals of 20 responses between blocks. So many “1 chance in a million” outcomes cannot be reconciled with the assumptions of iid.

These cases with perfect reversals involve data that are so clean and response patterns that are so different that it is easy to detect obvious and systematic changes in preferences between blocks. Such findings suggest that there might be subtler cases where people change between similar patterns of preference or where responses contain enough variability that one could not spot violations of iid without a statistical tool to detect them.

Birnbaum (2012) devised two statistical tests of iid based on the Monte Carlo procedure suggested by Smith and Batchelder (2008): One test uses the variance of preference reversals between pairs of trial blocks and the other uses the correlation between preference reversals and the separation between blocks (which is correlated with the intervening time between blocks). These tests, summarized in Appendix B, show that most participants in all three studies had significant violations of iid. In Exps 1 and 2 there were six tests of iid for each person (variance and correlation methods in *LH, LP*, and *PH* designs). We found that only 3 out of 51 and only 4 of 43 participants did not have at least one violation of iid significant at the .01 level of significance in Exps 1 and 2, respectively.

Only 4 of 42 in Exp 3 did not have at least one of two tests significant at the .01 level. These results show that the assumptions of iid must be rejected as empirical descriptions. Violations of iid suggest that people

Table 2: Raw data from Case #134 in the *LH*, *LP*, and *PH* Designs. Day indicates the day on which the participant completed each block, denoted “blk”. “Order” indicates where all 20 responses in a block were perfectly consistent with a transitive order. Note that all 60 responses are opposite between Block 7 and Block 15.

Day	blk	<i>LH</i> Design	<i>LH</i> Order	<i>LP</i> Design	<i>Order</i>	<i>PH</i> Design	<i>PH</i> Order
1	1	21112111222121112222		22222212212211221211		22221222112221122111	
1	2	22221121111222121121		21221222212211122221		11112111221222122222	
1	3	11122111122122212222		12122111221122212212		11211112112221122111	
1	4	21212111222122212222		11122211221122112212		22221222122211121111	
1	5	11112111221122212222	<i>ABCDE</i>	22111222222211121111		22221222112211111111	
1	6	11112111221122212222	<i>ABCDE</i>	12221222112211121111		22221222112211121111	<i>ONMLK</i>
1	7	11112111221122212222	<i>ABCDE</i>	22221222112211121111	<i>JHGF</i>	22221222112211121111	<i>ONMLK</i>
8	8	11112111221122212222	<i>ABCDE</i>	11112121222211121111		22221222112211121111	<i>ONMLK</i>
8	9	11112111221122212222	<i>ABCDE</i>	12111112212211122211		22221222112211121111	<i>ONMLK</i>
8	10	11112111221122212222	<i>ABCDE</i>	11112111221122212222	<i>FGHIJ</i>	22221222112211121111	<i>ONMLK</i>
8	11	22111212222221122122		11112121221122212222		11212111221122212222	
8	12	22221222112211121111	<i>EDCBA</i>	11112111221122212222	<i>FGHIJ</i>	11112111221122212222	<i>KLMNO</i>
8	13	22111222212211121211		11112111221122212222	<i>FGHIJ</i>	11112111221122212222	<i>KLMNO</i>
8	14	12221212222222222222		11112111221122212222	<i>FGHIJ</i>	11112111221122212222	<i>KLMNO</i>
8	15	22221222112211121111	<i>EDCBA</i>	11112111221122212222	<i>FGHIJ</i>	11112111221122212222	<i>KLMNO</i>
8	16	22221222212211122111		11112111221122212222	<i>FGHIJ</i>	11112111221122212222	<i>KLMNO</i>
8	17	22211222112211121111		11112111221122212222	<i>FGHIJ</i>	11112111221122212222	<i>KLMNO</i>
8	18	22221222112211121111	<i>EDCBA</i>	12221122121121222222		111121112222212222	
8	19	2212121122121212112		12122212222222222222		11112111221122212222	<i>KLMNO</i>
8	20	22221222111122122112		21111221211221221222		21122111221122212122	

are changing their “true” preferences during a study, in which case it could easily be misleading to analyze only marginal choice proportions. Instead, we should examine response patterns.

### 3.3 Analysis of modal response patterns

According to the family of LS models, it is possible to prefer  $A \succ B, B \succ C, C \succ D, D \succ E$ , and yet prefer  $E \succ A$ . This pattern is denoted 11112, where 1 (or 2) represents a preference response for the alphabetically higher (or lower) labeled gamble, in Choices *AB, BC, CD, DE*, and *AE*, respectively. The opposite pattern, 22221, is also intransitive. All other patterns for these five choice problems are compatible with transitivity.

In each block of trials testing transitivity, there are two such tests, each based on 5 choice problems, where the positions of the gambles are reversed in the two tests. For each person, we determined the most frequent response patterns for these five choice problems separately for each presentation order. Out of 408 possible cases (136 participants by 3 designs), 333 cases were consistent; that is, the same modal pattern was observed in both presentation arrangements.

The number of participants who showed each of the modal response patterns are shown in Table 3, for consistent cases in the *LH, LP*, and *PH* designs, respectively. There were only 7 matrices (involving just 6 participants) with intransitive, consistent modal patterns; 7 cases out of 333 represent only about 2%. Cases were numbered starting with #101, 201, and 301 in the three experiments, respectively, and case numbers for those people showing consistent modal violations of transitivity are listed in Table 3. Only one person (#214) showed intransitivity in two linked designs: 22221 and 22221 in *LH* and *LP* designs. This case will be reexamined later.

### 3.4 Analysis of the Priority Heuristic

The priority heuristic implies the patterns, 22221 and 11112 in Table 3 for *LH* and *PH* designs, respectively. No consistent case showed these modal patterns, so no one obeyed the predictions of the priority heuristic.

The priority heuristic was proposed as a theory to describe the process that (most) people use when making choices (Brandstätter, et al., 2006). It is supposed to fit modal preferences, averaged over participants. Averaging our data (see Appendix C), the median response



Table 3: The frequency of consistent, modal response patterns in *LH*, *LP*, and *PH* designs. To be consistent, the participant had to have the same modal response pattern, over repetition blocks, in both ways of presenting the choices. Patterns 11112 and 22221 are intransitive. There were 51, 43, and 42 participants in Experiments 1, 2, and 3 with three designs each; only 7 cases out of 333 consistent modal patterns were intransitive.

<i>LH</i> Pattern	Exp 1	Exp 2	Exp 3	Total
11111	25	19	10	54
11112	0	0	0	0
others	1	2	3	6
22221	1 (#120)	1 (#214)	1 (#311)	3
22222	14	16	11	41
<i>LP</i> Pattern	Exp 1	Exp 2	Exp 3	Total
11111	27	22	10	59
11112	0	0	0	0
others	0	2	1	3
22221	1 (#125)	1 (#214)	0	2
22222	16	14	19	49
<i>PH</i> Pattern	Exp 1	Exp 2	Exp 3	Total
11111	16	12	10	38
11112	0	1 (#239)	1 (#309)	2
others	0	0	0	0
22221	0	0	0	0
22222	30	25	21	76

proportions are perfectly consistent with both WST and the TI. The averaged data agree with the transitive orders implied by the TAX model with its prior parameters: *ABCDE*, *FGHIJ*, and *ONMLK*; i.e., 11111, 11111, and 22222, in designs *LH*, *LP*, and *PH*, respectively.

As shown in Appendix C, the priority heuristic correctly predicted the most often chosen gamble in the averaged data in only three out of ten choice problems in each sub-design. By predicting only 9 of 30 modal choice proportions correctly, this model performed significantly worse than a random coin toss, which would have a binomial probability of .98 of scoring 10 out of 30 or higher. Therefore, the priority heuristic not only failed to fit the data of any individual, it also failed to describe the averaged data in this study.

Table 4: Percentages of all response patterns in *LH*, *LP*, and *PH* Designs. Column sums may differ from 100, due to rounding.

<i>LH</i> Design	Exp 1	Exp 2	Exp 3
11111	39	39	21
11112	1	0	1
others	37	29	54
22221	3	2	4
22222	20	29	20
<i>LP</i> Design	Exp 1	Exp 2	Exp 3
11111	40	42	26
11112	1	1	1
others	28	23	41
22221	4	4	3
22222	27	31	30
<i>PH</i> Design	Exp 1	Exp 2	Exp 3
11111	22	23	19
11112	2	2	2
others	27	20	40
22221	2	1	3
22222	47	54	36

### 3.5 Individual response patterns

The finding in Table 3 that 98% of individual modal response patterns are transitive does not rule out the possibility that some individuals might have intransitive patterns of preferences as “true” patterns in a mixture of response patterns. In order to explore this possibility, we tabulated all response patterns (see Appendix D). The proportions are shown in Table 4, which shows that intransitive response patterns (11112 and 22221) amounted to 5% or less of all individual response patterns in all three designs of all three experiments. Because some of these responses could occur by random error, this low rate of intransitivity of individual response patterns provides little support for the notion that more than a small number have “true” intransitive patterns as part of a mixture of response patterns. A more detailed analysis is presented in Appendix D, which describes the search for individuals who might have intransitive patterns as secondary or tertiary patterns in a mixture. Appendix H further analyzes these response patterns with respect to iid and TE models.

### 3.6 WST and TI

Whereas Tables 3 and 4 analyze response patterns for five of the choice problems, WST and TI can be examined for all 10 choice problems, which might detect other violations of transitivity besides those implied by LS models.

Violations of iid (Appendix B) cast doubt on any analysis that focuses strictly on marginal choice proportions, including tests of WST and TI. Nevertheless, we examined WST and TI as a third tactic to search for evidence of intransitivity. There were 136 participants with 3 matrices per person, making 408 data matrices; of these, only 18 matrices (4.4%) violated (were not perfectly consistent with) both WST and TI (In *LH*, #120, 126, 137, 140, 151, 214, and 311; in *LP*, #102, 122, 125, 214, and 239; in *PH*, #137, 147, 202, 239, 309, and 338). If one were to apply statistical tests (a dubious procedure given the violations of iid), some of these 18 cases (out of 408) might be declared “nonsignificant;” but in the big picture, it matters little whether the rate of violation of WST and TI is 4.4% or, say, 2%. Further details are in Appendix E.

### 3.7 Intransitive individuals

Because violations of transitivity (by any of the definitions) are so rare in these data, it might be tempting to conclude that no one is ever intransitive. However, we think that such a conclusion is too strong for two reasons: First, it argues from failure to reject the null hypothesis to its “truth”. Second, some intransitive cases appear to be systematic. Table 5 shows response patterns (for the five choice problems of Table 3) for Participants #125, 214, and 309 in each block. Responses are shown for each block in each of two sessions (days), with responses reflected so that identical numbers represent consistent choices. Participant #125 shows evidence of intransitivity in the *LP* design, showing the exact, intransitive response pattern 22221 in 20 of 30 presentations, and showing the same intransitive pattern 7 times in both versions within a block.

One might argue that #125 was transitive in the *LP* design with a “true” pattern of 22222, but responded randomly on the last choice, which produced the modal pattern 22221. But we need to explain why, 27 out of 30 times, this person chose *F* over *J* (a response of 1 for the last choice listed); the binomial probability of 27 or more out of 30 with  $p = \frac{1}{2}$  is less than .00001. This binomial calculation assumes a single “true” pattern (22222) and independence of responses both within and between blocks.

A more complex analysis by means of the TE model (see Appendix F) allows all 30 transitive “true” patterns (but no intransitive ones). Fitting this more general model, the probability to observe the pattern of 22221

both times in a block is estimated to be only 0.15, which under the weaker independence assumptions of the TE model yields a probability of finding 7 or more repeated patterns of 22221 (out of 15 blocks) of only  $p = .003$ . So even with this mixture model, these data are unlikely to arise from such a transitive model.

This response pattern for #125 in the *LP* design, 22221, would be consistent with LPH, LHP, or HLP LS models, if  $\$16 \geq \Delta_L > \$4$ . According to these respective LS models, however, #125 should have shown a pattern of 22221, 22221, or 22222 in the *LH* design, respectively. Instead, #125 had the modal response pattern of 11111 in the *LH* design (also 20 of 30 times, with 8 repeats), which requires  $\Delta_L \leq \$4$  under any of these three models, contradicting the behavior in the *LP* design. Therefore, we cannot use any of the LS models to connect the modal response patterns of Case #125 in these two linked designs. So even if we conclude that Case #125 was truly intransitive in the *LP* design, the results in the *LH* design contradict the compatible linkages of the LS models.

Participant #214 is the only case in which a person had intransitive modal data in two linked designs that might be consistent with a LS model. This person showed the modal pattern 22221 in both the *LH* and *LP* designs and the transitive pattern, 22222, in the *PH* design. These modal patterns of behavior (22221, 22221, and 22222 in *LH*, *LP*, and *PH* Designs) are consistent with the LPH LS model with  $\$16 \geq \Delta_L > \$4$  and  $\Delta_P \leq 0.04$ .

If we argue that #214 was truly transitive with “true” patterns of 22222 in both *LH* and *LP* conditions, we need to explain why this person repeated the exact 22221 pattern on eleven blocks out of 22 opportunities and why the last choice is “1” 19 times out of 22 in the *LH* condition, and 19 of 22 in the *LP* task. Suppose the probability of choosing “1” in the last choice is 0.5; if so, the binomial probability to show 19 or more out of 22 is 0.0004, so it is unlikely that two sets of such data arose from true patterns of 22222, combined with independent random responding on the last choice. Case #214 is the only case of intransitive behavior in more than one design consistent with a single LS model. However, this same person violated a critical property of LS models, as shown in the next section, as did most of the other participants.

### 3.8 Testing interactive independence

The *LS* design of Experiments 2 and 3 provides critical tests of the family of LS models. Four tests of interactive independence (each consisting of two choices) per block of trials were constructed from the *LS* design. For example, consider these two choice problems:  $R = (\$95, .95; \$5)$  versus  $S = (\$55, .95; \$20)$  and  $R' = (\$95, .10; \$5)$

Table 5: Analysis of Participants #125, #214, and #309. *LH* and *LH2* show the response patterns for choice problems *AB*, *BC*, *CD*, *DE*, and *AE* when the alphabetically higher gamble was presented first or second. The patterns, 22221 and 11112 are intransitive (bold font).

S #125								S #214							
day	blk	<i>LH</i>	<i>LH2</i>	<i>LP</i>	<i>LP2</i>	<i>PH</i>	<i>PH2</i>	day	blk	<i>LH</i>	<i>LH2</i>	<i>LP</i>	<i>LP2</i>	<i>PH</i>	<i>PH2</i>
1	1	12111	21111	22211	22211	22222	22222	1	1	21121	21221	21211	21211	22222	22222
1	2	11111	12111	<b>22221</b>	22222	22222	22222	1	2	22222	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222
1	3	12111	21111	<b>22221</b>	22211	22222	22222	1	3	22222	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222
1	4	11111	21111	<b>22221</b>	22211	22222	22222	1	4	11111	21111	<b>22221</b>	<b>22221</b>	22222	22222
1	5	12111	21111	22211	22222	22222	22222	8	5	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222	22222
1	6	11111	11111	<b>22221</b>	22211	22222	22222	8	6	<b>22221</b>	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222
8	7	11111	11111	<b>22221</b>	<b>22221</b>	22222	22222	8	7	22222	<b>22221</b>	<b>22221</b>	22222	12222	<b>22221</b>
8	8	11111	11111	<b>22221</b>	<b>22221</b>	22222	22222	8	8	12211	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222
8	9	11111	11111	<b>22221</b>	<b>22221</b>	22222	22222	8	9	<b>22221</b>	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222
8	10	11111	11111	<b>22221</b>	<b>22221</b>	22212	22222	8	10	<b>22221</b>	22121	<b>22221</b>	22222	22222	22222
8	11	11111	11111	<b>22221</b>	<b>22221</b>	22222	22222	8	11	<b>22221</b>	<b>22221</b>	<b>22221</b>	<b>22221</b>	22222	22222
8	12	11211	11111	<b>22221</b>	22222	22222	22222								
8	13	11111	11111	<b>22221</b>	<b>22221</b>	22222	22222								
8	14	11111	11111	<b>22221</b>	<b>22221</b>	22222	22222								
8	15	11221	11111	22211	<b>22221</b>	22222	22222								

  

S #309							
day	blk	<i>LH</i>	<i>LH2</i>	<i>LP</i>	<i>LP2</i>	<i>PH</i>	<i>PH2</i>
1	1	11121	11111	21111	21111	12222	11222
1	2	11111	11111	11111	11111	12112	12222
1	3	11111	11111	11111	11111	11122	<b>11112</b>
8	4	11111	11111	11111	11111	11111	11111
8	5	11111	11111	11111	11111	11111	11111
8	6	11111	11111	11111	11111	11111	11111
8	7	11111	11111	11111	11111	<b>11112</b>	<b>11112</b>
8	8	11111	11111	11111	11111	<b>11112</b>	<b>11112</b>
8	9	11111	11111	11111	11111	<b>11112</b>	<b>11112</b>
8	10	11111	11111	11111	11111	<b>11112</b>	<b>11112</b>

versus  $S' = (\$55, .1; \$20)$ .<sup>2</sup>

For each of four tests, there are four possible response patterns,  $SS'$ ,  $SR'$ ,  $RS'$ , and  $RR'$  in each block. According to any LS model, a person should prefer either  $S$  and  $S'$  or  $R$  and  $R'$ ; that is,  $SS'$  and  $RR'$ . With any mixture

of LS models, a person might show a mixture of these two response patterns, but should not switch systematically from  $R$  in the first choice to  $S'$  in the second choice, denoted the  $RS'$  pattern. TAX with its prior parameters, implies this  $RS'$  reversal.

Consider again Case #214, whose modal data conformed to the LPH LS model in two linked designs. This person completed 11 blocks of trials with 4 tests of interactive independence each, making 44 possible tests. Out of 44 tests (two choice problems per test), this person had

<sup>2</sup>The four tests of interactive independence described here are as follows: Test 1:  $R = (\$95, .95; \$5)$ ,  $S = (\$55, .95; \$20)$ ,  $R' = (\$95, .1; \$5)$ ,  $S' = (\$55, .1; \$20)$ ; Test 2:  $R = (\$95, .90; \$5)$ ,  $S = (\$55, .90; \$20)$ ,  $R' = (\$95, .05; \$5)$ ,  $S' = (\$55, .05; \$20)$ ; Test 3:  $R = (\$99, .90; \$1)$ ,  $S = (\$40, .90; \$35)$ ,  $R' = (\$99, .10; \$1)$ ,  $S' = (\$40, .10; \$35)$ ; Test 4: same as Test 3 with positions reversed.

the exact response pattern of  $RS'$  in 43 of 44 tests. Therefore, the data of Case #214 cannot be represented by any LS model or mixture of LS models.

In Experiments 2 and 3, there were just four cases that showed consistent evidence of intransitivity: #214, 239, 309, and 311. For these cases, the scores were 43 to 0, 27 to 0, 11 to 0, and 48 to 0, comparing  $RS'$  reversals (predicted by interactive models) versus opposite reversals,  $SR'$ , respectively. Three other cases (from Exps 2 and 3) were identified with partial indicators of intransitivity: #202, 218, and 338. The scores for these cases on the tests of interactive independence are 37 to 0, 33 to 0, and 8 to 0, respectively. So even for those cases that seem most promising for evidence of intransitivity, the data refute interactive independence, which is implied by any LS model or mixture of LS models.

Most individuals, including those whose data appear compatible with transitivity, showed evidence of interaction: Out of the 85 participants in Experiments 2 and 3, there were 79 (93%) who had more response patterns of  $RS'$  against only 2 who had more of the opposite reversal and only 4 who had equal numbers or no reversals. Summed over participants and blocks, there were 1807 blocks with the  $RS'$  pattern compared to only 98 with the  $SR'$  pattern. Interaction rules out all LS models.

Another test from the LS design rules out a sub-class of LS models including the priority heuristic. Any person who uses a LS, starting with the four variables of the priority heuristic (lowest consequence, probability of the lowest consequence, highest consequence, probability of highest consequence), considered in any order, should prefer  $G4 = (\$99, 0.3; \$15, 0.65; \$14, 0.05)$  over  $F4 = (\$88, 0.12; \$86, 0.7; \$3, 0.18)$ , if  $\Delta_L$  and  $\Delta_H \leq \$11$  and  $\Delta_P \leq 0.13$ ). Instead, 84% of participants chose  $F4$  over  $G4$  more than half the time. If we retain a LS starting with any of these four attributes, we must conclude that  $\Delta_L > \$11$ ,  $\Delta_H > \$11$  and  $\Delta_P > 0.13$ , contrary to published parameters needed to account for previous data.

In order for LPH, LHP, and HLP LS models to mimic the transitive predictions of the prior TAX model (which are the most commonly observed patterns in the data), these LS models all require  $\Delta_L \leq \$4$ . These are the only LS models that mimic the TAX model this way, but we must reject them for those people who systematically prefer  $F4$  over  $G4$ , since that requires  $\Delta_L > \$11$ , which contradicts the assumption ( $\Delta_L \leq \$4$ ) needed in order to mimic that transitive model's predictions.

