RATIONALITY OR IRRATIONALITY OF PREFERENCES?
A QUANTITATIVE TEST OF INTRANSITIVE DECISION HEURISTICS

BY

YING GUO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Master's Committee:

Professor Michel Regenwetter, Advisor
Assistant Professor Hans-Friedrich Köhn

Abstract

In this paper, I present a comprehensive analysis of two decision heuristics that permit intransitive preferences: the lexicographic semiorder model and the similarity model. I also compare these two intransitive decision heuristics with transitive linear order models and two simple transitive heuristics. For each decision theory, I use two types of probabilistic specifications: distance-based models (which assume deterministic preferences and probabilistic response processes), and mixture models (which assume probabilistic preferences and deterministic response processes). I test 26 such probabilistic models on datasets from three different experiments using both frequentist and Bayesian order-constrained statistical methods. The frequentist goodness-of-fit tests show that the distance-based models with modal choice and the mixture models for all of the decision heuristics explain the participants' data fairly well for all stimulus sets. The frequentist analysis generates little evidence against transitivity. Model selection using Bayes factors suggests extensive heterogeneity across participants and stimulus sets.

To my family.

Acknowledgments

I thank Dr. Michel Regenwetter for his insights and assistance in the preparation of this paper. I thank Alice Huang for her help with my writing skills. I acknowledge funding through Dr. Regenwetter's National Science Foundation (NSF) grants SES # 08-20009, # 10-62045, and #14-59694, his 2009 Arnold O. Beckman Research Award from University Research Board of the University of Illinois at Urbana-Champaign, and his Extreme Science and Engineering Discovery Environment (XSEDE) grant SES #130016 (PI: M. Regenwetter). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and need not reflect the views of NSF, XSEDE, or the University of Illinois.

Table of Contents

List of Abbreviations . . . . **vi**

Chapter 1 Rationality or Irrationality of Preferences? A Quantitative Test of Intransitive Decision Heuristics **1**

1.1 Introduction 1
1.2 Intransitive Heuristic Models 2
1.2.1 Lexicographic Semiorder Models 2
1.2.2 Similarity Models 4
1.3 Transitive Models 6
1.3.1 Linear Order Models 6
1.3.2 Two Simple Transitive Heuristics 6
1.4 Probabilistic Specifications 6
1.4.1 Distance-Based Models 7
1.4.2 Mixture Models 8
1.4.3 Summary of Models 11
1.4.4 Statistical Methods 11
1.5 Experiments 12
1.6 Results 14
1.6.1 Distance-Based Model Results 14
1.6.2 Mixture Model Results 16
1.6.3 Model Comparison: Individual Level 18
1.6.4 Model Comparison: Group Level 19
1.7 Conclusions and Discussions 20
1.8 Tables 22
1.9 Supplement Materials 35

Chapter 2 Parsimonious Testing of Transitive or Intransitive Preferences: Reply to Birnbaum (2011) **57**

References **68**

List of Abbreviations

2AFC	Two-alternative forced-choice.
BF	Bayes factor.
GBF	Group Bayes factor.
LO	The linear order model.
LSO-Diff	The lexicographic semiorder model using an identity function $u(x) = x$ for utility.
LSO-Ratio	The lexicographic semiorder model using a log function $u(x) = log(x)$ for utility.
Payoff-only	The decision heuristic according to which any option with a larger reward is preferred to any option with a smaller reward.
Prob-only	The decision heuristic according to which any option with a larger probability of winning is preferred to any option with a smaller probability of winning.
SIM-Diff	The similarity model using an identity function $u(x) = x$ for utility.
SIM-Ratio	The similarity model using a log function $u(x) = log(x)$ for utility.

Chapter 1

Rationality or Irrationality of Preferences? A Quantitative Test of Intransitive Decision Heuristics

1.1 Introduction

To have *transitive preferences*, for any options x, y, and z, one who prefers x to y and y to z must prefer x to z. Transitivity of preferences plays an important role in many major contemporary theories of decision-making under risk or uncertainty, including nearly all normative, prescriptive, and even descriptive theories. Most theories use an overall utility value for each gamble and assume that a decision maker prefers gambles with higher utility values; in other words, most theories imply transitivity of preferences. These theories include expected utility theory (Bernoulli, 1738), prospect theory (Kahneman and Tversky, 1979), cumulative prospect theory (Tversky and Kahneman, 1992), and decision field theory (Busemeyer and Townsend, 1993). Transitivity of preferences is a fundamental element of utility, and abandoning it means questioning nearly all theories that rely on this element. Moreover, transitivity of preferences is important because when a decision maker's preferences are not transitive (i.e., *intransitive* or *irrational*), he risks becoming a "money pump" (Bar-Hillel and Margalit, 1988; Block et al., 2012) and losing his entire wealth.

In the past few decades, researchers have provided a great deal of empirical evidence that suggests that both human and animal decision makers violate transitivity of preferences (see, e.g., Tversky, 1969; Loomes and Sugden, 1987; Brandstätter et al., 2006; González-Vallejo, 2002). However, these studies contain pervasive methodological problems in collecting, modeling, and analyzing empirical data. Some common problematic approaches are pattern counting, pattern counting with hypothesis testing in which the hypotheses are wrongly specified, conducting multiple binomial tests, and using between-participant modal choice (see Section 2 of Guo (2018) for details on these methodological problems). Thus, there is still little evidence of intransitivity (Regenwetter et al., 2011a; Regenwetter and Davis-Stober, 2012; Davis-Stober et al., 2015). Transitivity of preferences is central to many prominent theories in psychology and economics, and we have to be very careful about claiming violations of transitivity of preferences. This paper reviews and tests two prominent intransitive decision heuristics, and compares these intransitive heuristics to the transitive linear order model and two simple transitive heuristics to find out if transitivity of preferences is violated and which model can best explain participants' behavior.

The rest of the paper is organized as follows: Section 1.2 describes two intransitive decision heuristics: lexicographic semiorder models and similarity models; Section 1.3 describes the transitive linear order model and introduces two simple transitive heuristics; Section 1.4 introduces two kinds of probabilistic specifications for the algebraic models: distance-based models and mixture models. It also describes the statistical tools; Section 1.5 describes the five stimulus sets used in this paper: Experiment I in Tversky (1969), Cash I and Cash II in Regenwetter et al. (2011a), and Session I and Session II in an experiment I conducted in 2012; Section 1.6 reports the data analysis results and Section 1.7 concludes the paper.

1.2 Intransitive Heuristic Models

In this section, I describe two intransitive heuristics, including the lexicographic semiorder model (Tversky, 1969) and the similarity model (Rubinstein, 1988; Leland, 1994). These two intransitive heuristics are illustrated using Tversky's (1969) stimulus set (see Panel A of Table 1.1). Tversky's stimulus set comprises five different gambles: a, b, c, d, and e. For example, Gamble a is written as $\left(\$5, \frac{7}{24}; \$0, \frac{17}{24}\right)$, which states that a decision maker has a $\frac{7}{24}$ chance of winning \$5 and a $\frac{17}{24}$ chance of winning nothing. The gambles are designed such that the expected values increase in the probabilities of winning, whereas they decrease in the payoffs. The probability of winning of each gamble increases in equal steps $\left(\frac{1}{24}\right)$, whereas the payoff of the corresponding gambles decreases in equal steps (\$0.25). Employing these gambles, Tversky attempted to learn whether intransitive preferences could be produced and whether the participants would satisfy a lexicographic semiorder model.