### 3.9 Tests of Cumulative Prospect Theory and the Priority Heuristic

Also included in the LS design were direct tests of CPT that also test the priority heuristic. For any monotonic utility function and any probability weighting function,

CPT implies that if  $R = (\$90, 0.05; \$88, 0.05; \$2, 0.9) \succ S = (\$45, 0.2; \$4, 0.2; \$2, 0.6) \Rightarrow R+ = (\$90, 0.1; \$3, 0.7; \$2, 0.2) \succ S- = (\$45, 0.1; \$44, 0.1; \$2, 0.8)$ . Note that  $R+$  stochastically dominates  $R$ , and that  $S$  stochastically dominates  $S-$ . CPT therefore allows the response pattern  $SR+$  but not the opposite,  $RS-$ , which is implied by TAX with its prior parameters (proofs in Birnbaum, 2008c).

The LPH LS and the priority heuristic imply that a person should choose  $S \succ R$  and  $R+ \succ S-$ , as long as  $\Delta_P \leq 0.30$ , so the priority heuristic implies the  $SR+$  pattern that is also compatible with CPT. The PHL and PLH LS models with  $\Delta_P \leq 0.30$  also imply the same pattern. The HLP, HPL, and LHP LS models imply the pattern  $RR+$  when  $\Delta_H \leq \$45$ . These parameter ranges are extremely large and include by a wide margin plausible values.

There were two tests of this type per block in Experiments 2 and 3. In the two tests, 38 of 43 and 35 of 42 participants in Experiments 2 and 3, respectively, showed more response patterns of  $RS-$  than of the opposite, against only 4 and 4 who showed more of the  $SR+$  pattern compatible with CPT and the priority heuristic. These findings rule out CPT and the priority heuristic as well as the other LS models (with wide parameter ranges) for those participants who systematically show the  $SR+$  pattern. Additional results in the LS design, including individual results, are presented in Appendix G.

## 4 Discussion

Our first experiment was initially designed to test whether those participants who showed evidence of intransitive behavior consistent with use of a LS model in one design would show evidence of linked intransitivity between designs. However, we were surprised by two results from that first study: First, few participants showed plausible evidence of intransitivity in even one design, and no one in that study had consistent evidence of linked intransitivity in two designs.

Second, several individuals completely reversed their preferences between blocks of trials, which refutes the assumption of iid that is required for meaningful analysis of marginal choice proportions, averaged over response patterns. These findings led to a second experiment with new participants and new "filler" tasks between blocks, which also included tests of critical properties of LS models.

The second experiment confirmed that some people completely and perfectly reversed preferences between blocks. Because this behavior has less than one chance in a million under the assumptions of iid, we must reject that assumption. Evidence of intransitive behavior was again quite minimal.

The third experiment was an attempt to alter our procedures more drastically in an attempt to "confuse" partici-

pants by intermixing many different types of trials within blocks and by including multiple “fillers” between related items in order to put a greater burden on memory, which was conjectured as the reason that people behaved transitively. It was suggested that these changes in procedure might also produce better satisfaction of iid. Although these procedures increased “error” and reduced the incidence of perfect reversals, they did not prevent them, nor did these changes in procedure increase the incidence of violations of transitivity. Targeted statistical tests indicated that iid was violated strongly in all three studies by all but a very small number of participants.

When iid can be assumed, it means that an investigator can simplify data analysis by examining only choice proportions. But when iid is dubious, it means that we need to also examine response patterns because choice proportions could easily misrepresent individual data. When testing transitivity, it means that choice proportions can appear transitive when the person’s data are perfectly intransitive and it means that choice proportions can appear intransitive when every single response pattern by the person was transitive.

In a search for intransitive patterns of the type consistent with LS models, only a few cases gave credible evidence of intransitivity. However, these cases also showed evidence of violation of critical properties of LS models, including systematic violation of interactive independence. Other tests led to contradictions in the value of difference thresholds required by LS models to handle the data.

Only one person showed intransitive behavior in two designs that could be linked by a LS model. Case #214 showed data consistent with the LPH LS model with  $\$16 \geq \Delta_L > \$4$  and  $\Delta_P \leq 0.04$ . However, this same person chose *F4* over *G4* 100% of the time, which means that  $\Delta_P > 0.13$ , contradicting the LS model that links these two designs. And this person also systematically violated the critical property of interactive independence 43 times in 44 tests, which means that no LS model can account for this person’s data.

Had we tested only transitivity in separate designs, we would have concluded that cases of intransitive preference are rare. Such findings might modify our assessment of the incidence of this behavior. Some studies claimed evidence of systematic violations (e.g., Tversky, 1969; Myung, et al., 2005) and others claimed that “significant” violations of transitivity might be due to chance (Regenwetter, et al., 2010). Based on these new data, the estimated incidence of violation of transitivity in these designs is below 5%, which is compatible with recent studies with *PH* designs.

Because LS models can handle transitive response patterns as well as intransitive ones, and because studies done to date have examined only a tiny region of the

space of all possible sets of choice problems, evidence concerning the incidence of violations of intransitivity says very little about the empirical standing of the class of LS models and it says little about the general validity of transitivity in the infinite space of all choice problems. The failure to find predicted intransitivity might mean only that the researchers did not yet do the right study.

#### 4.1 Refutation of LS models

However, by using linked designs and by including critical tests of the LS models, we can reach much stronger conclusions regarding the LS family; namely, these models can be rejected as descriptive for most people, including even those who appeared to show indications of intransitivity as well as for those whose data appear to be transitive. These findings agree with other tests of critical properties of LS models (Birnbbaum, 2010; Birnbbaum & Gutierrez, 2007; Birnbbaum & LaCroix, 2008)

If those few cases of systematic intransitivity we observed are “real” (and not due to statistical coincidence), then some other origin must be sought to account for them besides the family of LS models. One possibility that has been suggested is that people use an interactive, integrative model but have a tendency to “round off” via editing (i.e., they assimilate subjective values of attributes that are similar) in a choice problem before applying an interactive, integrative model (Kahneman & Tversky, 1979). Such a model could produce intransitive choices and also violate the critical property of interactive independence (Birnbbaum & Gutierrez, 2007).

This “rounding” or “editing” model should also cause linked violations as long as the rules for rounding stay the same in all designs; if so, it might describe the data of Case #214, who was the only case showing evidence of linked intransitivity in two designs, but not Case #125, whose intransitive data in one design contradicted the modal pattern in another design under any such interpretation. Although the “editing” operations were originally proposed as general descriptions, only one case (out of 136) appears to call for this “rounding” rule. Systematic violations of the editing rules of cancellation and combination have been observed in other studies (Birnbbaum, 2008b), so the empirical status of the editing rules remains doubtful.

#### 4.2 Refutation of the Priority Heuristic

These experiments were designed to produce robust, linked violations of transitivity if a person used the priority heuristic. The priority heuristic is a variant of LPH LS, with specified parameters that are part of the theory. As Birnbbaum (2008a) noted, the optimal parameters for this

model to describe certain published data are indeed close to the values postulated by Brandstätter, et al., (2006). However, neither averaged choice proportions nor individual data of any person agreed with the predictions of the priority heuristic. Furthermore, violations of interactive independence refute the priority heuristic along with the other LS models.

These failures of priority heuristic to predict new data are consistent with findings of other recent studies that tested other implications of this heuristic (Birnbaum, 2008a; 2010; Birnbaum & Bahra, 2012; Birnbaum and LaCroix, 2008; Fiedler, 2010; Gloeckner & Betsch, 2008; Gloeckner & Herbold, 2011; Hilbig, 2008; Rieskamp, 2008).

In response to previous failures of the priority heuristic, Brandstätter, Gigerenzer & Hertwig (2008) proposed that the priority heuristic is preceded by use of an “adaptive toolbox” of other procedures for comparing risky gambles. These other processes were proposed to handle cases where the priority heuristic was refuted by experiments designed to test its implications. For example, they suggested that the model does not apply when there is a large discrepancy in expected value (a ratio exceeding 2), when there is a no conflict resolution, when there are branches that might be cancelled, when two or more branches lead to the same consequence, etc. However, this study contains none of the “triggering conditions” yet postulated to invoke the other heuristics. The gambles compared in each set (Table 1) are very close in expected value, choices between them do not have “no conflict” resolutions, and all gambles have exactly two branches. None of the excuses yet published provide a reason for the model to fail in this study.

### 4.3 Refutation of CPT

The LS design of Experiments 2 and 3 also included individual tests of CPT and the priority heuristic. These tests extend previous findings (Birnbaum, 2008c; Birnbaum, 2010) and show that, analyzed at the level of individuals, the majority show systematic violations of CPT as well as of the priority heuristic. Because the property tested does not assume any specific functional form of the value function on money and it does not assume any particular probability weighting function, the refutation of CPT holds for all parametric versions of that model. This evidence against CPT is consistent with other studies that found systematic violations of that model (Birnbaum, 1999, 2004, 2008b, 2008c; Birnbaum & Bahra, 2012).

### 4.4 Refutation of iid assumptions

The present data show extremely strong evidence against iid. These violations were found even in Experiment 3, where multiple fillers separated related trials within blocks and more than 50 intervening trials separated blocks. Therefore, the assumption of iid should be considered as a dubious basis for determining whether or not formal properties in choice data satisfy structural properties such as transitivity.

The refutation of iid creates difficulties for the analysis of individual choice proportions, rather than response patterns. Regenwetter, et al. (2010, 2011) proposed that marginal choice proportions (averaged over response patterns) could be used in a general method for analyzing algebraic models with probabilistic data. However, the overwhelming and extreme violations of iid indicate that this method might, in principle, lead to wrong theoretical conclusions as well as erroneous statistical results.

There are two general forms of the TE model that require weaker assumptions than iid. These models typically violate iid, except in special cases. In *t*TET, a person is assumed to have the same “true” preferences within a block of trials, but it is allowed that a person might change “true” preferences from block to block. This model implies that iid can be violated within a person, if that person changes systematically during the study.

In a sense, the debate between the methods of Regenwetter, et al. (2011) and of Birnbaum (2011) is a debate about how often a person might change “true” preferences. The method of Regenwetter, et al. (2011) assumes that responses are resampled independently (“true” preferences can change) between every pair of trials, as long as there are several intervening filler trials, whereas the TE approach of Birnbaum (2011) assumes that a person’s true preferences last longer and are theorized to be constant within a block of trials. They are allowed to change between blocks.

In the TE models, iid occurs only in special cases, such as when a person maintains only one “true” preference pattern throughout a study. Because the TE model can allow iid to be satisfied or violated, we think the Regenwetter, et al. iid assumptions are stronger than those required by or implied by the TE model. Trials within the same block and between blocks are assumed to be independent in the iid approach, whereas the TE model assumes only that errors within and between blocks are independent, allowing choices within blocks to be dependent.

When the TE model holds, it can provide more information than is available when iid holds: one can estimate the distribution of “true” preferences in a mixture, whereas in the method of Regenwetter, et al. (2011) one can test mixture models but one cannot discover the distribution of preferences in a person’s mixture. A practical

difficulty of the TE model is that more data are required in order to fit this model to the data.

The correlations between behavior in different blocks represent a problem for both iid models and for a subclass of TE models in which “true” choice patterns are re-sampled independently between blocks. Instead, the present data appear more consistent with the idea that a person follows a model in which parameters change gradually but systematically throughout a study. For example, parameters of the TAX model might change gradually via a random walk from trial to trial and block to block.

It would be useful to produce a statistical method that detects when a person changes “true” preferences; such a test might provide a more accurate model and also allow one to compare the assumptions of iid and TE as special cases of a more general model. According to the iid models, this test should identify that people change “true” preference between every trial (or that they never change), and according to the TE models, this test should indicate that people change true preferences only between blocks.

The finding that iid assumptions are violated agrees with Birnbaum’s (2012) reanalysis of data from Regenwetter, et al. (2011), which concluded there were significant deviations from iid in those data as well. By using two repetitions of each choice problem within each block, and by separating blocks by more than 50 trials, the present studies provide stronger tests of iid than were possible in that study, where each choice problem was presented only once per block and only three trials separated blocks.

#### 4.5 Does transitivity hold everywhere?

Because so few “significant” violations of transitivity have been found in recent studies, including this one, some might argue that they should be dismissed as due to chance. If one uses a 5% level of significance, one expects 5 cases per hundred to be “significant” by chance, so finding a small number of “significant” violations of transitivity does not refute the null hypothesis that transitivity holds for everyone.

But even if one can retain the null hypothesis, it does not mean that the null hypothesis has been proved to be true, and there may indeed be some people who truly violate transitivity. Those particular individuals who significantly violated transitivity might continue to do so if they were tested again. A new study might discover stimuli in which violations would be more apparent. Another procedure might be found in which transitivity is easier to detect. Therefore, despite the weakness of evidence against transitivity in the literature, we think the case for transitivity is still open.

#### 4.6 Conclusions

In contrast, the cases against the priority heuristic, the family of LS models, CPT, and assumptions of iid are quite strong: This study found significant and systematic violation of all four of these theoretical ideas. We found that data systematically violated the predictions of the priority heuristic and no one satisfied them. Only one person showed intransitivity in two designs that could be linked via a LS model; and that person (along with the majority of others) also showed systematic violations in interactive independence, a critical property of LS models. Tests of CPT with any monotonic value and probability weighting functions led to systematic violations by most individuals. Finally, tests of iid showed extremely strong violations indicating that people are likely changing their true preferences during a long study.

#### References

- Bahra, J. P. (2012). *Violations of transitivity in decision making*. (Unpublished master’s thesis). California State University, Fullerton, CA, USA.
- Birnbaum, M. H. (1999). Paradoxes of Allais, stochastic dominance, and decision weights. In J. Shanteau, B. A. Mellers, & D. A. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 27–52). Norwell, MA: Kluwer Academic Publishers.
- Birnbaum, M. H. (2004). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology*, 48, 87–106.
- 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, 263–287.
- Birnbaum, M. H. (2008a). Evaluation of the priority heuristic as a descriptive model of risky decision making: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, 115, 253–260.
- Birnbaum, M. H. (2008b). New paradoxes of risky decision making. *Psychological Review*, 115, 253–262.
- Birnbaum, M. H. (2008c). New tests of cumulative prospect theory and the priority heuristic: Probability–Outcome tradeoff with branch splitting. *Judgment and Decision Making*, 3, 304–316.
- Birnbaum, M. H. (2010). Testing lexicographic semiorders 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 in choice data of the assumption that repeated choices are independently and identically distributed. *Judgment and Decision Making*, 7, 97–109.
- Birnbaum, M. H., & Bahra, J. P. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making*, 7, 402–426.
- Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes*, 104, 97–112.
- Birnbaum, M. H., & LaCroix, A. R. (2008). Dimension integration: Testing models without trade-offs. *Organizational Behavior and Human Decision Processes*, 105, 122–133.
- 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.
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Choices without tradeoffs. *Psychological Review*, 113, 409–432.
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2008). Risky Choice with Heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, & Willemsen (2008) and Rieger & Wang (2008). *Psychological Review*, 115, 281–289.
- Fiedler, K. (2010). How to study cognitive decision algorithms: The case of the priority heuristic. *Judgment and Decision Making*, 5, 21–32.
- Gloeckner, A., & Betsch, T. (2008). Do people make decisions under risk based on ignorance? An empirical test of the priority heuristic versus cumulative prospect theory. *Organizational Behavior and Human Decision Processes*, 107, 75–95.
- Gloeckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, 24, 71–98.
- González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, 109, 137–155.
- Hilbig, B. (2008). One-reason decision making in risky choice? A closer look at the priority heuristic. *Judgment and Decision Making*, 3, 457–462.
- Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences*, 10, 131–153.
- Loomes, G., Starmer, C., & Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica*, 59, 425–440.
- Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39, 641–648.
- Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24, 178–191.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Morrison, H. W. (1963). Testable conditions for triads of paired comparison choices. *Psychometrika*, 28, 369–390.
- Myung, J., Karabatsos, G., & Iverson, G. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, 49, 205–225.
- Regenwetter, M., Dana, J., & Davis-Stober, C. (2010). Testing Transitivity of Preferences on Two-Alternative Forced Choice Data. *Frontiers in Psychology*, 1, 148. <http://dx.doi.org/10.3389/fpsyg.2010.00148>.
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of Preferences. *Psychological Review*, 118, 42–56.
- Regenwetter, M., Dana, J., Davis-Stober, C. P., and Guo, Y. (2011). Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). *Psychological Review*, 118, 684–688.
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1446–1465.
- Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: A review of research on preferential choice. *Journal of Economic Literature*, 44, 631–661.
- Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*, 15, 713–731. <http://dx.doi.org/10.3758/PBR.15.4.713>.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Wakker, P. (2011). *Prospect theory: For risk and ambiguity*. Cambridge, UK: Cambridge University Press.



## Appendix A: Analysis of the LPH Lexicographic Semiorder

In the LPH LS model, the person is assumed to compare first the lowest consequences (L), then the probabilities (P), then the highest consequences (H). Table A.1 shows the predicted patterns of behavior for the LPH LS in the *LH* design for choices *AB, BC, CD, DE*, and *AE*; in the *LP* design for *FG, GH, HI, IJ*, and *FJ*; and in the *PH* design for *KL, LM, MN, NO*, and *KO*, respectively. The numbers 1 and 2 refer to preference for the alphabetically higher or lower alternative, respectively; and “?” designates that the model is undecided. Each row shows the results under different ranges of threshold parameters. According to the priority heuristic,  $\$4 < \Delta_L \leq \$16$  and  $0.04 < \Delta_P \leq 0.16$ , so the predicted patterns are 22221, ????1, and 11112 in the three designs, respectively. Other transitive patterns (e.g., 22111) are possible if people compare subjective values of the stimulus attributes, as in Footnote 1.

Table A. 1. Predicted preference patterns in LPH LS model.

Parameters		Stimulus Designs		
$\Delta_L$	$\Delta_P$	<i>LH</i>	<i>LP</i>	<i>PH</i>
$\Delta_L \leq \$4$	$\Delta_P \leq 0.04$	11111	11111	22222
$\Delta_L \leq \$4$	$0.04 < \Delta_P \leq 0.16$	11111	11111	11112
$\Delta_L \leq \$4$	$0.16 < \Delta_P$	11111	11111	11111
$\$4 < \Delta_L \leq \$16$	$\Delta_P \leq 0.04$	22221	22221	22222
$\$4 < \Delta_L \leq \$16$	$0.04 < \Delta_P \leq 0.16$	22221	????1	11112
$\$4 < \Delta_L \leq \$16$	$0.16 < \Delta_P$	22221	?????	11111
$\$16 < \Delta_L$	$\Delta_P \leq 0.04$	22222	22222	22222
$\$16 < \Delta_L$	$0.04 < \Delta_P \leq 0.16$	22222	22221	11112
$\$16 < \Delta_L$	$0.16 < \Delta_P$	22222	?????	11111

## Appendix B: Testing assumptions of independence and identical distribution (iid)

Birnbaum (2011, 2012) noted that the TE model implies iid only in the special case when a person has only a single “true” preference pattern; if a person changes from one “true” pattern to another from block to block, iid can be violated.

Birnbaum (2012) devised two tests that use Monte Carlo simulations suggested by Smith and Batchelder (2008). Both tests begin by computing the average number of preference reversals between each pair of repetition blocks and the variance of the number of preference reversals between blocks. Suppose a person completed 20 blocks: one counts the number of preference reversals between each of  $190 = 20 \times 19/2$  pairs of blocks, summed over the 20 choice problems in each design. If iid holds, the variance of preference reversals should not be large. In addition, the number of preference reversals between two blocks should not be systematically smaller between blocks that are closer together in time than between blocks that are farther apart in time.

Table B.1 shows the results of tests of iid in Experiments 1 and 2. The mean number of preference reversals between blocks is shown in columns labeled “*m*” for each person in each design. For example, the 3.77 for Case # 101 (first row) means that the average number of preference reversals (out of 20 choice problems) between two repetition blocks was 3.77; in other words, the mean number of agreements between two blocks was  $20 - 3.77 = 16.23$  out of 20 (81%) for this person in the *LH* design.

Table B.1. Analysis of iid assumptions in Experiments 1 and 2 in *LH*, *LP*, and *PH* designs ( $m$  = mean number of preference reversals between blocks,  $var$  = variance,  $p_v$  = simulated  $p$ -level of variance test,  $r$  = correlation,  $p_r$  = simulated  $p$ -level of correlation test).

sub	<i>LH</i> Design					<i>LP</i> Design					<i>PH</i> Design				
	$m$	$var$	$p_v$	$r$	$p_r$	$m$	$var$	$p_v$	$r$	$p_r$	$m$	$var$	$p_v$	$r$	$p_r$
101	3.77	9.80	0.00	-0.12	0.87	4.89	38.40	0.00	-0.40	0.51	2.21	4.24	0.00	-0.85	0.02
102	0.46	0.61	0.23	0.70	0.49	7.26	15.62	0.00	0.87	0.00	2.68	6.51	0.00	0.90	0.01
103	0.85	4.15	0.00	0.58	0.49	0.57	0.94	0.04	0.88	0.04	0.37	0.28	1.00	0.83	0.13
104	0.73	1.34	0.01	0.73	0.31	1.07	2.88	0.00	0.84	0.06	0.94	1.22	0.08	-0.53	0.51
105	0.98	1.99	0.00	0.85	0.07	1.60	7.29	0.00	-0.86	0.07	5.16	14.32	0.00	0.42	0.49
106	4.74	4.36	0.13	0.31	0.49	6.69	27.76	0.00	0.93	0.00	9.43	10.40	0.01	-0.33	0.25
107	5.81	18.06	0.00	0.93	0.00	8.43	54.87	0.00	-0.62	0.09	1.94	6.81	0.00	0.66	0.30
108	3.80	9.73	0.00	0.97	0.00	2.48	3.17	0.15	0.82	0.12	9.11	13.13	0.11	-0.20	0.64
109	9.61	37.65	0.00	0.94	0.00	6.17	26.21	0.00	0.46	0.37	9.45	56.60	0.00	0.98	0.00
110	2.96	2.81	0.34	0.41	0.47	0.66	1.56	0.02	0.86	0.25	0.22	0.17	1.00	-0.74	0.74
111	1.58	6.79	0.00	0.85	0.14	2.26	11.43	0.00	-0.82	0.18	1.04	1.49	0.07	0.40	0.74
112	7.46	15.63	0.00	-0.42	0.38	7.51	15.44	0.00	-0.56	0.18	9.18	11.54	0.00	-0.16	0.64
113	2.75	20.31	0.00	0.54	0.47	9.63	44.65	0.00	0.87	0.00	2.73	17.43	0.00	0.83	0.05
114	3.63	5.16	0.02	-0.27	0.73	9.25	12.20	0.25	0.07	0.85	4.59	7.94	0.00	0.85	0.01
115	0.50	0.33	1.00	0.95	0.02	0.00	0.00	1.00	0.00	1.00	0.18	0.15	1.00	0.82	0.19
116	7.42	11.03	0.00	0.46	0.23	2.56	21.68	0.00	0.70	0.26	3.09	21.70	0.00	0.73	0.18
117	0.75	2.08	0.00	0.71	0.23	1.27	6.85	0.00	0.84	0.07	0.21	0.38	0.05	0.73	0.39
118	2.18	2.73	0.08	0.84	0.06	2.78	4.12	0.01	0.69	0.20	3.14	7.26	0.00	0.59	0.38
119	1.82	6.63	0.00	0.82	0.25	1.16	1.72	0.12	0.54	0.65	0.83	1.38	0.07	0.75	0.34
120	5.07	13.49	0.00	-0.42	0.49	4.03	8.28	0.00	-0.41	0.59	2.68	6.37	0.00	0.92	0.01
121	1.26	1.00	0.21	0.73	0.04	0.29	0.23	1.00	0.83	0.18	1.20	3.92	0.00	0.82	0.07
122	2.48	24.17	0.00	0.81	0.26	4.54	12.41	0.00	0.86	0.04	0.36	0.27	1.00	0.65	0.61
123	0.66	1.11	0.01	0.75	0.25	3.17	9.54	0.00	0.92	0.00	0.47	0.66	0.04	-0.67	0.38
124	7.24	13.88	0.00	0.53	0.20	1.08	2.62	0.00	0.81	0.13	8.60	79.67	0.00	0.27	0.55
125	2.01	2.50	0.06	0.96	0.00	4.21	3.69	0.54	0.53	0.25	0.25	0.21	1.00	-0.77	0.35
126	2.92	2.29	0.65	-0.34	0.51	4.66	4.99	0.14	-0.03	0.96	3.46	4.46	0.01	0.94	0.00
127	3.26	8.28	0.00	0.02	0.99	4.24	18.20	0.00	-0.45	0.54	4.88	9.16	0.00	0.08	0.91
128	8.06	76.54	0.00	0.97	0.00	8.15	21.88	0.00	0.42	0.44	1.38	1.21	0.70	-0.73	0.49
129	2.54	3.17	0.01	0.46	0.43	8.24	14.38	0.00	0.49	0.17	2.04	3.33	0.00	-0.47	0.48
130	0.36	0.27	1.00	0.36	0.68	6.72	79.60	0.00	-0.24	0.67	0.58	0.56	0.40	-0.04	0.97
131	3.04	4.92	0.00	0.69	0.15	5.53	12.00	0.00	-0.34	0.55	1.88	2.06	0.17	-0.05	0.95
132	2.56	3.49	0.01	0.33	0.64	7.77	11.55	0.00	0.57	0.15	3.24	11.41	0.00	0.45	0.57
133	1.22	1.68	0.07	-0.88	0.02	1.23	4.01	0.00	0.70	0.25	0.47	0.50	0.39	0.70	0.33
134	9.65	42.46	0.00	0.80	0.00	8.84	35.23	0.00	0.81	0.00	9.79	70.34	0.00	0.72	0.00
135	3.00	12.38	0.00	0.93	0.00	1.03	2.31	0.00	0.80	0.11	1.41	4.89	0.00	0.88	0.01
136	2.50	4.44	0.00	0.85	0.08	0.59	0.74	0.32	0.76	0.50	0.20	0.16	1.00	-0.61	0.78
137	7.72	19.40	0.00	0.82	0.02	6.82	11.87	0.01	0.73	0.10	7.54	11.69	0.08	0.71	0.08
138	1.79	2.90	0.00	0.80	0.06	1.20	2.09	0.01	-0.43	0.62	4.52	10.86	0.00	0.01	0.99
139	2.05	2.18	0.04	0.66	0.15	2.51	5.17	0.00	0.82	0.05	4.61	11.80	0.00	0.62	0.19

sub	LH Design					LP Design					PH Design				
	<i>m</i>	var	<i>p<sub>v</sub></i>	<i>r</i>	<i>p<sub>r</sub></i>	<i>m</i>	var	<i>p<sub>v</sub></i>	<i>r</i>	<i>p<sub>r</sub></i>	<i>m</i>	var	<i>p<sub>v</sub></i>	<i>r</i>	<i>p<sub>r</sub></i>
140	5.42	12.18	0.00	0.78	0.03	1.51	5.13	0.00	0.68	0.36	1.11	1.26	0.22	0.87	0.05
141	5.66	17.53	0.00	0.96	0.00	3.73	15.84	0.00	0.96	0.00	5.27	52.39	0.00	-0.39	0.56
142	0.67	0.61	0.68	-0.75	0.28	0.57	0.94	0.05	0.84	0.13	0.47	0.84	0.02	-0.72	0.30
143	6.21	9.23	0.00	0.53	0.27	4.50	11.88	0.00	0.73	0.12	2.92	7.02	0.00	0.73	0.20
144	6.00	11.77	0.00	0.90	0.00	3.25	7.61	0.00	0.80	0.05	5.38	21.14	0.00	0.96	0.00
145	7.63	20.84	0.00	0.37	0.38	8.93	19.84	0.00	-0.18	0.63	4.61	11.80	0.00	0.62	0.20
146	4.01	7.87	0.00	0.62	0.28	8.73	12.64	0.00	0.34	0.39	8.21	12.64	0.01	0.43	0.36
147	4.58	4.96	0.05	0.59	0.19	2.53	3.68	0.01	0.85	0.01	8.66	13.84	0.00	0.70	0.03
148	2.73	2.85	0.09	-0.76	0.05	5.05	8.48	0.00	0.34	0.56	1.01	1.71	0.01	-0.04	0.97
149	8.73	73.96	0.00	0.55	0.11	0.85	0.98	0.23	-0.58	0.49	0.49	0.52	0.40	0.51	0.56
150	1.82	8.87	0.00	0.77	0.23	2.96	5.06	0.01	0.65	0.35	1.15	2.00	0.01	0.18	0.87
151	7.93	11.90	0.02	0.82	0.01	2.13	2.82	0.09	0.71	0.27	1.66	1.99	0.17	-0.64	0.41
201	4.72	26.52	0.00	-0.70	0.16	2.95	40.45	0.00	0.40	0.62	6.55	74.08	0.00	0.57	0.22
202	2.37	2.72	0.21	0.88	0.02	5.68	7.25	0.53	-0.38	0.59	6.86	13.44	0.00	-0.70	0.14
203	2.49	29.45	0.00	0.75	0.19	0.86	0.48	1.00	0.75	0.21	3.01	5.97	0.00	0.18	0.83
204	4.84	40.18	0.00	0.96	0.00	2.68	14.07	0.00	0.92	0.01	1.97	6.38	0.00	0.94	0.00
205	0.00	0.00	1.00	0.00	1.00	3.35	47.88	0.00	0.64	0.31	0.90	1.53	0.02	0.73	0.24
206	2.21	21.63	0.00	0.80	0.19	0.35	0.50	0.17	0.73	0.40	6.75	38.62	0.00	0.97	0.00
207	1.50	2.60	0.02	-0.73	0.37	2.31	4.55	0.00	0.93	0.00	0.79	2.07	0.00	-0.83	0.14
208	5.42	5.66	0.14	0.02	0.97	5.24	22.11	0.00	0.86	0.00	7.16	11.97	0.01	0.68	0.10
209	2.25	2.14	0.42	0.27	0.76	5.50	11.69	0.08	0.79	0.26	5.02	5.62	0.14	0.49	0.32
210	2.21	11.16	0.00	0.68	0.25	1.19	2.34	0.00	-0.40	0.67	5.77	44.76	0.00	0.38	0.51
211	7.09	9.81	0.00	-0.53	0.19	2.68	8.11	0.00	0.56	0.41	2.39	22.48	0.00	-0.81	0.09
212	5.51	43.04	0.00	0.95	0.00	1.90	4.51	0.00	0.81	0.15	8.66	90.72	0.00	0.98	0.00
213	2.48	12.27	0.00	0.83	0.04	1.05	1.88	0.01	0.77	0.14	1.69	3.89	0.00	0.64	0.27
214	5.72	27.02	0.00	0.62	0.30	3.07	8.64	0.00	0.89	0.02	0.50	1.25	0.01	-0.84	0.26
215	1.33	1.09	0.57	0.83	0.04	1.52	3.14	0.00	0.76	0.17	1.22	0.93	0.77	0.84	0.05
216	5.10	10.03	0.00	0.83	0.00	2.66	3.99	0.00	0.82	0.03	1.11	2.38	0.00	-0.71	0.35
217	1.56	3.74	0.00	0.89	0.00	3.28	7.48	0.00	0.96	0.00	0.84	3.06	0.00	0.82	0.13
218	5.50	25.15	0.00	0.94	0.00	1.12	3.26	0.00	0.82	0.17	0.20	0.16	1.00	0.84	0.45
219	0.33	0.89	0.00	-0.83	0.19	0.21	0.18	1.00	-0.87	0.07	0.10	0.09	1.00	-0.83	0.11
220	0.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00
221	3.90	3.32	0.91	-0.57	0.42	4.32	50.64	0.00	0.86	0.08	1.08	0.84	0.85	0.90	0.06
222	7.64	41.24	0.00	0.85	0.02	1.47	1.21	0.69	-0.56	0.61	7.21	59.92	0.00	0.99	0.00
223	0.08	0.07	1.00	0.69	0.50	5.44	15.82	0.00	0.79	0.00	0.00	0.00	1.00	0.00	1.00
224	3.29	5.25	0.01	0.73	0.14	1.51	1.69	0.18	-0.15	0.88	1.09	1.29	0.28	0.88	0.17
225	0.66	0.69	0.22	0.49	0.50	0.69	0.75	0.28	0.18	0.85	1.85	2.60	0.02	0.42	0.59
226	2.79	12.87	0.00	-0.55	0.49	0.93	5.67	0.00	-0.84	0.15	0.12	0.10	1.00	-0.79	0.25
227	3.09	16.99	0.00	0.86	0.03	4.66	25.45	0.00	0.93	0.00	4.85	6.53	0.00	0.60	0.17
228	7.58	19.95	0.00	0.87	0.00	7.46	23.47	0.00	0.99	0.00	6.25	26.93	0.00	0.81	0.01
229	4.33	20.75	0.00	0.95	0.00	1.26	1.82	0.03	0.32	0.73	0.89	0.80	0.44	0.51	0.54
230	3.12	5.44	0.00	0.98	0.00	1.90	2.76	0.07	0.59	0.52	8.53	55.68	0.00	-0.60	0.25

sub	LH Design					LP Design					PH Design				
	<i>m</i>	var	<i>p<sub>v</sub></i>	<i>r</i>	<i>p<sub>r</sub></i>	<i>m</i>	var	<i>p<sub>v</sub></i>	<i>r</i>	<i>p<sub>r</sub></i>	<i>m</i>	var	<i>p<sub>v</sub></i>	<i>r</i>	<i>p<sub>r</sub></i>
231	4.00	4.50	0.75	0.53	0.52	7.68	25.73	0.00	0.58	0.50	1.78	2.01	0.21	0.08	0.94
232	3.41	19.46	0.00	0.91	0.00	5.39	55.38	0.00	0.93	0.00	7.50	94.12	0.00	0.98	0.00
233	0.66	1.13	0.01	0.51	0.54	1.96	16.68	0.00	0.68	0.34	0.29	0.23	1.00	0.70	0.38
234	1.40	2.82	0.00	0.84	0.06	0.12	0.11	1.00	0.76	0.41	0.12	0.11	1.00	0.76	0.40
235	4.41	5.59	0.00	0.94	0.00	9.47	62.53	0.00	0.98	0.00	1.80	1.89	0.32	-0.54	0.55
236	1.90	3.08	0.04	0.90	0.04	0.31	0.24	1.00	0.16	0.95	9.26	60.03	0.00	-0.01	0.99
237	0.29	0.41	0.15	0.69	0.39	0.74	0.95	0.09	0.85	0.05	0.47	1.00	0.00	0.18	0.84
238	0.95	2.74	0.00	0.85	0.06	1.32	1.10	0.75	-0.65	0.56	5.78	18.80	0.00	-0.34	0.60
239	4.01	3.51	0.78	0.71	0.12	6.22	8.86	0.12	0.67	0.16	5.68	7.09	0.39	0.70	0.14
240	0.40	0.46	0.37	0.58	0.48	6.28	38.50	0.00	0.70	0.02	0.68	1.55	0.00	0.79	0.10
241	0.33	1.22	0.00	0.69	0.36	0.84	1.41	0.01	0.85	0.04	0.67	1.36	0.00	0.60	0.41
242	1.68	1.64	0.09	0.60	0.26	6.20	16.89	0.00	-0.34	0.58	2.32	2.64	0.24	-0.77	0.20
243	2.29	4.77	0.00	0.82	0.03	1.99	1.61	0.60	0.31	0.71	0.28	0.23	1.00	-0.64	0.59

The median numbers of preference reversals between blocks were 2.77, 2.68, and 1.91 in the *LH*, *LP*, and *PH* designs, respectively, corresponding to 86%, 87%, and 90% agreement. In Experiment 1, the medians were 2.96, 3.25, and 2.21, all higher than corresponding values in Experiment 2, which were 2.48, 1.99, and 1.78, respectively. Perhaps agreement between blocks was higher in Experiment 2 because there were fewer filler trials between blocks in Experiment 2 than in Experiment 1.

Next, we computed the mean number of preference reversals between successive blocks, between blocks that are separated by two blocks, by three, etc. These scores were then correlated for each person with the absolute difference between blocks. This correlation would be positive if a person’s behavior changed gradually and systematically from block to block. If iid assumptions hold, however, this correlation would be zero, aside from random fluctuations. These correlation coefficients were computed for each individual for each design of Experiments 1 and 2, and the results are shown under the columns labeled “*r*” in Table B.1.

Table B.1 shows that most of the correlations in Experiments 1 and 2 are positive. The median correlations in *LH*, *LP*, and *PH* designs were 0.71, 0.70, and 0.51, respectively. For individuals, 83%, 76%, and 66% were positive in the *LH*, *LP*, and *PH* designs, respectively, all significantly more than half the samples ( $z = 6.39, 4.95, \text{ and } 2.34$ , respectively). Correlation coefficients of these magnitudes represent serious violations of iid.

For each person, a significance test of the correlations using the Monte Carlo procedure of Smith and Batchelder (2008) was conducted. For each person, responses to each choice problem are randomly permuted between blocks and the correlation coefficient is recalculated for each random permutation. If iid holds, it should not matter how responses to a given choice problem are permuted among blocks. The estimated  $p_r$  value is then the proportion of simulations in which the absolute value of the simulated correlation is greater than or equal to the absolute value of the original correlation in the data. The use of absolute values means that this is a two-tailed test.

Based on 10,000 simulations per person per task, these  $p$ -levels for the correlations are shown under columns labeled “ $p_r$ ” in Table B.1. If iid holds, we expect that 5% of these should be “significant” at the .05 level, which means about 5 people out of 94 in Experiments 1 and 2). Instead, 32, 23, and 19 had  $p_r < 0.05$  in the *LH*, *LP*, and *PH* designs.

Analysis of iid in Experiment 3 is presented in Table B.2. Recall that in Experiment 3, all three transitivity designs (60 trials) were intermixed with each other, with the 16 trials of the *LS* design, and with 31 additional choice problems, making blocks of 107 trials. Each block was separated by at least 57 unrelated trials; in this procedure, two repetitions of the same exact choice problem were separated on average by 164 intervening trials. The median number of preference reversals in this study between blocks was 24.7 out of 107, corresponding to a median agreement rate of 77%. This figure is lower than agreement rates in Experiments 1 and 2 where the *LH*, *LP*, and *PH* designs were in separate blocks, rather than intermixed.

Table B.2. Analysis of iid assumptions in Experiment 3, as in Table B.1. Each block contains 107 choice problems, including *LH*, *LP*, and *PH* designs. Blocks were separated by a filler task with 57 choices.

sub	mean	var	$p_v$	$r$	$p_r$	blocks
301	31.06	95.52	0.00	0.77	0.01	20
302	28.36	139.75	0.00	0.32	0.61	10
303	14.38	36.92	0.02	0.94	0.00	11
304	13.06	31.13	0.03	0.99	0.00	10
305	27.10	97.18	0.00	0.96	0.00	19
306	16.53	84.48	0.00	0.97	0.00	11
307	10.18	22.40	0.04	0.97	0.00	11
308	25.40	154.14	0.00	0.64	0.23	10
309	11.06	71.61	0.00	0.97	0.00	10
310	10.98	130.55	0.00	0.94	0.00	21
311	13.42	63.21	0.00	0.93	0.00	13
312	17.53	152.08	0.00	0.49	0.39	20
313	25.04	93.29	0.06	0.58	0.32	10
314	31.54	141.29	0.00	0.96	0.00	17
315	10.35	25.19	0.00	0.95	0.00	18
316	24.37	204.95	0.00	0.98	0.00	14
317	31.90	157.25	0.00	0.00	1.00	16
318	20.12	99.54	0.00	-0.01	0.99	10
319	16.98	180.31	0.00	0.88	0.00	17
320	3.40	4.13	0.04	-0.56	0.33	23
321	40.08	167.67	0.00	0.88	0.00	19
322	17.54	43.20	0.73	-0.58	0.36	10
323	14.62	97.33	0.00	0.88	0.03	10
324	31.29	150.58	0.00	0.98	0.00	22
325	17.23	79.13	0.00	0.97	0.00	16
326	15.50	104.49	0.00	0.96	0.00	10
327	17.24	55.04	0.00	0.97	0.00	16
328	40.62	157.91	0.00	-0.09	0.86	15
329	30.16	95.43	0.00	0.84	0.00	18
330	18.86	47.72	0.01	0.91	0.00	15
331	41.36	203.51	0.04	0.70	0.08	11
332	25.44	104.98	0.00	0.79	0.06	10
333	22.39	114.92	0.00	0.82	0.01	13
334	49.80	222.60	0.00	0.33	0.34	16
335	26.60	199.23	0.00	0.91	0.00	13
336	41.82	227.59	0.00	-0.11	0.84	12
337	12.86	38.06	0.00	0.33	0.69	11
338	43.50	168.58	0.00	0.88	0.00	17
339	41.51	195.38	0.00	0.48	0.30	12
340	41.26	218.14	0.00	0.79	0.03	11
341	36.99	286.11	0.00	0.95	0.00	11
342	26.64	110.55	0.00	0.88	0.01	11

The median correlation between mean number of preference reversals (over 107 choice problems) and distance in blocks in Experiment 3 was 0.88; only 5 of 42 participants had negative correlations, significantly fewer than half ( $z = -4.94$ ). We would expect only about 2 of 42 should be significant ( $p < .05$ ), but as shown in Table B.2, 27 of 42 individuals had  $p_r < .05$ , highly unlikely under the null hypothesis of iid ( $z = 17.63$ ).

Birnbaum’s (2012) second test of iid compares the variance of the number of preference reversals between blocks against variances simulated via computer-generated permutations of the data. If people have different “true” preferences in different trial blocks, they could show a greater variance of preference reversals than would be found when data are randomly permuted between replication blocks. Even if a person randomly and independently sampled a new “true” pattern before each block of trials, the variance method could potentially detect such violations of iid, which the correlation method would not detect.