1.2.1 Lexicographic Semiorder Models

Tversky (1969) defined a *lexicographic semiorder model* as follows: a semiorder (Luce, 1956) or a just noticeable difference structure is imposed on a lexicographic ordering. Lexicographic semiorder models predict transitive and intransitive preferences.

A lexicographic semiorder works as follows. Suppose a decision maker is asked to choose between two alternatives x and y, where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. I use $x \succ_i y$ to denote that a decision maker prefers x to y on attribute i; I use $x \prec_i y$ to denote that the decision maker prefers y to x on attribute i; and I use $x \sim_i y$ to denote that the decision maker is indifferent between x and y on attribute i. I write $\succ$ for strict preference and $\sim$ for indifference. According to a lexicographic semiorder model:

1. The decision maker considers gamble attributes sequentially, for example, first payoffs and then proba-

bilities of winning, or first probabilities of winning and then payoffs. For each attribute i, the decision maker uses a threshold ϵ_i, and $\epsilon_i > 0$.

2. The decision maker stops the pairwise comparison decision process between two gambles whenever the values of the currently considered attribute i differ by more than the threshold ϵ_i. He then prefers the more attractive gamble on that attribute (either $x \succ_i y$ or $x \prec_i y$). Otherwise, the decision maker has no preference on that attribute ($x \sim_i y$) and proceeds to the next attribute $i + 1$.

3. If the decision maker cannot decide after comparing these two gambles for all attributes (i.e., the values on all attributes do not differ by more than their corresponding thresholds), then he is indifferent between x and y, that is, $x \sim y$.

Consider the ten gamble pairs that comprise all possible pairwise combinations of the five gambles in Tversky (1969). In Tversky's study, each gamble was displayed as a wheel of chance in which a shaded area represented the probability of winning and in which the value of payoff was shown on top of the shaded area. Because the probabilities were not displayed in the numerical form, it was not possible for decision makers to calculate the exact expected values. Tversky (1969) predicted that for "adjacent pairs," that is, for pairs (a, b), (b, c), (c, d), and (d, e), decision makers would prefer gambles with higher payoffs, because the probabilities of winning were visually very similar. In other words, the differences in the probabilities of winning may not have exceeded their thresholds. For the extreme pair, pair (a, e), however, he predicted that decision makers would prefer the gamble with higher probability of winning, because the difference in the probabilities would be large enough to exceed the corresponding threshold and the decision maker would determine his preference before even considering the reward sizes.

An example may serve to further clarify how a lexicographic semiorder model works. Assume that a decision maker considers, in order, first probabilities of winning and then payoffs for the ten gamble pairs in Tversky's stimulus set. Suppose that he uses an identity function for all attribute values, $u(x) = x$, and he uses $\frac{3.5}{24}$ as the threshold for the probabilities of winning for all pairs. Panel B in Table 1.1 shows the differences of probabilities of winning in all ten pairs in Tversky (1969). It shows that the decision maker prefers e to a for pair (a, e) based on the probability of winning, because the probability difference is $\frac{4}{24}$, larger than the threshold. For the remaining pairs, he does not have a preference based on the probability of winning, so he moves on to the next attribute, the payoff. Suppose he uses \$0.35 as the threshold for payoffs. Panel C in Table 1.1 shows the payoff difference in each pair. It shows that for pairs (a, c), (a, d), (b, d), (b, e), and (c, e), the differences between the payoffs exceed \$0.35; therefore, he prefers the gambles with higher payoffs for those pairs. For adjacent pairs, pairs (a, b), (b, c), (c, d), and (d, e), he still cannot make decisions after comparing the values of the two possible attributes; thus, he is indifferent on those pairs.

In Table 1.1, the table on the left side of Panel D shows one of the decision maker's binary preference relations (a *preference pattern*) for the ten gamble pairs in Tversky (1969)—if he uses a lexicographic semiorder model, considers the probability of winning before the payoff, uses a probability threshold of $\frac{3.5}{24}$, and a payoff threshold of \$0.35. The preference pattern for the ten gamble pairs is $a \sim b$, $a \succ c$, $a \succ d$, $a \prec e$, $b \sim c$, $b \succ d$, $b \succ e$, $c \sim d$, $c \succ e$, and $d \sim e$. In particular, $a \succ c$, $c \succ e$, and $e \succ a$ forms an intransitive preference cycle.

For any pair (x, y), the binary choice probability θ_{xy} is the probability of choosing x over y. When a decision maker strictly prefers x to y and performs deterministically, he chooses x over y all the time ($\theta_{xy} = 1$); when a decision maker prefers y to x and choose deterministically, he never chooses x over y ($\theta_{xy} = 0$); when a decision maker is indifferent about x and y, suppose for now, for simplicity, that he chooses x or y with probability one half ($\theta_{xy} = \frac{1}{2}$). The table on the right side of Panel D depicts the binary choice probabilities of a decision maker whose preference pattern is shown on the left.

The example above uses an identity function $u(x) = x$ for utility. One could posit, alternatively, that decision makers psychophysically transforms money amount in question via a log transformation (Anderson, 1970); e.g., instead of $x_i - y_i$, the difference becomes $log(x_i) - log(y_i)$ or $log\frac{x_i}{y_i}$; and in this case, a log utility function $u(x) = log(x)$ is used. In this paper, I consider two kinds of lexicographic semiorder models, one uses an identity function $u(x) = x$ for utility (represented as LSO-Diff), and the other one uses a log function $u(x) = log(x)$ for utility (represented as LSO-Ratio).

1.2.2 Similarity Models

Rubinstein (1988) proposed a type of intransitive heuristic model called a *similarity model* to explain some phenomena that cannot be explained by expected utility theory. Unlike a lexicographic semiorder model, which orders gamble attributes lexicographically, a similarity model assumes that the decision maker considers all attributes simultaneously.

Rubinstein (1988) defined two types of similarity, the ϵ-difference similarity and λ-ratio similarity. Suppose that $\epsilon > 0$ is the threshold. For any m, $n \in \mathbb{R}$, Rubinstein defined the difference similarity by $m \sim n$ if $|m - n| \leq \epsilon$, and the ratio similarity by $m \sim n$ if $\frac{1}{\lambda} \leq m/n \leq \lambda$. In other words, the difference similarity uses an identity function $u(x) = x$ for the utility of money rewards x, and the ratio similarity uses a log function $u(x) = log(x)$ for utility. Rubinstein described how a similarity model works for gambles with two outcomes as follows: Suppose there are two gambles, $x = (x_1, x_2)$ and $y = (y_1, y_2)$, where x_1, x_2, y_1, and y_2

are attributes of the gambles, e.g., the payoff or the probability of winning.

Step 1. If both $x_1 > y_1$ and $x_2 > y_2$, then $x \succ y$. Or, if both $x_1 < y_1$ and $x_2 < y_2$, then $x \prec y$. Otherwise, the decision maker proceeds to Step 2.

Step 2. If $x_2 \sim y_2$ and $x_1 > y_1$ (and not $x_1 \sim y_1$), then $x \succ y$. If $x_2 > y_2$ (and not $x_2 \sim y_2$) and $x_1 \sim y_1$, then $x \succ y$. Otherwise, the decision maker moves to Step 3, which is not specified in Rubinstein (1988).

Based on the procedures proposed by Rubinstein (1988), the similarity models I test in the current paper work as follows: a decision maker picks a threshold for each attribute of a gamble pair and forms a preference for that attribute. The decision maker derives his final preferences from integrating all preferences on all attributes. To illustrate, suppose the decision maker considers two gambles x and y, each with two attributes, Attributes 1 and 2, and proceeds through the following decision making process:

- $(x \succ_1 y \;\; and \;\; x \succ_2 y) \;\; or \;\; (x \succ_1 y \;\; and \;\; x \sim_2 y) \;\; or \;\; (x \sim_1 y \;\; and \;\; x \succ_2 y) \Rightarrow x \succ y$,
- $(x \prec_1 y \;\; and \;\; x \prec_2 y) \;\; or \;\; (x \prec_1 y \;\; and \;\; x \sim_2 y) \;\; or \;\; (x \sim_1 y \;\; and \;\; x \prec_2 y) \Rightarrow x \prec y$,
- $(x \succ_1 y \;\; and \;\; x \prec_2 y) \;\; or \;\; (x \prec_1 y \;\; and \;\; x \succ_2 y) \;\; or \;\; (x \sim_1 y \;\; and \;\; x \sim_2 y) \Rightarrow x \sim y$.

Here I show an example of how a similarity model works using Tversky's (1969) gambles: suppose a decision maker uses a similarity model with an identity function, $u(x) = x$. He uses $\frac{3.5}{24}$ as the threshold of probabilities of winning, and \$0.35 as the threshold of payoffs. He forms preferences for the ten gamble pairs regarding probabilities of winning and payoffs, as shown in the top two tables of Panel E in Table 1.1. When considering the probabilities of winning, he prefers e over a, and he is indifferent about the remaining pairs. When considering the payoffs, he is indifferent about the adjacent pairs and prefers the gambles with higher payoffs for the other pairs. After integrating his preferences on both attributes, the decision maker derives his final preferences, which are shown in the bottom table of Panel E in Table 1.1. The decision maker is indifferent about all adjacent pairs and the extreme pair, pair (a, e). Of the remaining pairs, the decision maker prefers the gambles with higher payoffs. For example, for pair (a, e), the decision maker prefers e to a ($a \prec e$) based on the probabilities of winning and prefers a to e ($a \succ e$) based on the payoffs. Thus, after integrating his preferences across both attributes, the decision maker is indifferent between a and e ($a \sim e$). Here, $a \succ c$, $c \succ e$, and $e \sim a$ form intransitive preferences.

In this paper, I consider two types of similarity models, one uses an identity function $u(x) = x$ for utility (represented as SIM-Diff), and the other one uses a log function $u(x) = log(x)$ for utility (represented as SIM-Ratio).

For a more detailed review of lexicographic semiorder models and similarity models, see Guo (2018).

1.3 Transitive Models

1.3.1 Linear Order Models

In this paper, I also test linear order models, which contain all permissible transitive strict linear orders. The five gambles in Tversky's experiment generate $5! = 120$ linear orders. All of these 120 linear orders are transitive. The linear order model does not consider gamble specifics and only depends on the number of gambles under consideration. Regenwetter et al. (2011a,b, 2017) tested linear order models on risky and intertemporal data, and reported that the linear order model could explain the participants' behavior very well.

1.3.2 Two Simple Transitive Heuristics

One simple transitive heuristic, labeled *Payoff-only*, is that a decision maker prefers the gamble with larger payoff, regardless of the probabilities of winning. For example, taking Tversky's gambles, this heuristic predicts that the decision maker's preference pattern is: $a \succ b$, $a \succ c$, $a \succ d$, $a \succ e$, $b \succ c$, $b \succ d$, $b \succ e$, $c \succ d$, $c \succ e$, and $d \succ e$ (Ranking *abcde*). One other simple transitive heuristic, labeled *Prob-only*, is that a decision maker prefers the gamble with larger probability of winning, regardless of the payoffs. For Tversky's gambles, this heuristic predicts that the decision maker's preference pattern is: $a \prec b$, $a \prec c$, $a \prec d$, $a \prec e$, $b \prec c$, $b \prec d$, $b \prec e$, $c \prec d$, $c \prec e$, and $d \prec e$ (Ranking *edcba*). Both of these preference patterns, Rankings *abcde* and *edcba*, are among the 120 linear orders. Both are also special cases of LSO-Diff, LSO-Ratio, SIM-Diff, and SIM-Ratio for Tversky's stimuli.

1.4 Probabilistic Specifications

What do rigorous tests of algebraic decision theories look like? To answer this question, I want to discuss the relationship between preferences and choices first. Preference is defined as people's attitude towards a set of items (Lichtenstein and Slovic, 2006). It is used by many theories in psychology and economics, and it is a theoretical concept that we cannot directly observe. What we can observe and study in an experimental paradigm are pairwise choices. As Tversky (1969) mentioned, when a person is faced with the same choice options repeatedly, he does not always choose the same option. Therefore, one needs to figure out how variable choices are related to underlying preferences.

To be more specific, transitivity of preferences is an algebraic property, and decision theories are usually stated in deterministic terms. At the same time, experimental research collects variable choice data. How can one test an algebraic theory using probabilistic data? Luce (1959, 1995, 1997) presented a two-fold challenge for studying algebraic decision theories. The first part of the challenge is to specify a probabilistic extension of an algebraic theory, a problem that has been discussed by many scholars (Carbone and Hey, 2000; Harless and Camerer, 1994; Hey, 1995, 2005; Hey and Orme, 1994; Loomes and Sugden, 1995; Starmer, 2000; Tversky, 1969). The second part of the challenge is to test the probabilistic specifications of the theory with rigorous statistical methods, which was only solved in the past decade with a breakthrough in order-constrained, likelihood-based inferences (Davis-Stober, 2009; Myung et al., 2005; Silvapulle and Sen, 2005). In order to perform an appropriate and rigorous test of transitivity of preferences, researchers have to solve Luce's challenge. However, very few studies in the existing literature offer convincing solutions.

Regenwetter et al. (2014) provided a general and rigorous quantitative framework for testing theories of binary choice, which one can use to test transitivity of preferences. To solve the first part of Luce's challenge, they presented two kinds of probabilistic specifications of algebraic models to explain choice variability: a *distance-based* probabilistic specification models preferences as deterministic and response processes as probabilistic, and a *mixture* specification models preferences as probabilistic and response processes as deterministic. Sections 1.4.1 and 1.4.2 provide details of these two probabilistic specifications. For the second part of Luce's challenge, Regenwetter et al. (2014) employed frequentist likelihood-based statistical inference methods for binary choice data with order-constraints on each choice probability (Iverson and Falmagne, 1985; Silvapulle and Sen, 2005; Davis-Stober, 2009). Myung et al. (2005) and Klugkist and Hoijtink (2007) provided Bayesian order-constrained statistical inference techniques. In this paper, I specify two kinds of probabilistic models for each algebraic theory and test those probabilistic models with both frequentist and Bayesian order-constrained statistical methods.