By means of the same type of permutations, the  $p_v$ -level was estimated as the proportion of 10,000 permutations of the data in which the variance of preference reversals was greater than or equal to the variance in the original data. Tables B.1 and B.2 show the variances in columns labeled “var”, and the estimated  $p_v$  values. In Experiments 1 and 2,  $p_v$  were “significant” (i.e.,  $p < 0.05$ ) for 67, 68, and 58 out of 94 participants in the *LH*, *LP*, and *PH* designs, respectively. In Experiment 3, all except two (#313 and 322) of the 42 participants had  $p_v < .05$ .

Even with the .01 level of significance, only 11 cases did not have a significant violation of iid in at least one of the tests.

The failure of iid means not only that statistical tests based on this assumption are inappropriate, but also that marginal choice proportions may not be representative of the actual patterns of behavior exhibited by participants, so one might reach wrong conclusions. It also means that we cannot not assume that participants are displaying a static single behavior, but rather that they are likely learning, changing, or shifting their behavior throughout the course of a long experiment.

### Appendix C: Analysis of overall choice proportions and the priority heuristic

Median choice proportions (averaged over all three experiments) are shown in Table C.1 for *LH*, *LP*, and *PH* designs in the upper, middle, and lower portions of the table, respectively. The numbers above the diagonal in each part of the table show the median proportion of responses preferring the column stimulus over the row. For example, the entry of .33 in Row *A* column *C* shows that on average, *C* was chosen over *A* 33% of the time (so in 67% of choices, *A* was chosen over *C*). Because all choice proportions above the diagonal are less than 50%, proportions in this table satisfy WST with the order *ABCDE*, which agrees with the prediction of the TAX model (and CPT) with their prior parameters. These proportions are also perfectly consistent with the TI.

Table C.1. Binary choice proportions (above diagonal) for each design, medians over all three experiments. Predictions of the priority heuristic are shown below diagonal; “?” indicates that the model is undecided.

<i>LH Design</i>					
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i> = (\$84, 0.5; \$24)	0.46	0.33	0.21	0.09	
<i>B</i> = (\$88, 0.5; \$20)		<i>B</i>	0.43	0.32	0.10
<i>C</i> = (\$92, 0.5; \$16)		<i>C</i>	<i>C</i>	0.48	0.17
<i>D</i> = (\$96, 0.5; \$12)		<i>A</i>	<i>D</i>	<i>D</i>	0.24
<i>E</i> = (\$100, 0.5; \$8)		<i>A</i>	<i>B</i>	<i>E</i>	<i>E</i>

  

<i>LP Design</i>						<i>PH Design</i>					
	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>		<i>K</i>	<i>L</i>	<i>M</i>	<i>N</i>	<i>O</i>
<i>F</i> = (\$100, 0.5; \$24)	0.42	0.40	0.37	0.33		<i>K</i> = (\$100, .50; \$0)	0.83	0.88	0.90	0.90	
<i>G</i> = (\$100, 0.54; \$20)	?		0.40	0.43	0.29	<i>L</i> = (\$96, .54; \$0)	<i>K</i>		0.77	0.85	0.90
<i>H</i> = (\$100, 0.58; \$16)	?	?		0.42	0.43	<i>M</i> = (\$92, .58; \$0)	<i>K</i>	<i>L</i>		0.82	0.83
<i>I</i> = (\$100, 0.62; \$12)	<i>F</i>	?	?		0.42	<i>N</i> = (\$88, .62; \$0)	<i>N</i>	<i>L</i>	<i>M</i>		0.77
<i>J</i> = (\$100, 0.66; \$8)	<i>F</i>	<i>G</i>	?	?		<i>O</i> = (\$84, .66; \$0)	<i>O</i>	<i>O</i>	<i>M</i>	<i>N</i>	

Both WST and TI are perfectly satisfied by the median choice proportions in the other two designs as well, shown in middle and lower sections of Table C.1. The majority choice proportions in these designs also agree with predictions of the TAX model with prior parameters: *FGHIJ*, and *ONMLK*.

The predicted majority choices of the priority heuristic are shown below the diagonal in Table C.1 for each design. For example, the priority heuristic predicts that the majority of people should choose *C* in the choice between  $A = (\$84, 0.5; \$24)$  and  $C = (\$92, 0.5; \$16)$  because the difference in the lowest outcome is less than \$10, so the choice should be determined by the highest consequences, which favor *C*. However, the median for this choice was 0.33, which shows that more than half the participants chose *A* over *C* more than half the time. The priority heuristic correctly predicted only three out of ten proportions in each table.

This type of analysis can be (justly) criticized because it is based on averaged choice proportions, which may or may not represent patterns of behavior by individuals. Nevertheless, it is worthwhile to show that the priority heuristic fails to describe averaged choice proportions. The priority heuristic was previously claimed to be an accurate model for predicting modal choice proportions (Brandstätter, et al., 2006, 2008). If this analysis were not presented, the idea might persist that the heuristic might provide a good description of averaged data, even if it fits no single person.

## Appendix D: Analysis of all response patterns

Table D.1 shows an analysis of five choices from the *LH* design: *AB*, *BC*, *CD*, *DE*, and *AE*. Responses are coded such that 1 indicates choice of the gamble represented by the alphabetically higher letter in the choice and 2 indicates choice of the other gamble; therefore, 11111 represents the transitive pattern *ABCDE*; 22222 matches the transitive pattern *EDCBA*. The priority heuristic implies the intransitive pattern, 22221; i.e.,  $E \succ D$ ,  $D \succ C$ ,  $C \succ B$ ,  $B \succ A$ , but  $A \succ E$ . This pattern is consistent with either LPH LS or PLH LS with  $\$16 \geq \Delta_L > \$4$  or with LHP LS with  $\$16 \geq \Delta_L > \$4$  and  $\Delta_H \leq \$4$ .

Table D.1 shows the number of individual trial blocks on which each response pattern on these five choices was observed. The last row in Table D.1 shows the totals. In Experiment 1, for example, participants completed a total of 801 trial blocks in the *LH* design, with two versions of each choice problem per block (there are 1602 responses per item).

Each choice problem was presented twice in each block (with positions counterbalanced); therefore, we can tabulate the frequencies of each possible response patterns when the gambles were presented in one arrangement, (e.g., *AB*), in the other arrangement (e.g., *BA*), or in both. For example, the 315 in the first row of the table (11111) under “ROW” shows that of the 801 blocks in Experiment 1, 315 times a person chose  $A \succ B$ ,  $B \succ C$ ,  $C \succ D$ ,  $D \succ E$ , and  $A \succ E$ , when the gambles were presented with the alphabetically higher-labeled gamble first (e.g., *AB*). The 311 under “COL” shows that 311 times people expressed these same preferences (by clicking opposite buttons) when they were presented with positions reversed (*BA*).

The column labeled “BOTH” shows the number of blocks in which individuals showed exactly the same preference pattern on both versions of the same choices within a block. That is, the person exactly matched 10 responses to show the same decisions on five choice problems presented twice. For example, the 235 under “BOTH” in the first row for Experiment 1 indicates that 235 times (out of 801 blocks), a person had all ten choices matching choice pattern 11111.

The most common response patterns in all three experiments are the transitive patterns, 11111 and 22222, which correspond to the orders, *ABCDE* and *EDCBA*, respectively. These were also the most frequently repeated patterns (BOTH positions), accounting for 88%, 88%, and 81% of the repeated (i.e., consistent) patterns.

We can define within-block, pattern self-consistency as the percentage of times that a person had the same response pattern in both presentations of each choice problem in the same block. Note that pattern self-consistency requires that responses to ten items agree in two presentations of five choice problems. Self-consistency was higher in Experiments 1 and 2 (405/801 is 51% and 438/645 is 68%, respectively), where each trial block had 25 or 26 trials, than it was in Experiment 3, where each trial block had 107 trials (197/591 corresponds to only 33%).

This finding of lower self-consistency in Experiment 3 would be consistent with the idea that people had more “error” (more “confusion”) in Experiment 3, when these different types of trials were intermixed than in the first two studies. It would also be consistent with the idea that people are less likely to maintain the same “true” preferences for 107 trials than for 25 trials.

Table D.1. Frequency of response patterns in tests of transitivity in *LH* Design. The pattern of intransitivity predicted by the priority heuristic is 22221.

pattern	Exp 1			Exp 2			Exp 3		
	ROW	COL	BOTH	ROW	COL	BOTH	ROW	COL	BOTH
11111	315	311	235	254	252	218	118	131	84
11112	4	5	0	2	4	0	5	1	0
11121	11	8	0	5	3	1	18	10	0
11122	2	1	0	1	0	0	5	1	0
11211	47	53	11	25	33	13	31	38	8
11212	2	6	0	1	2	0	7	3	0
11221	6	7	0	4	2	0	7	13	1
11222	4	0	0	2	0	0	5	4	0
12111	33	32	8	20	14	7	14	14	0
12112	4	4	0	0	0	0	9	5	0
12121	6	3	0	0	2	0	13	1	0
12122	4	1	0	8	2	1	10	2	0
12211	15	18	2	13	16	4	22	19	3
12212	1	3	0	2	2	0	6	6	0
12221	7	5	1	4	3	0	11	12	0
12222	7	11	1	10	10	1	24	15	1
21111	23	35	3	17	19	1	31	22	4
21112	3	3	0	0	0	0	7	1	0
21121	10	4	0	5	2	0	6	8	0
21122	3	3	0	3	0	0	9	5	1
21211	25	19	2	13	13	3	18	21	5
21212	5	1	0	2	0	0	3	5	0
21221	11	11	2	4	9	1	5	9	0
21222	8	7	0	6	10	3	7	8	0
22111	13	11	2	12	10	5	14	11	1
22112	3	3	0	0	2	0	4	6	0
22121	6	5	0	2	3	0	10	4	0
22122	8	10	1	8	12	3	15	20	2
22211	29	17	5	16	13	3	24	17	3
22212	18	6	2	3	4	1	10	16	0
22221	25	26	8	12	17	6	25	27	8
22222	143	172	122	191	186	165	98	136	76
Total	801	801	405	645	645	436	591	591	197

The intransitive response pattern predicted by the priority heuristic for the *LH* design, 22221, was observed only 51, 29, and 52 times in Experiments 1, 2, and 3 (3%, 2%, and 4%), and it was repeated by a person (BOTH) only 8, 6, and 8 times within a block (2%, 1%, and 4% of repeated behavior) in the three studies, respectively. These figures represent very small percentages of the overall data, and one should keep in mind that some of this intransitive behavior (though less likely in the BOTH data) might be the result of “error”; for example, cases where the “true” pattern was 22222 and an “error” occurred on the last listed choice.



Table D.2. Frequency of response patterns in tests of transitivity in LP Design.

pattern	Exp 1			Exp 2			Exp 3		
	ROW	COL	BOTH	ROW	COL	BOTH	ROW	COL	BOTH
11111	329	316	275	285	256	239	161	144	106
11112	11	7	2	4	4	1	9	1	0
11121	8	13	0	3	6	0	9	17	0
11122	9	7	2	4	2	0	7	5	0
11211	20	18	0	11	19	2	14	16	3
11212	5	2	0	4	2	1	4	4	0
11221	10	12	1	7	16	5	9	6	0
11222	3	5	0	4	2	0	5	2	0
12111	8	17	0	4	13	0	16	7	0
12112	4	7	0	1	2	0	1	1	0
12121	5	1	0	1	2	0	1	4	1
12122	1	7	0	0	0	0	6	2	0
12211	7	9	0	3	1	0	7	4	2
12212	8	2	1	1	0	0	5	6	0
12221	7	1	0	5	2	0	1	6	0
12222	16	12	2	4	7	1	19	17	1
21111	32	16	4	20	15	8	25	26	4
21112	3	7	0	2	3	0	5	1	0
21121	7	4	2	1	1	0	5	9	0
21122	1	1	0	0	0	0	8	5	0
21211	8	14	2	10	9	3	9	9	0
21212	4	1	0	2	3	0	4	4	0
21221	6	9	1	6	6	3	11	12	0
21222	7	11	0	6	6	3	13	11	1
22111	6	8	0	13	15	10	10	6	0
22112	3	2	0	0	0	0	10	1	0
22121	3	4	0	1	0	0	4	12	0
22122	10	6	0	7	4	1	14	9	2
22211	17	13	3	4	8	1	8	14	0
22212	8	14	0	6	10	1	20	15	0
22221	42	29	13	31	17	12	18	12	0
22222	198	231	176	188	206	177	153	203	122
Total	806	806	484	638	637	468	591	591	242

Although the vast majority of individual response patterns are transitive, there might be a few individuals whose behavior, at least during part of the study, was truly intransitive. These cases of “temporary intransitivity” are more likely to be “real” when the same person repeated the same intransitive pattern in both versions within a block. Four of the 8 cases in Experiment 1 (i.e., BOTH 22221) were produced by #120, who showed this intransitive pattern only in the first two blocks of each day; the last three blocks each day (out of 12 total) were perfectly consistent with the transitive order *EDCBA*. Participant #140 contributed only 1 repeated instance of this pattern, but had 6 other blocks in which this pattern appeared once (out of 14 blocks completed). Three others produced one repeated pattern each.

Table D.3. Frequency of response patterns in tests of transitivity in *PH* Design. The predicted pattern of intransitivity from the priority heuristic is 11112.

pattern	Exp 1			Exp 2			Exp 3		
	ROW	COL	BOTH	ROW	COL	BOTH	ROW	COL	BOTH
11111	180	179	152	149	144	130	121	109	83
11112	11	8	1	11	10	3	9	16	4
11121	19	8	0	5	8	1	16	8	0
11122	6	8	0	3	2	0	4	3	0
11211	12	15	0	5	7	0	8	7	1
11212	6	6	0	1	1	0	7	6	0
11221	6	7	1	1	4	0	6	2	1
11222	6	6	1	1	0	0	5	5	0
12111	10	6	1	8	3	1	10	11	0
12112	5	5	0	3	2	0	8	6	0
12121	3	4	0	3	6	0	4	6	0
12122	5	4	0	2	6	0	8	8	0
12211	6	2	0	5	1	0	6	3	0
12212	7	5	1	3	3	0	3	4	0
12221	10	13	7	10	10	4	9	3	0
12222	18	15	1	25	8	3	15	12	1
21111	7	7	0	3	4	0	13	13	1
21112	8	7	0	7	4	0	10	4	0
21121	7	1	0	0	0	0	7	8	0
21122	7	4	0	4	2	0	6	9	0
21211	4	5	0	2	4	1	4	6	0
21212	7	4	0	4	6	1	3	13	0
21221	2	5	1	1	1	0	3	8	0
21222	18	26	1	17	12	2	20	16	0
22111	1	3	0	0	1	0	9	4	1
22112	6	2	0	1	3	0	3	11	1
22121	4	3	0	0	2	0	4	5	0
22122	14	16	1	5	6	0	28	23	3
22211	3	2	0	0	3	0	5	2	0
22212	23	16	2	11	15	0	12	28	0
22221	14	15	0	3	6	0	17	16	0
22222	365	393	325	340	349	312	208	216	163
Total	800	800	495	633	633	458	591	591	259

In Experiment 2, #214 repeated the 22221 pattern in the *LH* design four times and had 5 other blocks with one instance of this pattern out of 11 blocks; #218 repeated this pattern twice out of 11 blocks completed, but the last 7 blocks were almost perfectly consistent with the transitive order, *EDCBA*.

In Experiment 3, one person (#311) accounted for 4 of the 8 repeated patterns of 22221 in the *LH* design; this person also showed two other blocks with a single instance of this pattern and violated both WST and TI. Four others contributed one repeated pattern each.

Table D.2 shows an analysis of response patterns in the *LP* design. Patterns 11111 and 22222 in this design represent

transitive choice patterns *FGHIJ* and *JIHGF*, respectively. The 22221 pattern represents these intransitive preferences:  $J \succ I, I \succ H, H \succ G, G \succ F$ , but  $F \succ J$ , which are implied by the LPH LS model when  $\$16 \geq \Delta_L > \$4$  and  $0.04 \leq \Delta_P$ ; LHP LS and HLP LS models can also imply this intransitive pattern, if  $\$16 \geq \Delta_L > \$4$  with any  $\Delta_P$ .

Only 13, 12, and 0 blocks with a repeated pattern of 22221 were observed in *LP* Design in Experiments 1, 2, and 3 (3%, 3%, and 0%), respectively. Of the 13 in Experiment 1, 7 were contributed by #125, who also had 6 other blocks with one instance out of 15 blocks; #122 contributed 2 repeats with 5 other instances in 10 blocks; #102 had two blocks repeating the opposite intransitive pattern, 11112, and three other blocks with one instance of that pattern. In Experiment 2, 7 of the 12 repeated patterns of 22221 were from #214, who also had 3 other blocks showing one instance of this pattern; five others contributed one repeated pattern each.

Table D.3 analyzes the *PH* design, where the priority heuristic predicts the intransitive pattern, 11112; i.e.  $K \succ L, L \succ M, M \succ N$ , and  $N \succ O$ , but  $O \succ K$ . This pattern would also be consistent with LPH LS or PLH LS with  $0.16 \geq \Delta_P > 0.04$ , or with PHL, with  $0.16 \geq \Delta_P > 0.04$  and  $\Delta_H \leq \$4$ . This pattern was repeated once, three times, and four times in Experiments 1, 2, and 3. The one repeated pattern in Experiment 1 was by #137, who had 3 other instances of this pattern. Two of the three in Experiment 2 came from #239. All four in Experiment 3 came from #309.

In summary, the analysis of Tables D.1, D.2, and D.3 has added very little, if any, evidence that there are individuals (besides those already identified) who displayed intransitive patterns systematically for large sub-portions of the study. Even if we assume that all observed intransitive response patterns are “real,” Tables D.1, D.2, and D.3 indicate that only 5% or fewer of all response patterns for these five choices in three studies could be described as intransitive.

## Appendix E: Individual Choice Proportions, WST and TI

Tables E.1, E.2, and E.3 show marginal choice proportions for each person in the *LH*, *LP*, and *PH* designs, respectively. Participants in Experiments 1, 2, and 3 were assigned three digit identifiers starting with 101, 201, and 301, respectively.

Table E.1. Binary choice proportions for each individual in *LH* design, WST= weak stochastic transitivity, TI = triangle inequality; “yes” means that the property is perfectly satisfied by the proportions; Order compatible with WST is listed; Blks is the number of blocks, each of which has two presentations of each choice.

sub	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	WST	TI	Order	Blks
101	0.88	0.98	0.95	0.90	0.90	0.90	0.83	0.93	0.85	0.73	yes	yes	EDCBA	20
102	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.04	0.00	0.00	yes	yes	ABCDE	12
103	0.02	0.05	0.02	0.00	0.05	0.02	0.00	0.02	0.02	0.00	yes	yes	ABCDE	20
104	1.00	0.97	1.00	0.97	0.93	0.93	1.00	1.00	1.00	1.00	yes	NO	EDCBA	15
105	0.97	1.00	1.00	0.97	1.00	0.97	1.00	0.97	0.97	0.90	yes	yes	EDCBA	15
106	0.30	0.25	0.10	0.00	0.43	0.25	0.00	0.38	0.02	0.02	yes	yes	ABCDE	20
107	0.79	0.89	0.89	0.71	0.86	0.86	0.82	0.64	0.79	0.86	yes	NO	EDCBA	14
108	0.95	0.91	0.91	0.64	0.91	0.91	0.77	0.86	0.91	1.00	yes	NO	EDCBA	11
109	0.64	0.56	0.61	0.50	0.61	0.58	0.61	0.50	0.50	0.58	yes	yes	EDCBA	18
110	0.61	0.06	0.06	0.00	0.44	0.00	0.00	0.28	0.00	0.00	yes	yes	BACDE	9
111	0.10	0.10	0.00	0.00	0.10	0.00	0.00	0.15	0.00	0.00	yes	yes	ABCDE	10
112	0.58	0.31	0.19	0.31	0.27	0.27	0.12	0.12	0.35	0.35	yes	yes	BACDE	20
113	0.95	0.92	0.92	0.92	0.87	0.92	0.92	0.95	0.95	0.92	yes	yes	EDCBA	19
114	0.17	0.07	0.10	0.10	0.07	0.17	0.03	0.30	0.03	0.10	yes	yes	ABCDE	15
115	0.05	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	yes	yes	ABCDE	10
116	0.43	0.43	0.43	0.15	0.55	0.48	0.15	0.60	0.10	0.10	NO	yes		20
117	0.92	1.00	1.00	0.97	1.00	0.97	0.95	0.97	1.00	1.00	yes	NO	EDCBA	19
118	0.09	0.05	0.00	0.00	0.27	0.09	0.05	0.14	0.00	0.00	yes	yes	ABCDE	11
119	0.05	0.05	0.05	0.05	0.05	0.09	0.05	0.05	0.00	0.09	yes	yes	ABCDE	11
120	0.88	0.79	0.71	0.42	0.96	0.83	0.71	0.88	0.96	0.92	NO	NO		12

sub	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	WST	TI	Order	Blks
121	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.38	0.05	0.00	yes	yes	ABCDE	20
122	0.05	0.05	0.10	0.05	0.10	0.00	0.05	0.15	0.05	0.10	yes	NO	ABCDE	10
123	0.02	0.00	0.02	0.02	0.05	0.00	0.00	0.02	0.02	0.00	yes	yes	ABCDE	20
124	0.40	0.37	0.13	0.13	0.53	0.27	0.23	0.37	0.10	0.27	yes	yes	ACBDE	15
125	0.13	0.07	0.03	0.00	0.13	0.13	0.00	0.07	0.00	0.03	yes	yes	ABCDE	15
126	0.75	0.00	0.00	0.00	0.04	0.00	0.00	0.54	0.04	0.63	NO	NO		12
127	0.35	0.10	0.10	0.05	0.05	0.00	0.05	0.20	0.05	0.10	yes	yes	ABCDE	10
128	0.40	0.30	0.30	0.30	0.30	0.30	0.30	0.35	0.25	0.15	yes	yes	ABCDE	10
129	0.11	0.08	0.05	0.00	0.21	0.03	0.00	0.29	0.00	0.00	yes	yes	ABCDE	19
130	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	yes	yes	ABCDE	20
131	0.21	0.11	0.00	0.03	0.11	0.11	0.13	0.13	0.13	0.00	yes	NO	ABCDE	19
132	0.27	0.07	0.00	0.00	0.07	0.10	0.03	0.20	0.00	0.07	yes	yes	ABCDE	15
133	0.02	0.00	0.00	0.05	0.05	0.05	0.00	0.05	0.08	0.02	yes	NO	ABCDE	20
134	0.58	0.48	0.38	0.38	0.43	0.45	0.45	0.55	0.43	0.43	yes	yes	BADCE	20
135	0.98	1.00	0.95	0.80	0.98	0.95	0.85	0.90	0.85	0.88	yes	NO	EDCBA	20
136	0.75	1.00	1.00	0.88	0.94	1.00	0.81	0.88	0.94	0.94	yes	NO	EDCBA	8
137	0.80	0.80	0.65	0.25	0.80	0.65	0.40	0.75	0.50	0.60	NO	NO		10
138	0.10	0.05	0.05	0.00	0.08	0.00	0.00	0.20	0.02	0.02	yes	yes	ABCDE	20
139	0.05	0.00	0.08	0.00	0.15	0.08	0.00	0.30	0.00	0.00	yes	yes	ABCDE	20
140	0.96	0.79	0.96	0.29	0.75	0.82	0.68	0.96	0.75	0.71	NO	NO		14
141	0.86	0.89	0.94	0.47	0.94	0.94	0.53	0.94	0.53	0.53	NO	yes		18
142	0.02	0.02	0.00	0.02	0.02	0.00	0.02	0.02	0.00	0.02	yes	yes	ABCDE	20
143	0.18	0.24	0.06	0.09	0.26	0.32	0.15	0.53	0.18	0.24	yes	yes	ABDCE	17
144	0.70	0.93	0.78	0.75	0.75	0.75	0.78	0.75	0.90	0.83	yes	NO	EDCBA	20
145	0.60	0.75	0.80	0.83	0.58	0.70	0.70	0.75	0.80	0.70	yes	yes	EDCBA	20
146	0.25	0.13	0.04	0.04	0.04	0.00	0.13	0.42	0.13	0.17	yes	yes	ABCDE	12
147	0.10	0.30	0.10	0.02	0.38	0.23	0.02	0.30	0.13	0.02	yes	yes	ABCDE	20
148	0.20	0.08	0.02	0.00	0.23	0.05	0.05	0.23	0.00	0.00	yes	yes	ABCDE	20
149	0.63	0.71	0.71	0.68	0.66	0.71	0.66	0.63	0.71	0.61	yes	yes	EDCBA	19
150	0.04	0.07	0.04	0.04	0.07	0.11	0.04	0.00	0.04	0.07	yes	NO	ABCDE	14
151	0.71	0.54	0.33	0.25	0.46	0.63	0.25	0.92	0.54	0.67	NO	NO		12
201	0.81	0.84	0.78	0.88	0.88	0.88	0.81	0.81	0.91	0.97	yes	NO	EDCBA	19
202	0.00	0.04	0.04	0.04	0.08	0.08	0.12	0.19	0.00	0.12	yes	NO	ABCDE	15
203	0.84	0.95	0.92	0.95	0.95	0.95	0.92	0.97	0.92	0.95	yes	yes	EDCBA	18
204	1.00	0.85	0.80	0.80	0.85	0.75	0.85	0.90	0.80	0.90	yes	yes	EDCBA	11
205	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	yes	yes	ABCDE	20
206	0.96	0.93	0.93	0.89	0.96	0.93	0.89	1.00	0.93	0.96	yes	NO	EDCBA	15
207	0.04	0.00	0.04	0.08	0.00	0.13	0.00	0.08	0.00	0.04	yes	NO	ABCDE	13
208	0.50	0.62	0.23	0.00	0.54	0.58	0.00	0.42	0.00	0.00	NO	yes		14
209	0.91	0.82	1.00	0.86	0.95	0.91	1.00	0.91	0.95	1.00	yes	NO	EDCBA	13
210	0.88	0.95	1.00	0.95	0.93	0.95	0.95	0.83	0.98	0.98	yes	NO	EDCBA	20
211	0.53	0.31	0.34	0.09	0.41	0.31	0.06	0.88	0.31	0.31	yes	yes	BADCE	20
212	0.82	0.86	0.86	0.79	0.89	0.82	0.79	0.89	0.79	0.79	yes	yes	EDCBA	12
213	0.97	0.95	0.95	0.95	0.76	0.92	0.92	0.95	0.95	0.97	yes	yes	EDCBA	20
214	0.91	0.77	0.77	0.14	0.82	0.77	0.18	0.82	0.77	0.86	NO	NO		11

sub	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	WST	TI	Order	Blks
215	0.06	0.14	0.03	0.06	0.03	0.00	0.03	0.06	0.00	0.00	yes	NO	ABCDE	17
216	0.60	0.57	0.50	0.00	0.73	0.60	0.00	0.17	0.00	0.00	yes	yes	CDBAE	17
217	0.11	0.03	0.03	0.03	0.08	0.00	0.00	0.17	0.00	0.00	yes	yes	ABCDE	18
218	0.96	0.79	0.79	0.63	0.83	0.83	0.67	0.92	0.83	0.92	yes	NO	EDCBA	11
219	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	yes	yes	ABCDE	17
220	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	yes	yes	ABCDE	13
221	0.35	0.05	0.05	0.00	0.10	0.10	0.05	0.20	0.15	0.20	yes	yes	ABCDE	10
222	0.55	0.80	0.85	0.80	0.75	0.75	0.70	0.55	0.75	0.60	yes	yes	EDCBA	10
223	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	yes	yes	ABCDE	20
224	0.08	0.15	0.04	0.04	0.19	0.12	0.04	0.15	0.04	0.19	yes	yes	ABCDE	13
225	0.05	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.10	0.00	yes	NO	ABCDE	20
226	0.08	0.04	0.00	0.04	0.15	0.15	0.04	0.15	0.12	0.04	yes	yes	ABCDE	15
227	1.00	0.89	0.82	0.86	0.89	0.86	0.93	1.00	0.93	0.93	yes	NO	EDCBA	16
228	0.73	0.80	0.78	0.45	0.88	0.88	0.50	0.80	0.48	0.43	yes	yes	DCBAE	19
229	1.00	0.92	0.83	0.86	0.72	0.83	0.78	0.97	0.86	0.86	yes	NO	EDCBA	16
230	0.20	0.00	0.10	0.00	0.20	0.05	0.05	0.25	0.15	0.00	yes	yes	ABCDE	11
231	0.21	0.43	0.29	0.00	0.07	0.07	0.00	0.64	0.07	0.07	yes	NO	ABCDE	11
232	0.96	0.93	0.93	0.86	0.89	0.86	0.82	1.00	0.82	0.93	yes	NO	EDCBA	15
233	0.02	0.00	0.02	0.00	0.05	0.00	0.00	0.02	0.05	0.00	yes	NO	ABCDE	20
234	0.13	0.00	0.00	0.00	0.07	0.03	0.00	0.13	0.00	0.03	yes	yes	ABCDE	18
235	0.03	0.13	0.16	0.00	0.50	0.38	0.03	0.59	0.06	0.03	yes	yes	ABDCE	17
236	0.05	0.00	0.10	0.05	0.10	0.05	0.00	0.05	0.10	0.05	yes	NO	ABCDE	13
237	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.02	0.00	0.00	yes	NO	ABCDE	20
238	0.00	0.04	0.00	0.00	0.08	0.00	0.00	0.08	0.00	0.08	yes	yes	ABCDE	12
239	0.38	0.08	0.04	0.00	0.29	0.21	0.00	0.29	0.13	0.04	yes	yes	ABCDE	13
240	0.95	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.95	yes	yes	EDCBA	20
241	1.00	0.98	1.00	0.98	1.00	0.98	1.00	1.00	0.98	1.00	yes	NO	EDCBA	20
242	0.08	0.00	0.00	0.04	0.00	0.00	0.00	0.42	0.00	0.08	yes	NO	ABCDE	14
243	0.90	0.83	0.93	1.00	0.97	0.97	1.00	0.63	1.00	1.00	yes	NO	EDCBA	12
301	0.63	0.55	0.53	0.02	0.45	0.43	0.05	0.40	0.20	0.25	yes	yes	BCDAE	20
302	0.20	0.25	0.15	0.00	0.40	0.05	0.00	0.60	0.05	0.05	yes	yes	ABDCE	10
303	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05	yes	yes	ABCDE	11
304	0.20	0.15	0.00	0.00	0.05	0.05	0.00	0.10	0.00	0.00	yes	yes	ABCDE	10
305	0.61	0.34	0.13	0.03	0.37	0.21	0.03	0.84	0.08	0.05	yes	NO	BADCE	19
306	0.55	0.82	0.77	0.64	0.82	0.86	0.55	0.82	0.64	0.59	yes	yes	EDCBA	11
307	0.05	0.00	0.00	0.00	0.09	0.00	0.00	0.05	0.00	0.05	yes	yes	ABCDE	11
308	0.80	0.95	0.80	0.85	0.80	0.95	0.85	0.80	0.80	0.60	yes	yes	EDCBA	10
309	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.05	0.05	yes	NO	ABCDE	10
310	0.98	0.95	1.00	1.00	1.00	1.00	1.00	0.93	0.95	1.00	yes	NO	EDCBA	21
311	0.62	0.19	0.19	0.00	0.73	0.46	0.00	0.69	0.12	0.42	NO	NO		13
312	0.83	0.90	0.93	0.70	0.83	0.95	0.48	0.73	0.68	0.73	yes	NO	EDCBA	20
313	0.75	0.90	0.85	0.75	0.95	0.80	0.75	0.65	0.80	0.90	yes	yes	EDCBA	10
314	0.56	0.59	0.32	0.15	0.44	0.35	0.06	0.65	0.21	0.24	NO	yes		17
315	0.06	0.08	0.08	0.03	0.08	0.06	0.00	0.36	0.03	0.03	yes	yes	ABCDE	18
316	0.11	0.18	0.11	0.18	0.14	0.14	0.07	0.21	0.11	0.14	yes	NO	ABCDE	14

sub	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	WST	TI	Order	Blks
317	0.84	0.66	0.72	0.38	0.59	0.66	0.34	0.34	0.69	0.66	NO	yes		16
318	1.00	0.95	1.00	0.95	1.00	0.95	1.00	1.00	0.90	0.95	yes	NO	EDCBA	10
319	0.21	0.24	0.06	0.06	0.12	0.12	0.06	0.15	0.00	0.21	yes	yes	ABCDE	17
320	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.06	0.00	0.00	yes	yes	ABCDE	24
321	0.66	0.74	0.63	0.63	0.63	0.61	0.55	0.63	0.63	0.76	yes	yes	EDCBA	19
322	0.95	1.00	1.00	0.90	1.00	1.00	0.95	1.00	0.90	1.00	yes	NO	EDCBA	10
323	0.20	0.05	0.15	0.05	0.25	0.10	0.00	0.20	0.20	0.25	yes	yes	ABCDE	10
324	0.73	0.84	0.82	0.80	0.91	0.82	0.75	0.73	0.86	0.93	yes	NO	EDCBA	22
325	0.78	0.94	1.00	0.91	0.97	1.00	0.97	0.88	0.91	0.94	yes	NO	EDCBA	16
326	1.00	1.00	1.00	0.90	1.00	1.00	0.85	0.95	1.00	0.95	yes	NO	EDCBA	10
327	0.44	0.31	0.19	0.00	0.25	0.31	0.00	0.59	0.16	0.13	yes	yes	ABDCE	16
328	0.47	0.77	0.80	0.60	0.90	0.93	0.60	0.77	0.67	0.57	yes	yes	EDCAB	15
329	0.58	0.86	0.97	0.67	0.94	0.94	0.83	0.97	0.69	0.67	yes	yes	EDCBA	18
330	0.03	0.23	0.07	0.03	0.10	0.07	0.07	0.30	0.07	0.07	yes	NO	ABCDE	15
331	0.68	0.59	0.50	0.45	0.55	0.64	0.50	0.59	0.23	0.45	yes	yes	DCBAE	11
332	0.60	0.65	0.70	0.70	0.70	0.65	0.50	0.60	0.80	0.90	yes	NO	EDCBA	10
333	0.92	0.96	0.96	1.00	0.92	1.00	1.00	0.77	0.92	0.81	yes	yes	EDCBA	13
334	0.69	0.63	0.56	0.66	0.44	0.59	0.59	0.59	0.56	0.66	yes	yes	EDBCA	16
335	0.50	0.15	0.19	0.08	0.27	0.15	0.08	0.46	0.08	0.19	yes	yes	ABCDE	13
336	0.46	0.54	0.33	0.54	0.38	0.50	0.33	0.33	0.67	0.50	NO	yes		12
337	0.50	0.05	0.05	0.00	0.18	0.05	0.05	0.23	0.14	0.18	yes	yes	ABCDE	11
338	0.50	0.62	0.79	0.71	0.59	0.76	0.65	0.74	0.59	0.74	yes	yes	EDCBA	17
339	0.33	0.42	0.25	0.21	0.50	0.25	0.29	0.33	0.25	0.29	yes	yes	ABCDE	12
340	0.41	0.73	0.50	0.18	0.73	0.71	0.36	0.73	0.45	0.32	yes	yes	DCABE	11
341	0.27	0.41	0.23	0.36	0.68	0.36	0.27	0.45	0.32	0.50	yes	yes	ACBDE	11
342	0.41	0.14	0.05	0.05	0.23	0.36	0.00	0.64	0.23	0.45	yes	yes	ABDCE	11

Table E.2. Binary choice proportions in the LP design, as in Table E.1.

sub	FG	FH	FI	FJ	GH	GI	GJ	HI	HJ	IJ	WST	TI	Order
101	0.10	0.10	0.23	0.08	0.18	0.23	0.20	0.10	0.18	0.13	yes	NO	FGHIJ
102	0.07	0.57	0.71	0.50	0.18	0.54	0.68	0.14	0.57	0.14	NO	NO	
103	0.05	0.02	0.00	0.00	0.00	0.02	0.02	0.02	0.00	0.00	yes	NO	FGHIJ
104	1.00	0.93	1.00	0.93	1.00	0.97	1.00	0.97	1.00	0.90	yes	NO	JIHGF
105	0.93	0.90	0.93	0.93	0.97	1.00	0.97	0.97	1.00	0.97	yes	yes	JIHGF
106	0.35	0.40	0.18	0.13	0.35	0.25	0.10	0.35	0.13	0.15	yes	yes	FGHIJ
107	0.32	0.36	0.29	0.29	0.32	0.29	0.29	0.36	0.39	0.32	yes	yes	FGHIJ
108	1.00	1.00	0.82	0.86	0.91	0.95	0.91	0.95	0.91	0.95	yes	NO	JIHGF
109	0.19	0.14	0.11	0.25	0.25	0.22	0.22	0.28	0.19	0.17	yes	yes	FGHIJ
110	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.06	0.00	0.00	yes	yes	FGHIJ
111	0.05	0.10	0.10	0.00	0.05	0.00	0.05	0.20	0.00	0.10	yes	NO	FGHIJ
112	0.25	0.38	0.25	0.34	0.28	0.13	0.22	0.25	0.25	0.38	yes	yes	FGHIJ
113	0.47	0.47	0.50	0.53	0.55	0.63	0.50	0.63	0.58	0.47	yes	yes	IJFHG
114	0.62	0.54	0.54	0.46	0.73	0.62	0.54	0.62	0.54	0.42	NO	yes	

sub	FG	FH	FI	FJ	GH	GI	GJ	HI	HJ	IJ	WST	TI	Order
115	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	yes	yes	FGHIJ
116	0.05	0.10	0.05	0.05	0.05	0.08	0.02	0.10	0.10	0.10	yes	yes	FGHIJ
117	0.97	0.95	1.00	0.97	0.97	0.92	0.95	0.97	0.95	1.00	yes	NO	JIHGF
118	0.32	0.23	0.05	0.05	0.05	0.00	0.05	0.18	0.00	0.00	yes	yes	FGHIJ
119	0.09	0.05	0.00	0.00	0.00	0.09	0.05	0.00	0.05	0.00	yes	NO	FGHIJ
120	0.79	0.88	0.92	0.92	0.88	0.88	0.88	0.96	0.92	0.75	yes	yes	JIHGF
121	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.02	yes	NO	FGHIJ
122	0.90	0.80	0.80	0.10	0.95	0.95	0.10	1.00	0.35	0.60	NO	NO	
123	0.88	0.95	0.90	0.80	0.95	0.93	0.88	0.95	0.90	0.95	yes	NO	JIHGF
124	0.03	0.13	0.03	0.00	0.00	0.00	0.00	0.07	0.03	0.00	yes	NO	FGHIJ
125	1.00	0.93	0.37	0.10	1.00	0.60	0.13	1.00	0.20	0.77	NO	NO	
126	0.58	0.12	0.31	0.04	0.04	0.00	0.00	0.42	0.15	0.54	yes	yes	GFHJI
127	0.20	0.05	0.20	0.05	0.15	0.10	0.05	0.25	0.30	0.05	yes	yes	FGHIJ
128	0.67	0.44	0.50	0.67	0.78	0.67	0.56	0.72	0.72	0.78	NO	yes	
129	0.85	0.80	0.75	0.63	0.70	0.63	0.60	0.65	0.50	0.58	yes	yes	JIHGF
130	0.28	0.20	0.23	0.20	0.25	0.20	0.20	0.20	0.20	0.20	yes	yes	FGHIJ
131	0.38	0.23	0.30	0.15	0.10	0.15	0.15	0.20	0.10	0.08	yes	yes	FGHIJ
132	0.30	0.33	0.37	0.17	0.23	0.33	0.20	0.33	0.33	0.37	yes	yes	FGHIJ
133	0.89	0.95	0.89	0.97	1.00	0.97	0.97	1.00	1.00	1.00	yes	NO	JIHGF
134	0.30	0.28	0.30	0.35	0.35	0.43	0.30	0.38	0.43	0.43	yes	yes	FGHIJ
135	0.98	0.95	0.98	0.93	0.95	1.00	1.00	1.00	1.00	0.95	yes	NO	JIHGF
136	1.00	0.94	0.94	1.00	1.00	1.00	1.00	1.00	1.00	0.94	yes	NO	JIHGF
137	0.65	0.75	0.85	0.70	0.55	0.90	0.70	0.90	0.55	0.70	yes	NO	JIHGF
138	0.00	0.02	0.05	0.08	0.02	0.05	0.02	0.08	0.00	0.00	yes	NO	FGHIJ
139	0.08	0.10	0.20	0.05	0.02	0.05	0.08	0.00	0.08	0.08	yes	NO	FGHIJ
140	0.89	0.96	0.96	0.89	0.96	1.00	0.96	0.96	1.00	0.96	yes	NO	JIHGF
141	1.00	0.94	0.94	0.69	0.89	0.94	0.69	1.00	0.86	0.83	yes	NO	JIHGF
142	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.00	yes	yes	FGHIJ
143	0.83	0.87	0.90	0.80	0.83	0.90	0.83	0.90	0.87	0.87	yes	yes	JIHGF
144	0.15	0.18	0.03	0.12	0.09	0.06	0.09	0.12	0.06	0.09	yes	yes	FGHIJ
145	0.25	0.33	0.30	0.50	0.45	0.43	0.65	0.33	0.53	0.60	yes	yes	JFGHI
146	0.40	0.33	0.40	0.27	0.40	0.17	0.33	0.40	0.47	0.40	yes	yes	FGHIJ
147	0.03	0.11	0.11	0.03	0.08	0.05	0.08	0.13	0.11	0.03	yes	NO	FGHIJ
148	0.38	0.18	0.08	0.02	0.18	0.20	0.08	0.18	0.13	0.25	yes	yes	FGHIJ
149	0.98	1.00	0.98	0.98	1.00	0.98	0.98	0.98	0.95	0.98	yes	yes	JIHGF
150	0.07	0.14	0.04	0.04	0.04	0.14	0.11	0.14	0.07	0.07	yes	NO	FGHIJ
151	0.86	0.96	0.89	0.89	0.96	0.96	0.93	0.96	0.96	1.00	yes	NO	JIHGF
201	0.05	0.10	0.05	0.13	0.10	0.10	0.08	0.10	0.08	0.08	yes	yes	FGHIJ
202	0.28	0.00	0.17	0.22	0.11	0.17	0.28	0.39	0.17	0.22	yes	NO	FGHIJ
203	1.00	0.97	1.00	0.94	1.00	1.00	1.00	0.97	0.97	0.91	yes	NO	JIHGF
204	0.95	0.91	0.95	1.00	0.73	0.82	0.91	1.00	0.95	0.95	yes	yes	JIHGF
205	0.10	0.10	0.10	0.08	0.15	0.13	0.08	0.10	0.10	0.10	yes	yes	FGHIJ
206	1.00	1.00	0.97	1.00	1.00	1.00	0.97	1.00	0.97	1.00	yes	NO	JIHGF
207	0.13	0.00	0.00	0.08	0.08	0.08	0.08	0.13	0.08	0.00	yes	NO	FGHIJ
208	0.43	0.23	0.30	0.07	0.10	0.17	0.03	0.27	0.20	0.03	yes	yes	FGHIJ