1.4.1 Distance-Based Models

A distance-based model, which is also called the error model, assumes that a decision maker has a fixed preference throughout the experiment. It allows the decision maker to make errors/trembles in a binary pair with some probability that is bounded by a maximum allowable error rate. Formally, a distance-based model requires binary choice probabilities to lie within some specified distances of a point hypothesis that represents a preference state. More precisely, let $\tau \in (0, 0.50]$ be the upper bound on the error rate for each probability. For any pair (x, y), the probability of choosing x over y, θ_{xy}, is given by

$$
\begin{aligned}
x \succ y \quad &\Leftrightarrow \quad \theta_{xy} \geq 1-\tau \\
x \prec y \quad &\Leftrightarrow \quad \theta_{xy} \leq \tau \\
x \sim y \quad &\Leftrightarrow \quad \tfrac{1-\tau}{2} \leq \theta_{xy} \leq \tfrac{1+\tau}{2}
\end{aligned}
$$

When a decision maker prefers x to y, he chooses x over y with probability at least $1-\tau$. When a decision maker prefers y to x, he chooses x over y with probability at most τ. As mentioned before, when a decision maker is indifferent about x and y and chooses without errors, the "true" probability θ_{xy} is $\frac{1}{2}$. When this decision maker chooses with errors and the upper bound on the error rate is τ, the probability of choosing x over y is bounded by $\frac{1-\tau}{2}$ and $\frac{1+\tau}{2}$. When $\tau = 0.50$, this is also named as *modal choice*, which assumes a decision maker has a deterministic preference and allows the decision maker to make errors on each pair with probability at most 0.50. In other words, when $\tau = 0.50$, it means that the modal choice for each pair is consistent with the predictions of an algebraic theory (up to sampling variability). When $\tau = 0.90$, the decision maker chooses the preferred prospect with probability at least 0.90. Consider the example of the lexicographic semiorder model shown in Panel D of Table 1.1. That lexicographic semiorder model predicts $a \sim b$, $a \prec e$, and $b \succ e$. The distance-based model with upper bound $\tau = 0.50$ means that a decision maker chooses a over b with probability ranging from 0.25 to 0.75, a over e with probability at most 0.50, and b over e with probability at least 0.50. However, a distance-based model with upper bound $\tau = 0.50$ assumes a decision maker chooses his preferred prospect more often than not and might be too lenient. To compensate for this, one could place a more restrictive constraint on τ for each binary pair. Still using $a \sim b$, $a \prec e$, and $b \succ e$ as an example, the distance-based model with upper bound $\tau = 0.10$ means that the decision maker chooses a over b with probability ranging from 0.45 to 0.55, a over e with probability at most 0.10, and b over e with probability at least 0.90. In this paper, I use three different upper bounds, $\tau = 0.50$, 0.25, and 0.10, on the error rate.

1.4.2 Mixture Models

A mixture model assumes that a decision maker's preferences are probabilistic. Variations in observed choice behavior are no longer due to errors but rather to decision makers' uncertain preferences. A decision maker might fluctuate in his preferences during the experiment, making a choice based on one of the decision theory's predicted preference patterns on each given trial. A mixture model treats parameters of algebraic theory as random variables with unknown joint distribution; it does not make any distributional assumptions regarding the joint outcomes of the random variables. Geometrically, a mixture model forms the convex hull of the point hypotheses that capture the various possible preference states.

Take LSO-Diff and Tversky's stimuli (given in Table 1.1, Panel A), for example. There are three different parameters to consider in the algebraic model:

- The gambles' attribute order. There are two possible orders:
  - first payoff then probability of winning,
  - first probability of winning then payoff.
- The threshold for the probability of winning (ϵ_{prob}). There are five possible scenarios for the threshold regarding the probability of winning (ϵ_{prob}):
  - $\epsilon_{prob} < 1/24$ (strict linear order according to the probability of winning),
  - $\epsilon_{prob} \geq 4/24$ (complete indifference according to the probability of winning),
  - $1/24 \leq \epsilon_{prob} < 2/24$, $2/24 \leq \epsilon_{prob} < 3/24$, $3/24 \leq \epsilon_{prob} < 4/24$ (i.e., three more semiorders according to the probability of winning).
- The threshold for the payoff (ϵ_{pay}). There are five possible scenarios for the threshold regarding the payoff (ϵ_{pay}):
  - $\epsilon_{pay} < .25$ (strict linear order according to the payoff),
  - $\epsilon_{pay} \geq 1$ (complete indifference according to the payoff),
  - $.25 \leq \epsilon_{pay} < .5$, $.5 \leq \epsilon_{pay} < .75$, $.75 \leq \epsilon_{pay} < 1$ (i.e., three more semiorders according to the payoff).

As one considers different attribute orders and different values for ϵ_{prob} and ϵ_{pay}, one obtains many preference patterns. I obtain 21 different preference patterns for Tversky's gambles (shown in Table 1.2), as I vary the sequence of attributes and the threshold values. Row 16 in Table 1.2 shows the preference pattern that is depicted on the left side of Panel D in Table 1.1.

A mixture model treats the three parameters in the lexicographic semiorder model (the attribute orders and the threshold values) as random variables with any joint distribution whatsoever, hence permitting all possible probability distributions over the various permissible preference patterns.

As mentioned before, I write $\succ$ for strict preference and $\sim$ for indifference. I define $\mathcal{LSO}$ as a set of lexicographic semiorders and $P(\succ_{LSO})$ as the probability of lexicographic semiorder $\succ_{LSO}$ in $\mathcal{LSO}$. According to the mixture model, for any pair (x, y), the binary choice probability θ_{xy} is

$$\theta_{xy} = \sum_{\substack{\succ_{LSO} \in \mathcal{LSO} \\ \text{in which } x \succ y}} P(\succ_{LSO}) + \frac{1}{2} \sum_{\substack{\succ'_{LSO} \in \mathcal{LSO} \\ \text{in which } x \sim y}} P(\succ'_{LSO}).$$

This equation shows that the probability of choosing x over y equals the total probability of those lexicographic semiorders in which x is strictly preferred to y plus half of the probability of those lexicographic semiorders in which x is indifferent to y.

The mixture LSO-Diff model for Tversky's gambles can be cast geometrically as the convex hull (polytope) of 21 vertices in a suitably chosen 10-dimensional unit hypercube of binary choice probabilities. Each vertex encodes the binary choice probabilities when the probability mass is concentrated on a signal lexicographic semiorder. I provide a minimal description of the mixture polytope of LSO-Diff for Tversky's gambles in terms of its facet-defining equalities and inequalities, via the public-domain software PORTA [1]:
Equalities:

$$\theta_{ab} = \theta_{bc} = \theta_{cd} = \theta_{de}, \tag{1.1}$$

$$\theta_{ac} = \theta_{bd} = \theta_{ce}, \tag{1.2}$$

$$\theta_{ad} = \theta_{be}. \tag{1.3}$$

Inequalities:

$$0 \leq \theta_{ae}, \theta_{be}, \theta_{ce}, \theta_{de} \leq 1, \tag{1.4}$$

$$0 \leq \theta_{be} + \theta_{ce} - 2\theta_{de} \leq 2, \tag{1.5}$$

$$0 \leq \theta_{ae} + \theta_{ce} - 2\theta_{de} \leq 2, \tag{1.6}$$

$$0 \leq \theta_{ae} + \theta_{be} - 2\theta_{de} \leq 2, \tag{1.7}$$

$$0 \leq \theta_{ae} + \theta_{be} - 2\theta_{ce} \leq 2, \tag{1.8}$$

$$0 \leq -\theta_{ae} + 2\theta_{be} - 2\theta_{ce} + 2\theta_{de} \leq 2. \tag{1.9}$$

Equalities 1.1 to 1.3 show equal probabilities for certain gamble pairs. For example, Equality 1.1 shows equal probabilities for adjacent pairs in Tversky's stimuli. Equalities 1.1 to 1.3 show that this mixture polytope has four free parameters, θ_{ae}, θ_{be}, θ_{ce}, and θ_{de}, which are restricted by Inequalities 1.4 to 1.9. In this case, the mixture model is not full dimensional. It is a 4-dimensional polytope within in a 10-D space. I cannot test this mixture model with frequentist order-constrained statistical methods because the frequentist methods only work for full dimensional models. The Bayesian methods, on the other hand, can handle non full dimensional polytopes, such as the mixture LSO-Diff model described above.