sub	FG	FH	FI	FJ	GH	GI	GJ	HI	HJ	IJ	WST	TI	Order
209	0.08	0.25	0.25	0.42	0.67	0.67	0.92	0.92	0.83	0.92	yes	yes	JIHGF
210	0.95	0.98	0.95	0.95	1.00	0.98	0.93	0.95	1.00	0.98	yes	NO	JIHGF
211	0.97	0.95	0.92	0.92	0.97	0.84	0.89	1.00	0.84	0.92	yes	NO	JIHGF
212	1.00	1.00	0.96	0.96	0.88	0.92	0.92	0.96	0.96	0.92	yes	yes	JIHGF
213	1.00	0.95	0.98	1.00	0.98	0.90	0.98	0.98	1.00	0.95	yes	NO	JIHGF
214	1.00	0.91	0.73	0.14	0.91	0.91	0.23	1.00	1.00	0.91	NO	NO	
215	0.03	0.03	0.13	0.03	0.00	0.05	0.03	0.08	0.05	0.00	yes	NO	FGHIJ
216	0.97	1.00	0.88	0.03	1.00	0.75	0.06	0.22	0.03	0.06	yes	yes	HIGFJ
217	0.14	0.07	0.04	0.00	0.11	0.07	0.07	0.32	0.11	0.11	yes	yes	FGHIJ
218	0.95	1.00	1.00	0.95	1.00	0.91	0.86	1.00	1.00	1.00	yes	NO	JIHGF
219	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	yes	yes	FGHIJ
220	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	yes	yes	FGHIJ
221	0.90	0.70	0.90	0.90	0.90	0.90	0.85	0.85	0.90	0.90	yes	NO	JIHGF
222	0.06	0.00	0.00	0.00	0.13	0.06	0.00	0.06	0.13	0.00	yes	NO	FGHIJ
223	0.40	0.05	0.00	0.00	0.18	0.18	0.05	0.48	0.48	0.45	yes	yes	FGHIJ
224	0.11	0.00	0.04	0.00	0.07	0.00	0.00	0.11	0.07	0.04	yes	yes	FGHIJ
225	0.00	0.02	0.02	0.00	0.02	0.05	0.00	0.00	0.05	0.00	yes	NO	FGHIJ
226	1.00	1.00	1.00	1.00	0.96	0.93	0.96	1.00	0.93	0.96	yes	NO	JIHGF
227	0.88	0.88	0.78	0.69	0.84	0.84	0.97	0.81	0.88	0.94	yes	NO	JIHGF
228	0.32	0.45	0.32	0.11	0.37	0.24	0.13	0.34	0.32	0.21	yes	yes	FGHIJ
229	0.97	0.94	1.00	1.00	0.94	0.97	1.00	0.91	1.00	0.91	yes	yes	JIHGF
230	0.10	0.00	0.10	0.00	0.00	0.15	0.00	0.15	0.05	0.00	yes	yes	FGHIJ
231	0.70	0.80	0.60	0.50	0.60	0.60	0.70	0.30	0.60	0.40	yes	yes	HIGFJ
232	0.18	0.14	0.18	0.07	0.21	0.11	0.18	0.21	0.18	0.18	yes	yes	FGHIJ
233	0.05	0.05	0.08	0.05	0.05	0.03	0.03	0.08	0.03	0.08	yes	yes	FGHIJ
234	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	yes	yes	FGHIJ
235	0.41	0.35	0.35	0.35	0.47	0.41	0.41	0.62	0.44	0.41	yes	yes	FGHIJ
236	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.04	0.00	0.00	yes	NO	FGHIJ
237	0.00	0.02	0.02	0.00	0.00	0.08	0.00	0.05	0.00	0.02	yes	NO	FGHIJ
238	0.09	0.00	0.00	0.00	0.05	0.09	0.05	0.05	0.00	0.05	yes	yes	FGHIJ
239	0.20	0.70	0.85	0.95	0.10	0.75	0.80	0.35	0.75	0.60	NO	NO	
240	0.75	0.65	0.60	0.70	0.70	0.63	0.63	1.00	0.93	0.95	yes	NO	JIHGF
241	0.95	0.95	0.95	0.98	1.00	0.93	1.00	1.00	1.00	1.00	yes	yes	JIHGF
242	0.19	0.08	0.15	0.08	0.12	0.31	0.23	0.50	0.38	0.23	yes	yes	FGHIJ
243	0.79	0.00	0.08	0.04	0.00	0.04	0.00	0.08	0.13	0.04	yes	yes	GFHIJ
301	0.40	0.08	0.10	0.02	0.23	0.10	0.10	0.08	0.13	0.18	yes	yes	FGHIJ
302	0.75	0.65	0.80	0.75	0.65	0.75	0.40	0.50	0.70	0.60	yes	yes	JIHGF
303	0.91	0.91	0.82	0.73	0.82	0.86	0.82	0.91	0.82	0.91	yes	yes	JIHGF
304	0.40	0.15	0.00	0.00	0.05	0.10	0.05	0.30	0.20	0.05	yes	yes	FGHIJ
305	0.84	0.74	0.74	0.63	0.87	0.79	0.58	0.87	0.79	0.66	yes	NO	JIHGF
306	0.14	0.00	0.14	0.05	0.05	0.05	0.00	0.14	0.05	0.18	yes	yes	FGHIJ
307	0.05	0.18	0.18	0.00	0.00	0.05	0.05	0.09	0.00	0.05	yes	NO	FGHIJ
308	0.75	0.70	0.80	0.80	0.70	0.90	0.90	0.85	0.80	0.90	yes	yes	JIHGF
309	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	yes	NO	FGHIJ
310	0.86	0.81	0.86	0.86	0.83	0.83	0.86	0.88	0.88	0.88	yes	yes	JIHGF



sub	FG	FH	FI	FJ	GH	GI	GJ	HI	HJ	IJ	WST	TI	Order
311	1.00	1.00	1.00	0.92	1.00	0.96	0.92	0.96	0.96	0.96	yes	NO	JIHGF
312	0.15	0.03	0.10	0.02	0.05	0.08	0.05	0.03	0.08	0.10	yes	NO	FGHIJ
313	0.90	1.00	0.90	1.00	0.85	0.95	0.80	1.00	0.90	0.95	yes	NO	JIHGF
314	0.88	0.82	0.88	0.71	0.71	0.94	0.79	0.94	0.82	0.85	yes	NO	JIHGF
315	0.03	0.03	0.03	0.00	0.00	0.03	0.00	0.00	0.06	0.03	yes	NO	FGHIJ
316	0.11	0.14	0.14	0.04	0.04	0.14	0.14	0.14	0.11	0.00	yes	NO	FGHIJ
317	0.28	0.22	0.13	0.13	0.09	0.13	0.13	0.09	0.28	0.13	yes	NO	FGHIJ
318	0.95	0.90	0.95	0.85	1.00	0.90	0.90	0.95	0.85	0.85	yes	NO	JIHGF
319	0.12	0.09	0.06	0.09	0.15	0.06	0.03	0.09	0.06	0.09	yes	yes	FGHIJ
320	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.02	0.00	0.02	yes	yes	FGHIJ
321	0.74	0.74	0.76	0.74	0.82	0.87	0.79	0.76	0.87	0.84	yes	yes	JIHGF
322	1.00	0.95	1.00	0.90	1.00	0.95	0.95	1.00	1.00	0.90	yes	NO	JIHGF
323	0.95	0.85	0.75	1.00	0.95	0.75	0.70	1.00	0.90	0.95	yes	NO	JIHGF
324	0.77	0.75	0.77	0.75	0.68	0.61	0.75	0.68	0.77	0.68	yes	yes	JIHGF
325	0.97	1.00	0.94	0.97	0.97	0.94	0.94	0.97	0.97	0.94	yes	NO	JIHGF
326	0.90	0.90	0.95	0.95	0.90	1.00	1.00	0.95	1.00	0.85	yes	yes	JIHGF
327	1.00	0.97	0.94	0.91	0.94	0.97	0.91	1.00	0.97	1.00	yes	NO	JIHGF
328	0.67	0.77	0.63	0.47	0.67	0.73	0.53	0.53	0.50	0.50	NO	yes	
329	0.86	0.92	0.86	0.92	0.97	0.89	0.83	0.92	0.81	0.78	yes	yes	JIHGF
330	0.20	0.03	0.10	0.07	0.13	0.20	0.03	0.60	0.03	0.17	yes	yes	FGIHJ
331	0.64	0.59	0.55	0.77	0.73	0.68	0.50	0.68	0.55	0.77	yes	yes	JIHGF
332	0.95	1.00	0.95	0.90	0.95	0.95	0.90	1.00	0.85	0.90	yes	NO	JIHGF
333	0.58	0.58	0.58	0.42	0.46	0.65	0.54	0.62	0.62	0.58	NO	yes	
334	0.55	0.41	0.50	0.53	0.50	0.44	0.47	0.44	0.44	0.50	NO	yes	
335	0.27	0.15	0.08	0.15	0.15	0.23	0.12	0.42	0.15	0.35	yes	yes	FGHIJ
336	0.42	0.25	0.42	0.33	0.38	0.25	0.29	0.38	0.33	0.42	yes	yes	FGHIJ
337	0.09	0.05	0.05	0.00	0.23	0.05	0.00	0.00	0.09	0.05	yes	NO	FGHIJ
338	0.53	0.68	0.71	0.71	0.65	0.85	0.71	0.65	0.85	0.68	yes	yes	JIHGF
339	0.67	0.54	0.67	0.46	0.58	0.54	0.50	0.54	0.38	0.46	yes	yes	IHGfJ
340	0.73	1.00	0.82	0.55	0.68	0.64	0.73	0.55	0.64	0.68	yes	NO	JIHGF
341	0.82	0.77	0.73	0.82	0.82	0.73	0.68	0.68	0.95	0.77	yes	NO	JIHGF
342	0.48	0.32	0.05	0.00	0.05	0.00	0.09	0.32	0.18	0.41	yes	yes	FGHIJ

Table E.3. Individual binary choice proportions in the PH Design, as in Table E.1.

sub	KL	KM	KN	KO	LM	LN	LO	MN	MO	NO	WST	TI	Order
101	0.05	0.02	0.15	0.10	0.02	0.02	0.02	0.10	0.05	0.08	yes	NO	KLMNO
102	0.79	1.00	1.00	1.00	0.67	0.96	0.96	0.83	1.00	0.92	yes	yes	ONMLK
103	0.98	1.00	1.00	0.95	0.98	1.00	1.00	1.00	1.00	1.00	yes	NO	ONMLK
104	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.13	0.03	0.03	yes	NO	KLMNO
105	0.30	0.17	0.00	0.03	0.13	0.13	0.20	0.23	0.17	0.30	yes	yes	KLMNO
106	0.53	0.48	0.38	0.40	0.45	0.48	0.43	0.70	0.53	0.40	yes	yes	LKNOM
107	0.00	0.00	0.07	0.00	0.07	0.07	0.04	0.14	0.04	0.14	yes	yes	KLMNO
108	0.46	0.58	0.50	0.63	0.54	0.42	0.71	0.58	0.33	0.67	NO	yes	

sub	KL	KM	KN	KO	LM	LN	LO	MN	MO	NO	WST	TI	Order
109	0.39	0.47	0.36	0.47	0.42	0.44	0.39	0.42	0.33	0.39	yes	yes	KLMNO
110	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	yes	NO	ONMLK
111	0.95	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.95	0.85	yes	yes	ONMLK
112	0.38	0.45	0.58	0.40	0.40	0.28	0.43	0.38	0.45	0.40	NO	yes	
113	0.11	0.05	0.08	0.03	0.05	0.05	0.05	0.13	0.11	0.11	yes	yes	KLMNO
114	0.93	0.87	0.87	0.90	0.70	0.93	0.93	0.77	0.83	0.73	yes	yes	ONMLK
115	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	yes	NO	ONMLK
116	0.08	0.08	0.10	0.05	0.13	0.05	0.13	0.10	0.10	0.10	yes	yes	KLMNO
117	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00	yes	NO	KLMNO
118	0.77	0.95	0.95	1.00	0.68	0.95	1.00	0.86	0.91	0.91	yes	yes	ONMLK
119	0.95	0.91	1.00	1.00	0.95	1.00	0.95	1.00	1.00	1.00	yes	yes	ONMLK
120	0.86	0.95	0.95	1.00	0.77	0.95	0.91	1.00	0.91	0.86	yes	yes	ONMLK
121	0.00	0.08	0.02	0.02	0.05	0.00	0.00	0.00	0.05	0.08	yes	NO	KLMNO
122	0.95	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	yes	NO	ONMLK
123	0.95	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.98	0.98	yes	NO	ONMLK
124	0.33	0.33	0.33	0.33	0.33	0.30	0.30	0.30	0.30	0.30	yes	yes	KLMNO
125	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.97	yes	NO	ONMLK
126	0.58	0.96	0.96	1.00	0.73	1.00	1.00	0.85	0.92	0.77	yes	yes	ONMLK
127	0.95	1.00	0.90	0.95	0.60	0.80	0.75	0.90	0.75	0.60	yes	yes	ONMLK
128	0.94	1.00	0.94	0.94	1.00	1.00	0.94	1.00	0.89	0.94	yes	NO	ONMLK
129	0.98	1.00	0.93	1.00	0.90	0.98	1.00	0.85	0.88	0.90	yes	NO	ONMLK
130	1.00	0.97	1.00	1.00	0.95	1.00	0.97	0.95	1.00	1.00	yes	yes	ONMLK
131	1.00	0.98	0.95	0.93	0.88	0.98	0.93	0.93	0.98	0.95	yes	NO	ONMLK
132	0.93	0.93	0.93	0.97	0.83	0.93	0.90	0.87	0.87	0.90	yes	yes	ONMLK
133	1.00	1.00	0.98	0.95	1.00	0.98	1.00	1.00	0.98	1.00	yes	NO	ONMLK
134	0.45	0.45	0.43	0.40	0.40	0.43	0.48	0.50	0.50	0.45	yes	yes	KLMNO
135	1.00	0.98	0.98	0.95	0.95	0.93	0.93	0.98	0.98	0.98	yes	yes	ONMLK
136	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	yes	yes	ONMLK
137	0.15	0.75	0.75	0.90	0.45	0.55	0.65	0.30	0.55	0.30	NO	NO	
138	0.85	0.78	0.95	0.88	0.88	0.88	0.85	0.78	0.85	0.83	yes	yes	ONMLK
139	0.93	0.90	1.00	0.95	0.90	0.90	0.95	0.88	0.95	0.93	yes	yes	ONMLK
140	0.96	1.00	1.00	1.00	0.96	0.96	0.96	0.92	1.00	0.92	yes	yes	ONMLK
141	0.14	0.14	0.11	0.19	0.19	0.11	0.17	0.19	0.14	0.19	yes	yes	KLMNO
142	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.95	1.00	0.95	yes	NO	ONMLK
143	0.07	0.10	0.10	0.17	0.90	0.90	0.93	0.93	0.97	0.97	yes	yes	KONML
144	0.28	0.25	0.13	0.08	0.20	0.20	0.15	0.13	0.10	0.23	yes	yes	KLMNO
145	0.08	0.08	0.05	0.10	0.18	0.08	0.25	0.25	0.18	0.23	yes	yes	KLMNO
146	0.46	0.67	0.71	0.71	0.38	0.50	0.71	0.75	0.71	0.79	NO	yes	
147	0.31	0.66	0.56	0.47	0.28	0.31	0.34	0.38	0.38	0.28	NO	NO	
148	0.95	1.00	0.98	1.00	1.00	0.98	0.98	0.88	0.98	1.00	yes	NO	ONMLK
149	1.00	1.00	0.97	1.00	1.00	1.00	0.97	1.00	1.00	0.92	yes	NO	ONMLK
150	1.00	0.96	1.00	1.00	0.96	0.96	0.93	0.93	1.00	0.93	yes	NO	ONMLK
151	0.96	0.89	0.96	0.93	0.96	0.93	1.00	1.00	0.93	0.96	yes	NO	ONMLK
201	0.18	0.18	0.23	0.20	0.20	0.20	0.23	0.23	0.23	0.23	yes	yes	KLMNO
202	0.22	0.67	0.94	0.89	0.39	0.61	0.78	0.22	0.67	0.50	NO	NO	

sub	KL	KM	KN	KO	LM	LN	LO	MN	MO	NO	WST	TI	Order
203	0.79	0.89	0.92	0.89	0.95	0.97	0.95	0.97	0.95	0.82	yes	yes	ONMLK
204	0.92	1.00	1.00	0.96	0.79	1.00	1.00	0.96	0.96	0.83	yes	yes	ONMLK
205	1.00	0.93	0.98	0.98	0.95	1.00	0.95	1.00	1.00	0.98	yes	NO	ONMLK
206	0.06	0.13	0.13	0.06	0.41	0.31	0.38	0.31	0.38	0.34	yes	yes	KLMNO
207	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05	0.14	yes	yes	KLMNO
208	0.85	0.95	0.85	0.90	0.45	0.40	0.45	0.65	0.50	0.50	yes	yes	LNMOK
209	0.00	0.04	0.08	0.00	0.58	0.23	0.23	0.23	0.19	0.54	yes	yes	KMLON
210	0.83	0.78	0.80	0.80	0.85	0.85	0.80	0.88	0.88	0.75	yes	yes	ONMLK
211	0.95	0.93	0.95	0.95	0.88	0.93	0.95	0.95	0.95	0.93	yes	yes	ONMLK
212	0.31	0.31	0.31	0.31	0.31	0.35	0.31	0.35	0.31	0.35	yes	yes	KLMNO
213	0.94	1.00	0.97	1.00	0.97	0.79	0.94	1.00	0.91	0.97	yes	NO	ONMLK
214	0.95	0.95	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	yes	NO	ONMLK
215	0.95	0.95	0.98	1.00	0.93	1.00	0.98	0.98	1.00	0.95	yes	yes	ONMLK
216	0.97	0.97	0.97	0.97	0.94	0.97	0.97	1.00	0.97	0.97	yes	yes	ONMLK
217	1.00	1.00	0.97	1.00	0.94	1.00	0.94	0.97	0.97	0.97	yes	yes	ONMLK
218	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	yes	yes	ONMLK
219	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	yes	yes	ONMLK
220	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	yes	yes	ONMLK
221	1.00	1.00	1.00	0.95	1.00	1.00	0.95	0.95	0.90	0.95	yes	yes	ONMLK
222	0.33	0.11	0.22	0.39	0.28	0.22	0.22	0.22	0.28	0.22	yes	yes	KLMNO
223	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	yes	yes	ONMLK
224	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	yes	NO	KLMNO
225	0.95	0.95	0.98	0.95	0.93	0.95	0.98	1.00	0.95	0.85	yes	yes	ONMLK
226	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	yes	NO	ONMLK
227	0.35	0.35	0.42	0.46	0.88	0.96	1.00	1.00	0.92	0.92	yes	yes	KONML
228	0.29	0.32	0.21	0.34	0.08	0.24	0.18	0.16	0.13	0.13	yes	yes	KLMNO
229	0.97	0.91	1.00	1.00	0.97	1.00	1.00	0.97	0.97	0.97	yes	NO	ONMLK
230	0.38	0.31	0.38	0.31	0.31	0.31	0.25	0.50	0.31	0.38	yes	yes	KLMNO
231	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.75	0.92	0.75	yes	yes	ONMLK
232	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	yes	yes	KLMNO
233	1.00	0.98	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.98	yes	NO	ONMLK
234	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	yes	yes	KLMNO
235	0.96	0.93	1.00	1.00	0.96	0.93	0.89	0.96	0.96	0.89	yes	NO	ONMLK
236	0.63	0.67	0.63	0.71	0.63	0.54	0.58	0.58	0.54	0.54	yes	yes	ONMLK
237	1.00	1.00	1.00	1.00	1.00	0.98	0.98	1.00	0.95	0.98	yes	NO	ONMLK
238	0.70	0.80	1.00	0.95	0.65	0.80	0.90	0.70	0.75	0.70	yes	yes	ONMLK
239	0.15	0.75	0.90	0.95	0.45	0.85	0.90	0.30	0.85	0.50	NO	NO	
240	0.00	0.00	0.00	0.00	0.10	0.00	0.02	0.02	0.00	0.08	yes	yes	KLMNO
241	0.02	0.02	0.00	0.00	0.00	0.02	0.02	0.08	0.02	0.02	yes	yes	KLMNO
242	0.90	0.95	1.00	0.95	0.75	0.95	0.95	0.95	1.00	0.90	yes	yes	ONMLK
243	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	yes	NO	KLMNO
301	0.90	0.88	0.78	0.80	0.68	0.73	0.78	0.70	0.63	0.60	yes	yes	ONMLK
302	0.85	0.90	0.90	0.85	0.70	0.75	0.80	0.75	0.80	0.70	yes	yes	ONMLK
303	1.00	1.00	1.00	1.00	1.00	0.95	1.00	0.95	0.86	0.95	yes	NO	ONMLK
304	1.00	1.00	1.00	0.90	0.85	0.80	0.95	0.80	0.85	0.70	yes	NO	ONMLK