Unlike a lexicographic semiorder model which has three parameters, a similarity model has two param-

[1] For more information, please see `http://comopt.ifi.uni-heidelberg.de/software/PORTA/`)

eters: the threshold for the payoff (ϵ_{pay}) and the threshold for the probability of winning (ϵ_{prob}). Take SIM-Diff and Tversky's gambles as an example, as one varies the values for ϵ_{pay} and ϵ_{prob}, the SIM-Diff model permits 21 preference patterns (not the same 21 patterns as those predicted by the LSO-Diff model). The mixture SIM-Diff model treats these two parameters (ϵ_{pay} and ϵ_{prob}) in the similarity model as random variables with any joint distribution whatsoever, hence permitting all possible probability distributions over these 21 preference patterns. I provide the minimal descriptions of the mixture polytope for each decision heuristic in the supplemental materials.

1.4.3 Summary of Models

Table 1.3 summarizes all of the models in this paper. The first column lists the model names. For the model names, I use the word *noisy* for distance-based models, and the word *random* for mixture models. The second column lists the core theory for each model, and the third column gives a label for each core theory. This paper tests seven core theories, including four intransitive decision heuristics (LSO-Diff, LSO-Ratio, SIM-Diff, and SIM-Ratio) and three transitive heuristics (LO, Prob-only, and Payoff-only). In addition to these seven decision heuristics, I also consider a *saturated* model that is unconstrained that places no constraints whatsoever on binary choice probabilities. The fourth column describes the utility function for each intransitive heuristic. The fifth and sixth columns summarize whether preferences and responses are each deterministic or probabilistic. For each distance-based model, I consider three different upper bounds on the error rate. Because Prob-only and Payoff-only predict only one preference pattern each, there are no mixture models for these two heuristics. Altogether I test 26 models in this paper.

1.4.4 Statistical Methods

In the current study, I report results using both frequentist (Davis-Stober, 2009; Iverson and Falmagne, 1985; Silvapulle and Sen, 2005) and Bayesian (Myung et al., 2005) order-constrained statistical inference methods. For frequentist tests, the decision models under consideration are null hypotheses, and I report frequentist goodness-of-fit test results with a significance level of 0.05. For the distance-based models, the predicted preference pattern with the largest p-value is called a *best-fitting preference pattern*. For each participant, the frequentist test finds the best-fitting preference pattern and tests whether the data are compatible with the constraints on binary choice probabilities.

For Bayesian tests, I compute Bayes factors (BF, Kass and Raftery, 1995) for each model. The Bayes factor measures the empirical evidence for each decision model while appropriately penalizing the *complexity* of the model. The complexity of a model refers to the volume of the parameter space that a decision theory occupies relative to the saturated model.

For distance-based models, the order constraints are orthogonal within each model, and the priors on each dimension are independent and conjugate to the likelihood function. Thus, I can obtain analytical solutions for the Bayes factors of the distance-based models, compared to the saturated model. For mixture models, the order constraints are not orthogonal, so I use a Monte Carlo sampling procedure. I use supercomputing resources to complete the analyses in this paper[2].

I use Bayes factors to compare each model to the saturated model and select among models at both individual and group levels. To interpret the individual level Bayes factor results, I use the rule-of-thumb cutoffs for "substantial" evidence and "decisive" evidence, according to Jeffreys (1998). I use BF_A to represent the Bayes factor of model A; I use BF_B to represent the Bayes factor for model B; and I use $BF_{AB} = \frac{BF_A}{BF_B}$ to represent the Bayes factor for model A over model B. When $BF_{AB} > 3.2$, it means that there is "substantial" evidence in favor of model A; when $BF_{AB} > 100$, it means that there is "decisive" evidence in favor of model A. I will say that a decision model "fails" if its Bayes factor against the saturated model is less than 1.0; I will say that a decision model "substantially fits" if its Bayes factor against the saturated model is larger than 3.2; I will say that a decision model "decisively fits" if its Bayes factor against the saturated model is higher than 100; I will say that a decision model is "best" (or a "winner") if its Bayes factor against the saturated model is higher than 3.2 and it has the highest Bayes factor among the models under consideration.

For the group level comparison, I use the group Bayes factor (GBF, Stephan et al., 2007) to select among models. The GBF aggregates *likelihoods* across participants and is the product of individual-level Bayes factors. The model with the highest GBF is the one that best accounts for all participants' data jointly.

1.5 Experiments

In this paper, I analyze datasets from three different studies: Experiment I in Tversky (1969), Cash I and Cash II in Regenwetter et al. (2011a), and Session I and Session II in an experiment I conducted in 2012.

Experiment I in *Tversky (1969).* In this experiment, Tversky used five gambles, shown in Table 1.1. Each gamble was displayed on a card with a wheel of chance in which the black area represented the probability. The experiment used a 2AFC paradigm. Tversky pre-selected eight participants who made cyclical choices in a preliminary session. All eight participants then made repeated choices for each gamble pair over five sessions, four times each session.

[2] I ran analyses on Pittsburgh Supercomputer Center's Blacklight, Greenfield, and Bridges supercomputers, as an Extreme Science and Engineering Discovery Environment project (see also (Towns et al., 2014)). The analyses in this paper used about 140,000 CPU hours on the supercomputer.

Cash I and Cash II in *Regenwetter et al. (2011a).* This study replicated the study in Tversky (1969), except: (a) in the set labeled Cash I, the authors adjusted the amount of payoffs to their current dollar equivalent by adjusting for inflation; (b) in the set labeled Cash II, the authors created a new set of monetary gambles that each have an expected value equal to \$8.80 (see Table 1.4). Participants were 18 undergraduates at the University of Illinois at Urbana-Champaign. Gambles were presented as wheels of chance on computers, similar to Figure 1.1. Each gamble pair was repeated 20 times, separated by decoys to minimize memory effects.

Session I and Session II in an experiment I conducted in 2012. This experiment was conducted over two sessions held on two consecutive days. Session II replicated Session I. In Session I, 67 adults participated; of these, 54 returned for Session II. The stimulus set had 20 gamble pairs, ten gamble pairs from Cash I and ten gamble pairs from Cash II in Regenwetter et al. (2011a). Participants made repeated choices (20 times for each pair per session) over gamble pairs that were presented via computers using a 2AFC paradigm. Each gamble was displayed as a wheel of chance (see Figure 1.1), with colored areas to represent probabilities and numbers next to the wheels to represent payoffs. These 20 gamble pairs are only a fraction of all stimuli used in this experiment. The analysis results of another stimulus set in this experiment were published in Guo and Regenwetter (2014). From now on, I refer to this experiment from in 2012 as the Guo and Regenwetter (2014) experiment.