sub	KL	KM	KN	KO	LM	LN	LO	MN	MO	NO	WST	TI	Order
305	0.97	1.00	0.97	1.00	0.92	0.92	0.95	0.89	0.82	0.92	yes	yes	ONMLK
306	0.05	0.00	0.00	0.00	0.09	0.05	0.00	0.09	0.05	0.09	yes	yes	KLMNO
307	0.95	1.00	0.91	1.00	0.91	0.86	1.00	0.95	0.86	0.73	yes	NO	ONMLK
308	0.95	0.85	0.75	0.90	0.70	0.85	0.80	0.60	0.75	0.95	yes	NO	ONMLK
309	0.00	0.10	0.65	0.70	0.15	0.20	0.15	0.15	0.20	0.20	NO	NO	
310	0.02	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	yes	yes	KLMNO
311	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	0.96	1.00	yes	NO	ONMLK
312	0.10	0.03	0.05	0.08	0.05	0.03	0.00	0.05	0.02	0.05	yes	NO	KLMNO
313	0.60	0.75	0.80	0.75	0.95	0.85	0.75	0.65	0.75	0.85	yes	yes	ONMLK
314	0.88	0.91	0.85	0.94	0.94	0.91	0.94	0.82	0.88	0.91	yes	yes	ONMLK
315	0.06	0.03	0.03	0.11	0.00	0.06	0.00	0.00	0.00	0.03	yes	NO	KLMNO
316	0.04	0.11	0.07	0.07	0.36	0.29	0.39	0.43	0.43	0.36	yes	yes	KLMNO
317	0.13	0.13	0.31	0.16	0.19	0.06	0.09	0.16	0.19	0.16	yes	NO	KLMNO
318	0.85	0.85	0.80	0.75	0.85	0.95	0.90	0.85	0.80	0.90	yes	yes	ONMLK
319	0.91	1.00	0.91	0.88	0.94	0.91	0.97	0.97	0.94	0.91	yes	NO	ONMLK
320	1.00	0.98	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	yes	NO	ONMLK
321	0.74	0.84	0.84	0.82	0.61	0.89	0.61	0.82	0.76	0.76	yes	NO	ONMLK
322	0.85	0.85	0.85	0.95	1.00	0.95	0.90	0.85	0.90	0.95	yes	yes	ONMLK
323	0.90	0.95	1.00	1.00	1.00	0.95	1.00	0.90	1.00	0.95	yes	yes	ONMLK
324	0.18	0.14	0.11	0.11	0.27	0.25	0.25	0.07	0.16	0.20	yes	yes	KLMNO
325	0.97	1.00	0.97	0.94	0.91	0.88	0.81	0.84	0.72	0.88	yes	yes	ONMLK
326	0.05	0.00	0.15	0.15	0.15	0.00	0.10	0.00	0.05	0.10	yes	NO	KLMNO
327	0.97	1.00	1.00	0.97	0.97	1.00	1.00	1.00	1.00	1.00	yes	NO	ONMLK
328	0.77	0.60	0.53	0.60	0.57	0.50	0.43	0.43	0.60	0.60	NO	yes	
329	0.56	0.53	0.42	0.44	0.33	0.25	0.11	0.19	0.20	0.42	yes	yes	LMKNO
330	0.83	0.90	0.90	0.93	0.77	0.90	0.93	0.90	0.90	0.97	yes	yes	ONMLK
331	0.73	0.73	0.77	0.77	0.68	0.64	0.55	0.59	0.73	0.73	yes	yes	ONMLK
332	1.00	0.95	0.95	0.90	0.90	0.95	0.95	1.00	0.95	0.85	yes	NO	ONMLK
333	0.69	0.31	0.00	0.04	0.46	0.38	0.19	0.42	0.23	0.54	yes	NO	LKMON
334	0.50	0.50	0.59	0.47	0.41	0.53	0.47	0.56	0.44	0.50	yes	yes	NKLMO
335	0.92	0.88	0.92	0.92	0.96	0.85	0.92	0.88	0.88	0.92	yes	yes	ONMLK
336	0.29	0.13	0.38	0.29	0.17	0.25	0.17	0.46	0.21	0.13	yes	yes	KLMNO
337	0.05	0.00	0.00	0.09	0.05	0.00	0.09	0.05	0.00	0.09	yes	NO	KLMNO
338	0.71	0.38	0.62	0.65	0.71	0.59	0.68	0.53	0.56	0.71	NO	NO	
339	0.71	0.92	0.75	0.88	0.63	0.71	0.75	0.67	0.75	0.67	yes	yes	ONMLK
340	0.82	0.77	0.55	0.82	0.77	0.73	0.82	0.77	0.64	0.77	yes	yes	ONMLK
341	0.64	0.77	0.68	0.73	0.82	0.77	0.91	0.64	0.68	0.77	yes	yes	ONMLK
342	0.36	0.18	0.32	0.27	0.23	0.18	0.41	0.23	0.18	0.18	yes	NO	KLMNO

The choice proportions for Case #101 are shown in the first row. The entry in the last column of Table 2 indicates that #101 completed 20 blocks. Because each block included two presentations of each choice, proportions are based on 40 responses to each choice problem by Participant #101. All ten choice proportions in the first row are greater than 1/2; therefore, this person’s data are perfectly consistent with WST (indicated by the “yes”) and the transitive order, EDCBA. The choice proportions of #101 are also perfectly consistent with the TI, indicated by the “yes” under TI.

A “NO” displayed for WST or TI in Tables E.1, E.2, or E.3 indicates that choice proportions for a given person are not perfectly compatible with these properties, respectively. These do not represent tests of significance. For example,

proportions for case #104 (fourth row of Table 2) are all greater than  $\frac{1}{2}$ , so this case is perfectly compatible with WST and the order *EDCBA*. However, these choice proportions are not perfectly compatible with the TI, indicated by the “NO” in column TI, because, for example, the choice proportions show that  $P(AB) + P(BE) - P(AE) = 1 + 1 - .97 = 1.03$ , which is not between 0 and 1. The data for #104 are based on 30 responses per choice (15 blocks), so this violation would not have appeared if the one response out of 30 when this person chose *A* over *E* had been different ( $1/30 = .03$ ).

There were 107 people out of 136 (79%) whose choice proportions were perfectly consistent with WST in all three designs. There were only 13, 10, and 10 cases in which WST was not perfectly satisfied in the *LH*, *LP*, and *PH* designs, respectively (10%, 7%, and 7%). There was no one whose proportions violated (were not perfectly consistent with) WST in all three designs; only 4 were not perfectly compatible with WST in two designs (#137, 214, 239, and 328).

Violations of WST can easily occur when a person has a mixture of transitive patterns (Birnbaum & Gutierrez, 2007; Regenwetter, et al., 2010, 2011). Although the TI has the advantage (over WST) that it is consistent with a mixture of transitive orders, the TI can be violated by tiny deviations when a person is otherwise highly consistent with transitivity, and it can also be satisfied when a person has a mixture that includes systematic violations of transitivity (Birnbaum, 2011). Both TI and WST therefore might be misleading when a person has a mixture of preference patterns. When iid is violated, both of these properties can be misleading, and one should examine response patterns.

## Appendix F: Analysis of transitivity in *i*TET model

This section presents an individual TE model for the five choice problems that test the intransitive prediction of the LS models. For example, in the *LP* design, these are the *FG*, *GH*, *HI*, *IJ*, and *FJ* choices. Under an LS model, it is possible to show the intransitive data patterns 22221 (or 11112), which represent observed preferences for  $G \succ F$ ,  $H \succ G$ ,  $I \succ H$ ,  $J \succ I$  but  $F \succ J$ , (or their reverses). All other response patterns are compatible with transitivity. Suppose that within each block of trials, a person has the same “true” preferences but may show random “errors” in discovering or reporting his or her true preferences.

The probability of observing the response pattern 22221 in both tests and the “true” pattern is 22222 on a block is given by the following:

$$P_{22222}(22221, 22221) = p_{22222}(1 - e_1)^2(1 - e_2)^2(1 - e_3)^2(1 - e_4)^2(e_5)^2$$

Where  $P_{22222}(22221, 22221)$  is the probability responding 22221 on both tests and having the true pattern of 22222;  $p_{22222}$  is the probability that the “true” pattern is the transitive order *EDCBA*, and  $e_1, e_2, e_3, e_4$ , and  $e_5$  are the probabilities of “errors” on the five respective choice problems, which are assumed to be mutually independent and less than  $\frac{1}{2}$ . Note that the error terms are squared because this expression calculates the probability of observing the same response pattern on both items within a block (two responses for each of five choice problems).

The overall predicted probability in the TE model for the observed pattern 22221 on both tests,  $P(22221, 22221)$  is the sum of 32 terms including that above for the 32 possible “true” patterns, each with the appropriate error terms to create each response pattern given each true pattern. Transitivity is defined as the special case of this TE model in which the two intransitive probabilities of zero; that is, that  $p_{22221} = p_{11112} = 0$ .

A maximum likelihood solution of the transitive TE model to the 15 blocks of response patterns for case #125 (Table 5) yielded  $e_1 = e_2 = e_3 = 0$ ;  $e_4 = 0.21$ ,  $e_5 = 0.50$ ,  $p_{22211} = 0.038$ , and  $p_{22222} = 0.962$ ; all other parameters were zero. According to this solution,  $P(22221, 22221) = 0.15$ . However, the data showed 7 blocks with repeats of 22221 out of 15 blocks. From the binomial, the probability to observe 7 or more response patterns of 22221 and 22221 out of 15 is .003. Therefore, one can reject the assumption that the “true” probability of 22221 is zero. The binomial in this case assumes only that blocks and errors are independent, it does not assume or imply the stronger iid assumptions that responses within a block are independent (See Appendix H for more on this distinction).

When all parameters are free, the maximum likelihood solution yields  $e_1 = e_2 = e_3 = 0$ ;  $e_4 = 0.21$ ,  $e_5 = 0.10$ ,  $p_{22211} = 0.038$ , and  $p_{22221} = 0.962$ ; all other parameters were zero. In this solution,  $P(22221, 22221) = 0.49$ , which is compatible with the data showing 7 out of 15 repeated intransitive patterns of this type. In sum, the TE model indicates that we should reject the assumption of transitivity for #125 in this design. As a further note, Case #125 also chose  $G \succ J$  on 26 of 30 choices, creating another intransitive cycle of a type consistent with an LS model.

### Appendix G: Tests of interactive independence

Individual choice proportions for the *LS* Design of Experiments 2 and 3 are shown in Table G.1. *X1*, *X2*, *X3*, *X4*, and *X5* refer to the choices between  $R = (\$95, p; \$5, 1 - p)$  and  $S = (\$55, p; \$20, 1 - p)$ , where  $p = 0.95, 0.9, 0.5, 0.1$ , or  $0.05$ , respectively, which test interactive independence. Similarly, *Y1*, *Y2*, and *Y3* in Table 9 represent the choices between  $S = (\$99, p; \$1, 1 - p)$  versus  $R = (\$40, p; \$35, 1 - p)$ , where  $p = 0.9, 0.5$ , or  $0.1$ , averaged over two presentations with either *S* or *R* presented first. These also test interactive independence.

According to any *LS* model, the value of  $p$  should have no effect, because it is the same in both *R* and *S*. Even if subjective probability is a function of objective probability (as in Footnote 1), the common probability term drops out. Under any order of examining the attributes and with any difference thresholds, a person should either choose *R* or *S*, in all choices, for any common  $p$ . Therefore, any mixture of *LS* models should also show no effect of  $p$ .

According to the priority heuristic, a person should always choose *S*, because it has the higher lowest outcome (by \$15 in *X1* to *X5* and by \$34 in *Y1* to *Y3*). According to interactive models such as *TAX*, *CPT* and *EU*, however, as  $p$  decreases, the tendency to choose *S* should increase, because it provides the better lowest consequence, whose value is multiplied by a function of  $1 - p$ . In agreement with predictions of *TAX*, *CPT*, and *EU*, and contrary to all *LS* models (including priority heuristic) and mixtures thereof, the median choice proportions for *S* increase from 0.15 to 0.91 from *X1* to *X5* and from 0.15 to 0.93 in *Y1* to *Y3*. Out of 85 participants, 70 and 72 had  $X5 > X1$  and  $Y3 > Y1$ , respectively.

Table G.1. Individual choice proportions in the *LS* design (Experiments 2 and 3).

sub	<i>X1</i>	<i>X2</i>	<i>X3</i>	<i>X4</i>	<i>X5</i>	<i>Y1</i>	<i>Y2</i>	<i>Y3</i>	<i>Z1</i>	<i>Z2</i>	<i>Z3</i>	<i>Z4</i>	<i>W</i>
201	0.94	0.94	0.94	0.94	0.94	0.95	1.00	1.00	0.00	1.00	0.05	1.00	0.85
<b>202</b>	<b>0.15</b>	<b>0.31</b>	<b>0.69</b>	<b>1.00</b>	<b>1.00</b>	<b>0.17</b>	<b>0.50</b>	<b>0.89</b>	<b>0.56</b>	<b>1.00</b>	<b>0.44</b>	<b>0.78</b>	<b>1.00</b>
203	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.11	0.05	0.53
204	0.00	0.00	0.00	0.90	0.90	0.00	0.00	1.00	0.67	0.83	0.50	0.92	0.92
205	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.98	0.30	0.40	0.45	0.70	0.55
206	0.00	0.00	0.14	0.43	0.29	0.00	0.25	0.94	0.44	0.56	0.19	0.62	0.63
207	0.08	0.00	0.92	1.00	1.00	0.00	0.79	0.96	0.18	0.91	0.27	0.91	0.82
208	0.92	0.85	0.92	1.00	1.00	0.73	0.97	0.93	0.10	1.00	0.00	0.90	1.00
209	0.00	0.00	0.09	0.64	0.64	0.17	0.25	0.42	0.31	0.54	0.38	0.62	0.46
210	0.00	0.00	0.00	0.05	0.05	0.13	0.10	0.10	0.15	0.30	0.80	0.80	0.30
211	0.31	0.69	0.69	0.94	0.81	0.11	0.87	1.00	0.50	0.80	0.40	0.80	0.65
212	0.00	0.00	0.07	0.21	0.21	0.00	0.00	0.25	0.00	0.08	0.23	0.15	0.31
213	0.05	0.00	0.00	0.05	0.11	0.03	0.03	0.10	0.18	0.71	0.12	0.82	0.94
<b>214</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>1.00</b>	<b>1.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.95</b>	<b>0.55</b>	<b>0.82</b>	<b>0.73</b>	<b>1.00</b>	<b>1.00</b>
215	1.00	1.00	1.00	1.00	0.89	1.00	1.00	1.00	0.80	0.90	0.95	0.95	0.90
216	0.13	0.27	1.00	1.00	1.00	0.09	1.00	1.00	0.12	1.00	0.00	1.00	1.00
217	0.28	0.17	0.94	1.00	0.89	0.04	1.00	1.00	0.33	0.78	0.28	0.94	0.94
<b>218</b>	<b>0.00</b>	<b>0.17</b>	<b>0.50</b>	<b>0.83</b>	<b>0.83</b>	<b>0.09</b>	<b>0.23</b>	<b>0.95</b>	<b>0.44</b>	<b>0.89</b>	<b>0.44</b>	<b>0.89</b>	<b>1.00</b>
219	0.88	0.94	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	0.89	1.00	0.78
220	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.60	1.00	0.60	1.00	1.00
221	0.00	0.00	0.10	0.10	0.20	0.00	0.00	0.05	0.00	0.10	0.00	0.20	1.00
222	0.10	0.20	0.30	0.60	0.40	0.13	0.25	0.50	0.33	0.56	0.33	0.44	0.89
223	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	0.15	1.00	0.05	1.00	0.85
224	0.77	0.62	1.00	1.00	1.00	0.71	1.00	1.00	0.13	0.88	0.12	0.87	0.88
225	0.95	1.00	1.00	1.00	0.95	1.00	1.00	1.00	0.15	1.00	0.05	1.00	1.00
226	0.00	0.15	0.85	1.00	1.00	0.07	0.96	1.00	0.75	1.00	0.94	0.94	0.88
227	0.21	0.00	0.21	0.79	1.00	0.00	0.06	0.88	0.69	0.77	0.62	0.77	0.54

sub	X1	X2	X3	X4	X5	Y1	Y2	Y3	Z1	Z2	Z3	Z4	W
228	0.05	0.10	0.35	0.95	0.95	0.03	0.21	0.76	0.68	1.00	0.74	0.95	1.00
229	0.06	0.11	0.11	0.78	0.78	0.09	0.09	0.91	0.88	0.63	0.81	0.75	1.00
230	0.80	1.00	0.90	1.00	0.80	0.85	0.80	1.00	0.75	0.88	0.62	0.87	0.75
231	0.71	0.86	1.00	1.00	0.86	0.30	0.90	0.90	0.50	0.83	0.33	1.00	1.00
232	0.00	0.07	0.07	0.14	0.14	0.11	0.18	0.18	0.13	0.13	0.31	0.25	0.25
233	1.00	0.95	1.00	1.00	0.95	0.95	1.00	1.00	0.05	0.95	0.10	1.00	1.00
234	0.93	0.93	0.93	1.00	1.00	1.00	1.00	1.00	0.27	1.00	0.07	0.93	1.00
235	0.81	0.94	1.00	1.00	1.00	0.85	1.00	1.00	0.36	0.57	0.43	0.64	0.43
236	0.20	0.30	1.00	1.00	0.90	0.21	0.92	1.00	0.17	1.00	0.25	1.00	0.83
237	0.05	0.00	0.15	0.80	0.85	0.08	0.15	0.93	0.00	0.00	0.05	0.15	0.90
238	0.77	0.92	1.00	1.00	1.00	0.82	1.00	1.00	1.00	0.80	1.00	1.00	0.70
<b>239</b>	<b>0.42</b>	<b>0.67</b>	<b>0.92</b>	<b>0.92</b>	<b>1.00</b>	<b>0.15</b>	<b>1.00</b>	<b>1.00</b>	<b>0.60</b>	<b>0.50</b>	<b>0.40</b>	<b>0.60</b>	<b>0.80</b>
240	0.00	0.00	0.00	0.35	0.45	0.00	0.03	0.15	0.00	0.35	0.00	0.20	0.20
241	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.03	0.00	0.05	0.00	0.05	0.00
242	0.08	0.31	1.00	1.00	1.00	0.54	1.00	1.00	0.40	0.70	0.30	0.90	1.00
243	0.00	0.00	0.13	0.80	0.87	0.08	0.13	0.83	0.23	0.46	0.23	0.77	0.69
301	0.55	0.40	0.60	0.40	0.60	0.23	0.33	0.33	0.15	0.40	0.05	0.35	0.85
302	0.30	0.00	0.80	1.00	0.80	0.05	0.95	1.00	0.50	0.40	0.60	0.80	1.00
303	0.45	0.64	1.00	0.91	1.00	0.59	1.00	1.00	0.73	0.73	0.91	0.91	0.64
304	0.10	0.30	1.00	1.00	1.00	0.35	1.00	1.00	0.60	0.80	0.50	0.60	1.00
305	0.63	0.74	0.89	0.95	1.00	0.97	0.97	1.00	0.37	1.00	0.26	1.00	0.63
306	0.00	0.00	0.09	0.09	0.18	0.00	0.05	0.18	0.00	0.18	0.00	1.00	0.64
307	0.91	0.82	1.00	1.00	0.91	1.00	1.00	0.95	0.55	1.00	0.36	0.73	0.82
308	0.00	0.00	0.00	0.80	0.90	0.15	0.00	0.90	0.90	0.30	1.00	0.10	1.00
<b>309</b>	<b>0.60</b>	<b>0.80</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>0.70</b>	<b>1.00</b>	<b>0.95</b>	<b>0.30</b>	<b>0.90</b>	<b>0.60</b>	<b>1.00</b>	<b>1.00</b>
310	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.05	0.05	0.10
<b>311</b>	<b>0.00</b>	<b>0.08</b>	<b>0.85</b>	<b>1.00</b>	<b>1.00</b>	<b>0.12</b>	<b>0.96</b>	<b>1.00</b>	<b>0.85</b>	<b>0.92</b>	<b>0.54</b>	<b>0.85</b>	<b>0.54</b>
312	0.05	0.05	0.15	0.05	0.20	0.10	0.20	0.20	0.00	0.15	0.05	0.25	0.70
313	0.10	0.00	0.00	0.20	0.40	0.05	0.00	0.30	0.30	0.20	0.10	0.30	0.40
314	0.53	0.76	0.88	0.76	0.94	0.76	0.85	0.79	0.76	0.65	0.53	0.82	0.82
315	0.06	0.00	1.00	0.94	1.00	0.03	0.97	1.00	0.00	0.94	0.17	1.00	0.78
316	0.93	1.00	0.86	0.93	0.93	0.96	0.96	1.00	0.21	0.93	0.57	0.93	0.07
317	0.13	0.31	0.56	0.69	0.75	0.09	0.50	0.81	0.63	0.88	0.25	0.69	0.81
318	0.00	0.00	0.00	0.80	0.90	0.00	0.05	0.85	0.50	0.50	0.70	0.80	0.70
319	0.82	0.88	0.94	1.00	1.00	0.97	0.97	0.97	0.24	1.00	0.06	1.00	0.94
320	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.95	0.90	0.55	0.40
321	0.16	0.16	0.26	0.42	0.37	0.21	0.29	0.71	0.68	0.53	0.74	0.74	0.79
322	0.00	0.00	0.00	0.30	0.30	0.05	0.05	0.40	0.00	0.40	0.20	0.60	0.70
323	0.00	0.10	1.00	1.00	1.00	0.15	1.00	1.00	0.60	0.90	0.80	1.00	0.90
324	0.00	0.00	0.05	0.25	0.30	0.13	0.13	0.20	0.35	0.45	0.70	0.85	0.30
325	0.00	0.00	0.00	0.56	0.50	0.00	0.25	0.78	0.19	0.38	0.06	0.56	1.00
326	0.00	0.00	0.00	0.90	1.00	0.00	0.00	1.00	0.70	0.80	0.60	0.90	1.00
327	0.06	0.25	1.00	1.00	1.00	0.47	1.00	0.97	0.25	1.00	0.50	1.00	0.75
328	0.07	0.07	0.33	0.27	0.33	0.13	0.03	0.53	0.27	0.47	0.27	0.80	0.60
329	0.11	0.00	0.00	0.17	0.22	0.47	0.31	0.56	0.06	0.28	0.06	0.44	0.89