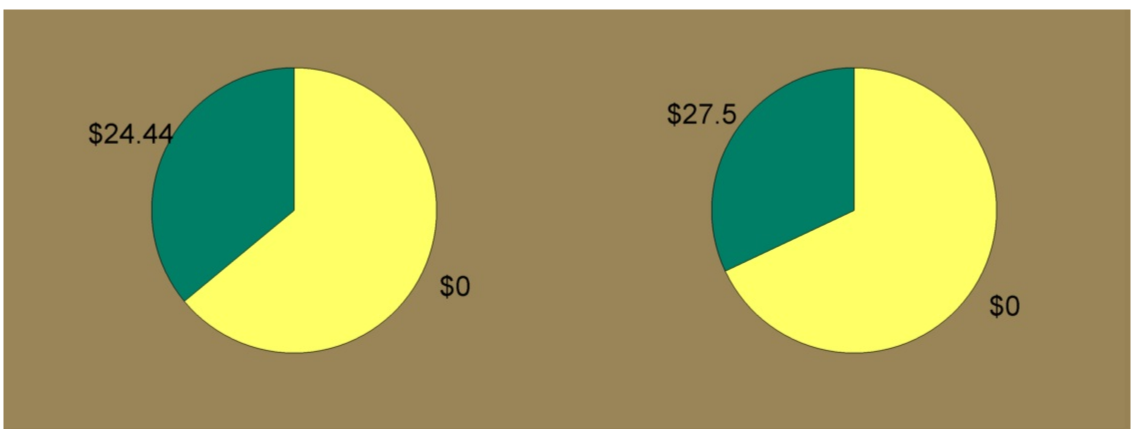

Figure 1.1: A gamble pair displayed in the experiment that I conducted in 2012.

1.6 Results

1.6.1 Distance-Based Model Results

Tables 1.5, 1.6, and 1.7 summarize the results for the distance-based models using both frequentist and Bayesian methods (Tables 1 - 16 in the supplemental materials provide individual-level p-values and Bayes factors for each stimulus set). The first two columns of Tables 1.5, 1.6, and 1.7 display the core theory and the upper bound τ on the error rate; Columns 3 - 5 and 7 - 8 report the total number of people who are fit by the distance-based models for Tversky's data, Cash I, Cash II, Session I, and Session II; Column 6 reports the number of people who are simultaneously fit for Cash I and Cash II; and Column 9 reports the number of people who are simultaneously fit for Session I and Session II.

Table 1.5 shows that, as expected, for each decision theory, the number of people who are fit is the highest for the distance-based models with $\tau = 0.50$ and decreases when the upper bound τ on the error rate decreases. Overall, the distance-based models with $\tau = 0.50$ for LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO perform very well and fit the data of almost all participants. Please note that the distance-based model with $\tau = 0.50$ for LO is also labeled *weak stochastic transitivity*, which is one of the most influential probabilistic models used for testing transitivity of preferences in the literature (Tversky, 1969). The results show that the data of almost all participants in all stimulus sets satisfy weak stochastic transitivity, and imply very little evidence against transitivity. When $\tau = 0.10$, the distance-based models for LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO account for almost none of Tversky's data and for the data of about half of the participants in the other stimulus sets. Thus, the number of people who are fit by the distance-based models decreases a lot when the upper bound τ decreases to 0.10 for all stimulus sets.

The noisy-Payoff-only and noisy-Prob-only models fit the data of fewer participants compared to the other distance-based models. These two models explain almost none of Tversky's data. For Cash I, the noisy-Prob-only models fit at most 13 (out of 18) participants' data, while the noisy-Payoff-only models fit at most three (out of 18) participants' data. For Cash II, Session I, and Session II, the noisy-Payoff-only and noisy-Prob-only models explain at most half of the participants' data. This result shows that there are some participants in all stimulus sets who might take "shortcuts" and form their preferences based on only one gamble attribute.

For Session I and Session II, the linear order model lives in 20-dimensional space, and it has 14,400 linear orders. There is a total number of $(67 + 54) \times 14,400 \times 3 = 5,227,200$ order-constrained frequentist tests for the noisy-LO model with three different upper bounds on the error rate for all participants. Computing all of these tests is computationally expensive. For each participant, instead of computing all frequentist tests, I use the following procedure: first, I pre-select the linear orders which substantially fit according to the Bayes factor analysis; second, I find the best-fitting linear order with the highest p-value among the preselected linear orders (note that the p-value of the best-fitting vertex is also the highest among all the 14,400 linear orders); and last, I check if the highest p-value is larger than the significance level of 0.05, and if so, I count it as a fit. Take the noisy-LO model with $\tau = 0.50$ for Session I as an example, the Bayes factor analysis shows that the noisy-LO model substantially wins over the saturated model for 67 (out of 67) participants. Of those 67 participants, the frequentist tests show that this noisy-LO model fits the data of 66 participants. For Session II, the noisy-LO model with $\tau = 0.50$ fits the data of all 54 participants. Again, these results show that the data of almost all of the participants in Sessions I and II satisfy weak stochastic transitivity.

When the frequentist tests of the distance-based models show that a participant is best described by a model with the same set of parameter values in two stimulus sets, I call it a *consistent fit*. For an intransitive heuristic, I count the number of people who are consistently fit by a model for two stimulus sets; and for a transitive heuristic, I count the number of people who are simultaneously fit by the same preference pattern predicted by a decision heuristic for two stimulus sets. Columns 6 and 9 in Table 1.5 report such results. Take the noisy-LSO-Diff model with $\tau = 0.50$ for Cash I and Cash II as an example, 18 (out of 18) participants in Cash I and 18 (out of 18) in Cash II are fit by the noisy-LSO-Diff model with $\tau = 0.50$, but only eight (out of 18) replicate across Cash I and Cash II. For the four intransitive models and the linear order model, the number of participants who replicate across Cash I and Cash II is much smaller than the number of participants who are fit in each set of Cash I and Cash II separately. In other words, when a model fits the data of some participants in Cash I, the estimated best-fitting parameters of that model need not predict the data of the same participants in Cash II. This shows that there might be some degree of 'over-fitting' for the distance-based models for LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO for Cash I and Cash II. The number of participants who replicate across Session I and Session II do not differ much from the number of participants who are fit in separate sessions. This result shows that the distance-based models for Session I and Session II do not seem to 'over-fit'. One interpretation might be that the distance-based models for Session I and Session II live in 20-dimensional space, and these models are much more parsimonious and are less likely to 'over-fit'.

Tables 1.6 and 1.7 shows the Bayes factor analysis results for the distance-based models. Panel A shows the results with substantial evidence and Panel B shows the results with decisive evidence. The results of the Bayes factor analyses with substantial evidence for the distance-based models are in alignment with the results of the corresponding frequentist analyses. When I consider the decisive evidence, the distance-based models with $\tau = 0.50$ for LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO fit for none of the participants in Cash I and Cash II; and the distance-based models with $\tau = 0.75$ or $\tau = 0.90$ for these five heuristics fit for about half of the participants in Cash I and Cash II. These results might be explained by the fact that the Bayes factor rewards parsimonious models and penalizes complex models. Thus, the distance-based model with $\tau = 0.50$ gets penalized for being more complex than the distance-based models with $\tau = 0.75$ or $\tau = 0.90$.

For the Bayes factor analyses, I also count the number of people who are simultaneously fit by the same model for two stimulus sets. Columns 6 and 9 in Tables 1.6 and 1.7 summarize such results. The number of fits that replicate across two stimulus sets is similar to the number of fits for separate sets. As I mentioned earlier, the frequentist analysis shows some evidence of 'over-fitting' for some distance-based models. In contrast, the Bayes factor analysis seems to be less forgiving. One interpretation is that the Bayes factor takes model complexity into account and successfully penalizes the more complex models.

1.6.2 Mixture Model Results

Table 1.8 shows the mixture model analysis results. It is made up of three panels. Each panel lists the number of permissible preference patterns, the number of inequality constraints, whether a polytope is full dimensional, the number of people who are successfully fit using frequentist methods, and the number of people who are substantially (and decisively) fit using Bayes factor methods. Because Prob-only and Payoff-only predict only one preference pattern each, there are no mixture models for these two heuristics. No frequentist tests of the random-LSO-Diff and random-SIM-Diff models for Tversky's set and Cash I are performed because their polytopes are not full dimensional. I cannot consider decisive evidence for the random-LO model for Tversky's set, Cash I and Cash II, because the maximum possible Bayes factor for that model is less than 100.

Panel A reports the results for Tversky's set. The frequentist analyses show that the random-LSO-Ratio, random-SIM-Ratio, and random-LO models all account for the data of more than half of the participants. The Bayesian analyses show that the mixture models for the four intransitive heuristics substantially fit for more than half of the participants, whereas the random-LO model only substantially fits for two (out of eight) participants. It seems that the random-LO model gets penalized by the Bayes factor for being too complex. The random-LSO-Diff and random-SIM-Diff models fit for the highest number of participants both substantially (eight out of eight participants) and decisively (three out of eight participants). These results show that, when using an identity function for utility, the mixture models for the intransitive heuristics fit for more participants than those with a log function for utility.

Panel B reports the results for Cash I and Cash II in Regenwetter et al. (2011a). The frequentist analyses show that the mixture models for LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO perform well and account for at least half of the participants' data, except that the random-SIM-Ratio model fits the data of seven (out of 18) participants for Cash II. The random-LO model fits the data of the highest number of participants (17 out of 18) for each set of Cash I and Cash II, suggesting very little evidence against transitivity. The Bayesian analyses show that the random-LO fits 12 participants for each set of Cash I and Cash II. Again, it seems like that random-LO model is penalized for being too complex.

Panel B also shows the number of participants who are simultaneously fit for both Cash I and Cash II. The random-LO model accounts for the data of the highest number of participants (17 out of 18) by the frequentist standard and beats the saturated model substantially for eight participants. The Bayes factor analyses show that the random-LSO-Diff, random-SIM-Diff, and random-LO models substantially fit for at least half of the participants for Cash I and Cash II simultaneously. When considering decisive evidence, the mixture models of all four intransitive heuristics fit for almost none of the participants.

Panel C reports the results for Session I and Session II in the Guo and Regenwetter (2014) experiment. The frequentist tests and Bayes factor analyses with substantial evidence show that the random-LO model performs the best and fits the data of almost all participants for each session. These results mean that almost all participants in Session I and Session II behave consistently with transitivity from the frequentist test point of view. The Bayes factor analyses with substantial evidence show that all five mixture models perform well and explain the data of more than half of the participants in each session. The mixture models for the two similarity heuristics for Cash I and Cash II decisively fit for more participants than the mixture models for the other three decision heuristics.

Panel C also shows the number of participants who are simultaneously fit by the mixture models for both sessions. The number of fits that replicate across sessions is similar to the number of fits for each session. Using the frequentist tests and the Bayes factor analyses with substantial evidence, the random-LO model simultaneously fits across two sessions for the most participants (51 out of 54 for frequentist test and 48 out of 54 for Bayes factor analysis with substantial evidence). The random-SIM-Diff and random-SIM-ratio models beat the saturated model decisively for the most participants (28 out of 54) for both Session I and Session II simultaneously.

Overall, I find a close alignment of results between the frequentist methods and the Bayesian methods, no matter whether I consider distance-based models or mixture models, although these statistical methods involve dramatically distinct concepts and computational procedures.

1.6.3 Model Comparison: Individual Level

I use Bayes factors to compare models. As I discuss in Section 1.4.4, for each participant, a decision model is "best" (or a "winner") if its Bayes factor against the saturated model is higher than 3.2 and it has the highest Bayes factor among a group of models. This section reports the best model at the individual level for each stimulus set.

Table 1.9 shows the best models for Tversky's experiment (top panel) and Regenwetter et al.'s experiment (bottom panel). For each panel, the first column shows the participant ID. The second column shows the core theory of the best model. The third column shows the stochastic form and the upper bound τ on the error rate (when applicable). I use "Fixed" to represent the distance-based model and "Random" to represent the mixture model. This column also reports the upper bound τ on the error rate for the distance-based model. The fourth column shows the Bayes factor for the best model compared to the saturated model. The fifth column shows the Bayes factor between the best and second-best models. I refer to LSO-Diff, LSO-Ratio, SIM-Diff, and SIM-Ratio as "intransitive" theories because they permit intransitive preference patterns (as well as transitive ones)

For Tversky's experiment, the core theories of the best models for all eight participants are models that permit intransitive preferences. Four of the eight best models are lexicographic semiorder models, and four are similarity models. For Cash I, among the core theories of the best models for all 18 participants, ten are transitive theories (of which, eight are Prob-only, and two are Payoff-only) and seven are intransitive theories (of which, six are similarity models, and one is a lexicographic semiorder model). For Cash II, among the core theories of the best models for all 18 participants, 11 are transitive theories (of which, five are Prob-only; four, Payoff-only; and two, LO) and seven are intransitive theories (of which, two are similarity models, and five are lexicographic semiorder models). For Participant 4 in Cash I, no models under consideration win over the saturated model substantially. For both Cash I and Cash II, four participants are simultaneously best fit by Prob-only as core theory; two participants, Payoff-only; and one participant, SIM-Ratio. Therefore, six participants in Regenwetter et al. (2011)'s experiment prefer the gambles with larger reward or prefer the gambles with larger probability all the time.

Regarding probabilistic specifications, seven out of eight winners are mixture models for Tversky's sets, five out of 18 for Cash I, and eight out of 18 for Cash II. The distance-based models win out less often than the mixture models for Tversky's set, but more often for Cash I and Cash II. These results suggest that across different stimulus sets, there are a lot of individual indifferences regarding their choice behavior.

Overall, no core theory is the best across the board. For Tversky's set, all participants are best fit by the intransitive heuristics. Almost all participants in Tversky's experiment seem to employ the mixture model, that is, they have variable preferences and make no mistakes when making choices during the experiment. For Regenwetter et al.'s stimuli, the transitive theories win out the most. Unlike Tversky's participants, most of the participants in Regenwetter et al.'s experiment tend to match the distance-based models, according to which they have deterministic preferences but make errors when making choices during the experiment. The results show that the participants in Tversky's experiment behave much differently from the participants in Regenwetter et al.'s experiment. The participants in Tversky's experiment were pre-selected for making cyclical choices in the preliminary sessions. It is not surprising that the intransitive heuristics explain Tversky's data well.