sub	X1	X2	X3	X4	X5	Y1	Y2	Y3	Z1	Z2	Z3	Z4	W
330	0.87	1.00	0.87	1.00	0.93	0.73	0.97	0.97	0.20	0.67	0.20	1.00	0.87
331	0.27	0.09	0.27	0.73	0.64	0.14	0.45	0.73	0.45	0.64	0.64	0.73	0.55
332	0.10	0.10	0.30	0.60	0.50	0.05	0.25	0.55	0.30	0.70	0.10	0.70	1.00
333	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.08	0.00	0.00	0.00	0.15
334	0.38	0.38	0.50	0.63	0.50	0.16	0.44	0.59	0.56	0.50	0.25	0.44	0.63
335	0.62	0.54	1.00	0.92	1.00	0.23	0.85	0.96	0.31	1.00	0.38	0.85	1.00
336	0.25	0.08	0.50	0.08	0.08	0.21	0.17	0.38	0.33	0.50	0.25	0.92	0.58
337	0.91	0.91	1.00	1.00	1.00	1.00	1.00	1.00	0.09	1.00	0.09	0.91	0.27
<b>338</b>	<b>0.18</b>	<b>0.18</b>	<b>0.24</b>	<b>0.12</b>	<b>0.29</b>	<b>0.29</b>	<b>0.21</b>	<b>0.18</b>	<b>0.41</b>	<b>0.35</b>	<b>0.18</b>	<b>0.29</b>	<b>0.65</b>
339	0.50	0.50	0.92	0.67	0.92	0.58	0.58	0.79	0.25	0.42	0.50	0.92	0.92
340	0.64	0.18	0.45	0.64	0.82	0.62	0.73	0.77	0.55	0.55	0.64	0.64	0.55
341	0.55	0.55	0.73	0.82	0.64	0.09	0.14	0.14	0.45	0.73	0.27	0.91	0.73
342	0.82	0.91	1.00	1.00	1.00	0.50	0.68	0.68	0.09	1.00	0.00	0.91	0.91

In Table G.1, Z1 and Z2 represent choices between  $R = (\$90, 0.05; \$88, 0.05; \$2, 0.9)$  versus  $S = (\$45, 0.2; \$4, 0.2; \$2, 0.6)$  and between  $R+ = (\$90, 0.1; \$3, 0.7; \$2, 0.2)$  versus  $S- = (\$45, 0.1; \$44, 0.1; \$2, 0.8)$ , respectively. Note that  $R+$  stochastically dominates  $R$  and  $S$  stochastically dominates  $S-$ . According to CPT,  $R \succ S \Rightarrow R+ \succ S-$ . According to the priority heuristic, most people should choose  $S \succ R$  (because of the smaller probability to receive the lowest consequence) and  $R+ \succ S-$  (again because of the probabilities to receive the lowest outcome).

According to TAX with its prior parameters, however, a person would have the opposite preferences:  $R \succ S$  and  $S- \succ R+$ . Consistent with TAX and contrary to CPT, priority heuristic, EU, and EV, the median choice proportions show 69% preference for  $R \succ S$  and 73% preference for  $S- \succ R+$ .

Z3 and Z4 present a similar test with positions counterbalanced:  $R2 = (\$80, 0.1; \$78, 0.1; \$3, 0.8)$  versus  $S2 = (\$40, 0.4; \$5, 0.1; \$4, 0.5)$  and  $R3+ = (\$80, 0.2; \$4, 0.7; \$3, 0.1)$  versus  $S3- = (\$40, 0.2; \$39, 0.2; \$3, 0.5)$ . In this test,  $R3+$  dominates  $R2$  and  $S2$  dominates  $S3-$ ; CPT implies that  $R2 \succ S2 \Rightarrow R3+ \succ S3-$ . The priority heuristic implies that the majority should choose  $S2 \succ R2$  and  $R3+ \succ S3-$ . However, median choice percentages again contradict CPT, priority heuristic, EU, and EV: median choice percentages are 70% choosing  $R2 \succ S2$  and 82% choosing  $S3- \succ R3+$ , showing the reversal predicted by TAX with prior parameters.

These results are also representative of the majority of individuals: Of the 85 participants in Experiments 2 and 3, 62 (73%) showed both  $Z1 < Z2$  and  $Z3 < Z4$ , which is significantly more than half of the sample ( $z = 4.23$ ).

The last column, W, in Table G.1 shows the proportion of responses favoring  $F4 = (\$88, 0.12; \$86, 0.7; \$3, 0.18)$  over  $G4 = (\$99, 0.3; \$15, 0.65; \$14, 0.05)$ . According to the priority heuristic, people should choose  $G4$  because the lowest outcome is better by \$11. In addition, notice that the probability to receive the lowest consequence of  $G4$  is also better (by 0.13), as is the highest consequence of  $G4$  (by \$11), as is the probability to win the highest consequence (by 0.18). Thus, all four of the features that are compared in the priority heuristic favor  $G4$ . However, the median choice proportion was 0.82 for  $F4$ , and only 14 of 85 individuals (16%) chose  $G4$  over  $F4$  half or more than half of the time. These 71 people (84% who chose  $F4$  over  $G4$  more than half the time) represents significantly more than half of all participants,  $z = 6.18$ ), contradicting priority heuristic and other LS models using these four variables with the same thresholds. The TAX model with its prior parameters correctly predicted this result.

The most promising cases in Experiments 2 and 3 for evidence of a LS model are #202, 214, 218, 239, 309, 311, and 338, who either showed response patterns consistent with intransitive LS models, or who violated WST and TI, or both. Those cases (marked in bold font in Table G.1 and described in the main text) show that even these people appear to systematically violate implications of the LS models. Consequently, most people violate the implications of the family of LS models, and no one showed evidence of intransitivity who also appeared to satisfy the LS models in these tests of



## Appendix H: Fit of True and Error and IID models to response patterns

The frequencies of response patterns from Tables D.1, D.2, and D.3 were fit to two models. The iid model assumes that the probability of showing a response pattern is the product of the marginal probabilities of showing each response. The TE model assumes that each block may have a different “true” pattern and independent “errors”. Both of these models allow transitive and intransitive patterns, and both models are oversimplified, but their application is instructive.

Table H.1 shows the fit of these two models to the *LH* design, aggregated over three experiments. “Rows” indicates the number of times that a person showed each response pattern when the gambles were presented with the alphabetically higher gamble first, and “Cols” shows the frequency when the same choice problem was presented with the gambles reversed. “Both” indicates cases where a person made the same choice responses on both presentations within a block. The U–C values represent the average frequency of showing the response pattern in either one arrangement or the other but not both. The models were fit to minimize the Chi-Square between the obtained and predicted frequencies in these  $2 \times 32 = 64$  cells.

The iid model assumes that the probability of showing a response pattern is simply the product of the individual choice probabilities. Consequently, the predicted frequency of response pattern 11111 is proportional to the product of the marginal choice proportions of choosing Response 1 in the five choice problems making up this pattern:  $P(AB)P(BC)P(CD)P(DE)P(AE)$ , where  $P(AB)$  is the proportion choosing *A* in the choice problem between *A* and *B*.

The iid model fails badly because people are more consistent than this model allows them to be. When a response pattern of 11111, for example, is observed within a block in one arrangement, it is highly likely that the same response pattern is observed when the stimuli are presented in the other arrangement within the same block, even though this requires pressing opposite buttons on randomly ordered trials. Consequently, responses agree within blocks to a much greater degree than predicted by the iid model.

For example, the response pattern 11111 occurred in the *LH* design 687 and 694 times in the two arrangements, and this same response pattern was shown with both arrangements 537 times within blocks. According to iid, however, the predicted frequency of showing this pattern in both versions of a block is only 7.4. Summed over response patterns, the iid model implies that out of the 2037 blocks, there should have been agreement in only 74 cases where a response pattern was repeated twice within a block. Instead, the actual number of cases where response patterns were consistent within blocks was 1038.

To understand why independence implies so little self-agreement, keep in mind that each response pattern is built of five responses. Therefore, ten responses have to fall in place to produce a match. The marginal probabilities to respond “1” on the five choices are 0.54, 0.54, 0.51, 0.61, and 0.66, respectively, so the probability to show all ten responses—assuming iid—is the product of these values squared, which is only .0036. This figure is much smaller than observed  $.26 = 537/2037$ , so the assumption of iid is not an accurate description of this behavior.

The TE model does a better job since it predicts agreement in 994 cases (where the actual was 1038). The TE model violates independence because it assumes that the true preferences are the same within blocks and only the errors are independent. Note that the data reveal even greater dependence than predicted by this TE model. The estimated error rates in the TE model in the *LH* design were 0.10, 0.06, 0.09, 0.06, and 0.04 for the five choice problems, respectively. The estimated “true” rate for the 11111 pattern was 0.452.

The assumption that errors are independent means that the probability to show the “true” pattern in both repetitions is the product of the probabilities of not making an error on all ten choice problems. Subtract each error term from one, square it, and find the product, which for the pattern 11111 is 0.472. If the “true” probability of the pattern 11111 is 0.452, the probability to repeat this true pattern is then 0.213, which is much greater than that predicted by iid and closer to the observed proportion. (The TE model also allows that a person can show this pattern by having other “true” patterns and making the required errors to match this pattern, but these only contribute a very small amount to the prediction in this case).

Tables D.5 and D.6 show predicted frequencies of these two models for the *LP* and *PH* designs, respectively. These cases again show that the iid model is not accurate at all in predicting self-agreement within blocks and the TE model does better but is not completely accurate.

The Chi-Squares of fit for the iid model are (obviously) off the charts, ranging from over 350,000 to more than 1.5 million. The Chi-Squares for the TE model with 27 df are 245.2, 253.2, and 299.7, all significant. The Chi-Squares testing the special case of TE that assumes transitivity, with 2 df, are 116.2, 251.7, and 37.2. The estimated rates of intransitivity are as follows: In the *LH* design,  $p_{22221} = .03$ ; in the *LP* design,  $p_{22221} = .03$ , and in the *PH* design,  $p_{11112} = .01$ . Assuming this TE model (which is dubious due to its lack of fit), one would conclude that these tiny rates of intransitivity are significant.

This TE analysis likely understates the dependence in the data because it does not properly handle individual differences, which contribute to its lack of fit. Despite that caveat, however, the estimated rates of intransitivity are similar to the rates based on separate analyses of individual data. In other words, besides the individuals identified as intransitive in Table 3 and Appendix D, there little additional evidence, if any, for mixtures containing partial or temporary intransitivity of these types in the individual block data by others.

Table H.1. Fit of Two Models to Frequency Data of *LH* Design. TE = True and Error model; IID = Independent and Identically Distributed model. Both models allow intransitivity. U–C = average frequency of a response pattern in either position arrangement but not both. Both = frequency of showing the same response pattern in both arrangements.

Pattern	Data				TE Model		IID Model	
	Rows	Cols	Both	U–C	Pred both	Pred U–C	Pred both	Pred U–C
11111	687	694	537	153.5	434.6	205.4	7.4	115.0
11112	11	10	0	10.5	0.9	28.2	2.0	61.1
11121	34	21	1	26.5	2.0	41.6	3.1	76.4
11122	8	2	0	5	0.1	3.3	0.8	40.1
11211	103	124	32	81.5	39.7	81.7	6.8	110.7
11212	10	11	0	10.5	0.8	6.1	1.8	58.7
11221	17	22	1	18.5	2.9	11.9	2.9	73.4
11222	11	4	0	7.5	0.1	4.5	0.8	38.6
12111	67	60	15	48.5	17.2	54.0	5.2	98.2
12112	13	9	0	11	6.8	6.8	1.4	51.9
12121	19	6	0	12.5	3.8	7.7	2.2	64.9
12122	22	5	1	12.5	2.0	7.6	0.6	34.0
12211	50	53	9	42.5	22.4	26.8	4.8	94.4
12212	9	11	0	10	0.2	6.7	1.3	49.9
12221	22	20	1	20	4.2	12.7	2.0	62.4
12222	41	36	3	35.5	3.9	46.5	0.5	32.7
21111	71	76	8	65.5	9.0	73.4	5.2	97.8
21112	10	4	0	7	0.0	4.5	1.4	51.7
21121	21	14	0	17.5	3.9	9.7	2.2	64.7
21122	15	8	1	10.5	4.7	6.3	0.6	33.9
21211	56	53	10	44.5	21.5	28.1	4.8	94.0
21212	10	6	0	8	0.8	4.9	1.3	49.7
21221	20	29	3	21.5	13.9	14.6	2.0	62.2
21222	21	25	3	20	3.7	30.7	0.5	32.6
22111	39	32	8	27.5	13.4	19.7	3.7	83.3
22112	7	11	0	9	0.2	6.0	1.0	43.9
22121	18	12	0	15	0.5	9.8	1.6	55.0
22122	31	42	6	30.5	7.2	43.8	0.4	28.7
22211	69	47	11	47	39.8	30.4	3.4	80.1
22212	31	26	3	25.5	3.6	31.3	0.9	42.2
22221	62	70	22	44	33.0	39.3	1.4	52.8
22222	432	494	363	100	297.6	138.9	0.4	27.6
Totals	2037	2037	1038	999	994.1	1042.9	74.4	1962.6

Table H.2. Fit of two models to frequency data *LP* design.

Pattern	Data				TE Model		IID Model	
	Rows	Cols	Both	U-C	Pred both	Pred U-C	Pred both	Pred U-C
11111	775	716	620	125.5	513.4	179.3	4.2	87.9
11112	24	12	3	15	3.4	38.4	2.0	61.9
11121	20	36	0	28	2.0	43.6	2.9	73.6
11122	20	14	2	15	15.6	8.1	1.4	51.6
11211	45	53	5	44	6.3	48.4	3.9	84.8
11212	13	8	1	9.5	4.2	5.4	1.9	59.7
11221	26	34	6	24	23.9	13.0	2.7	71.0
11222	12	9	0	10.5	0.2	6.2	1.3	49.8
12111	28	37	0	32.5	1.5	37.5	3.0	75.6
12112	6	10	0	8	4.7	4.5	1.5	53.1
12121	7	7	1	6	2.0	3.6	2.1	63.2
12122	7	9	0	8	0.1	4.7	1.0	44.3
12211	17	14	2	13.5	5.6	7.8	2.8	73.0
12212	14	8	1	10	5.0	5.7	1.4	51.2
12221	13	9	0	11	0.3	8.0	1.9	61.0
12222	39	36	4	33.5	4.4	40.3	0.9	42.7
21111	77	57	16	51	18.4	57.1	3.8	84.4
21112	10	11	0	10.5	3.0	5.7	1.8	59.4
21121	13	14	2	11.5	3.5	7.3	2.6	70.6
21122	9	6	0	7.5	0.1	4.5	1.3	49.5
21211	27	32	5	24.5	16.5	14.0	3.6	81.5
21212	10	8	0	9	3.0	5.2	1.7	57.3
21221	23	27	4	21	13.5	12.7	2.4	68.1
21222	26	28	4	23	4.6	30.4	1.2	47.8
22111	29	29	10	19	18.3	12.5	2.8	72.6
22112	13	3	0	8	0.1	4.6	1.3	51.0
22121	8	16	0	12	0.2	6.9	1.9	60.7
22122	31	19	3	22	3.3	35.0	0.9	42.5
22211	29	35	4	28	24.6	16.1	2.6	70.1
22212	34	39	1	35.5	1.6	34.6	1.2	49.1
22221	91	58	25	49.5	33.3	42.5	1.8	58.5
22222	539	640	475	114.5	410.3	144.3	0.9	41.0
Totals	2035	2034	1194	840.5	1147.1	887.9	66.7	1968.3

Table H.3. Fit of two models to frequency data *PH* design.

Pattern	Data				TE Model		IID Model	
	Rows	Cols	Both	U-C	Pred both	Pred U-C	Pred both	Pred U-C
11111	450	432	365	76	314.0	103.4	0.1	12.7
11112	31	34	8	24.5	11.1	25.0	0.3	23.5
11121	40	24	1	31	1.7	30.5	0.2	22.2
11122	13	13	0	13	9.2	7.1	0.9	41.0
11211	25	29	1	26	1.4	26.4	0.2	22.2
11212	14	13	0	13.5	13.0	7.4	0.9	40.8
11221	13	13	2	11	4.9	6.0	0.8	38.7
11222	12	11	1	10.5	2.3	6.4	2.7	70.7
12111	28	20	2	22	2.5	28.9	0.2	21.4
12112	16	13	0	14.5	11.3	7.7	0.8	39.5
12121	10	16	0	13	5.1	6.7	0.7	37.5
12122	15	18	0	16.5	12.4	9.3	2.5	68.5
12211	17	6	0	11.5	4.4	5.8	0.7	37.3
12212	13	12	1	11.5	2.9	6.9	2.5	68.3
12221	29	26	11	16.5	24.0	11.2	2.2	64.8
12222	58	35	5	41.5	5.7	45.4	7.7	117.0
21111	23	24	1	22.5	1.4	23.7	0.2	21.5
21112	25	15	0	20	25.4	11.2	0.8	39.6
21121	14	9	0	11.5	11.4	6.2	0.7	37.5
21122	17	15	0	16	3.9	8.6	2.5	68.6
21211	10	15	1	11.5	12.1	6.1	0.7	37.4
21212	14	23	1	17.5	5.3	9.5	2.5	68.4
21221	6	14	1	9	2.2	5.7	2.2	64.9
21222	55	54	3	51.5	3.5	57.5	7.7	117.1
22111	10	8	1	8	6.1	4.2	0.7	36.2
22112	10	16	1	12	1.7	7.9	2.3	66.3
22121	8	10	0	9	2.4	5.2	2.1	62.8
22122	47	45	4	42	4.4	52.6	7.2	113.6
22211	8	7	0	7.5	0.1	4.7	2.1	62.6
22212	46	59	2	50.5	3.4	60.5	7.2	113.2
22221	34	37	0	35.5	1.3	38.2	6.4	107.5
22222	913	958	800	135.5	661.4	216.1	22.2	189.8
Totals	2024	2024	1212	812	1171.9	852.1	90.9	1933.1