Table 1.10 shows the best model for each participant in Session I and Session II. For Session I, among the 67 winners, 28 are transitive theories (of which, 11 are Prob-only; 16 are Payoff-only; and one is LO) and 38 are intransitive theories (of which, 29 are similarity models, and nine are lexicographic semiorder models). For Session II, among the 54 winners, 21 are transitive theories (of which, seven are Prob-only; 11 are Payoff-only; and four are LO) and 32 are intransitive theories (of which, 27 are similarity models, and five are lexicographic semiorder models). For both Session I and Session II, 10 (out of 54) participants are simultaneously best fit by transitive theories (of which, six are Payoff-only and four are Prob-only) and 17 by intransitive theories (of which, 16 are similarity models, and one is a lexicographic semiorder model). For Participant 33, no substantive models beat the saturated model substantially for Session I. Therefore, more participants in Session I and Session II are best fit by the intransitive theories. The models that best fit the data of the most participants are the similarity models (with $u(x) = x$ in Session I and with $u(x) = log(x)$ in Session II).

As for the probabilistic specifications, for Session I, 40 out of 67 participants are best fit by the distance-based models and 27 by the mixture models; and for Session II, 40 out of 54 participants are best fit by the distance-based models and 14 by the mixture models. For Session I and Session II, there are more participants who seem to employ the distance-based models than the mixture models.

It is notable that for all three studies, when the intransitive heuristics are the best models, the probabilistic specifications are often the mixture models. In other words, when a participant employs an intransitive heuristic, he tends to vary his preferences during the experiment. There is no single core theory or probabilistic specification that is robust across all participants and all stimulus sets.

1.6.4 Model Comparison: Group Level

Table 1.11 reports the results of the model comparison at the group level using the group Bayes factor (GBF). The first column shows the model name; the second column shows the upper bound τ on the error rate, which is only applicable to the distance-based model; Columns 3 - 7 report the ranking of each model from the best (highest GBF) to worst (lowest GBF) for each stimulus set. The model with the highest group Bayes factor is the model that will generalize best to data from a randomly selected participant in a group for a stimulus set. For both Tversky's set and Cash I, the random-LSO-Diff and random-SIM-Diff models are among the top three models. For Cash II, Session I, and Session II, the noisy-SIM-Diff and noisy-SIM-Ratio models with $\tau = 0.75$ are among the top three models. The noisy-LO models with $\tau = 0.75$ and $\tau = 0.90$ and all noisy-Payoff-only and noisy-Prob-only models perform very badly; because they do not beat the saturated model for any of the stimulus sets. For a stimulus set, the distance-based Payoff-only and Prob-only models could best fit for some individual participants, but they could not fit for some other participants at all. With these huge individual differences, the noisy-Payoff-only and noisy-Prob-only models do not generate well to data from a randomly selected participant in a group. Overall, the results reveal that the similarity model and the lexicographic semiorder model are the core theories of the top three most generalizable models for all five stimulus sets.

1.7 Conclusions and Discussions

Transitivity of preferences is essential for nearly all normative, prescriptive, and descriptive theories of decision making. Almost any theory that uses utility functions implies transitivity. There are studies reporting intransitive choice behavior in the literature. However, most of those studies contain pervasive methodological problems as explained in Guo (2018). To explain the intransitive choice behavior, several contemporary theories are developed in the literature. The lexicographic semiorder model and the similarity model are two examples of those theories permitting intransitive preferences. This paper presents a comprehensive analysis of the lexicographic semiorder model and the similarity model and compares them to the transitive linear order model and two simple transitive heuristics. This paper tries to find out if there is much evidence against transitivity and which model can explain human choice behavior better, transitive models or intransitive models.

In this paper, I employ a rigorous quantitative framework for testing decision theories. I consider two types of probabilistic specifications of algebraic theories: the distance-based model and the mixture model. The distance-based model assumes that the decision maker has a deterministic preference and makes errors when making choices. I use three upper bounds τ on the error rate. The mixture model assumes that the decision maker has probabilistic preferences and chooses deterministically when making choices. The mixture model allows any probability distribution whatsoever over preference patterns that are consistent with the decision theory or the algebraic structure of interest. When a mixture model is rejected, it means that there does not exist a probability distribution over those preference patterns that would describe well the decision maker's data. All in all, I test 26 different probabilistic models in this paper.

I use both frequentist and Bayesian order-constrained statistical methods. The frequentist order-constrained method provides a goodness-of-fit test for the probabilistic model from a classical statistical perspective. I find some evidence of 'over-fitting' for some distance-based models using the frequentist analysis. The Bayesian order-constrained method allows me to put all 26 probabilistic models in direct comparison with one another at both the individual and group levels. Moreover, the Bayes factor measures the empirical evidence for each model while appropriately penalizing for the complexity of the model. The Bayes factor analysis is less forgiving than the frequentist methods.

I test all 26 models on the data from three different experiments. The frequentist goodness-of-fit tests show that the distance-based models for all seven decision heuristics with modal choice well-describe the participants' data in all stimulus sets. The mixture model analyses show that all five decision theories (LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO) perform well and can explain the data of more than half of the participants. The Bayesian analysis with substantial evidence provides similar results to the frequentist analysis.

The model comparison at the individual level shows that for Tversky's set, the intransitive heuristics win out for all participants; for Cash I and Cash II, the transitive heuristics win out for most participants; and for Session I and Session II, the intransitive heuristics win out for most participants. This result shows heterogeneity across participants and stimulus sets. Moreover, I do not find a single core theory, type of preference, or type of response process that best explains all participants' data in all stimulus sets. This reinforces earlier warnings that one needs to be cautious about a "one-size-fits-all" approach, as pointed out previously by Davis-Stober et al. (2015), Hey (2005), Loomes et al. (2002), and Regenwetter et al. (2014).

The model comparison at the individual level also shows that Payoff-only and Prob-only are the core theories of the best models for some participants in Cash I, Cash II, Session I, and Session II. This result means that there is a small group of participants who simplify the task and prefer the gambles with a higher payoff or the gambles with a higher probability of winning during the entire experiment. Unlike Cash I, Cash II, Session I and Session II, all of the best models in Tversky's experiment are intransitive. This result could be explained by the fact that all eight participants in Tversky's experiment were pre-selected for making cyclical choices in a preliminary session. The model comparison at the group level tells a somewhat different story: for all five stimulus sets, the similarity model and the lexicographic semiorder model are the core theories of the top three most generalizable models for all five stimulus sets.

Looking at the frequentist results, the linear order model explains well almost all participants' data in all stimulus sets. The frequentist tests of the random-LO model on Cash I and Cash II replicate the results in Regenwetter et al. (2011a). Thus, from a classical statistical perspective, I do not find much evidence against transitivity. However, the linear order model hardly wins out in the Bayesian model comparison. The results show that even when a participant doesn't violate transitivity from the frequentist test point of view, the intransitive heuristics can still give more parsimonious explanations of the participant's behavior than the linear order model. The results show that even though the lexicographic semiorder model and the similarity model allow intransitivity, they are not just models of intransitivity; both transitive and intransitive preferences can be consistent with these models. This speaks directly to Birnbaum (2011)'s concern about model mimicry. My analyses show that many participants are fit by both the intransitive heuristics and the linear order model. One explanation for this finding might be that many preference patterns predicted by the intransitive heuristics are transitive, and some are linear orders. Regenwetter et al. (2011b) report that the lexicographic semiorder model can mimic parts of the linear order model, and both models fit a large proportion of the participants. Future research might use more diagnostic stimuli to minimize overlap between intransitive decision heuristics and the linear order model.

1.8 Tables

Table 1.1: Tversky's (1969) gambles. Panel A shows the probabilities of winning, payoffs, and expected values for each of the five gambles. Panel B shows the differences in the probabilities of winning among pairs. Panel C shows the differences of the payoffs among pairs. Panel D shows an example of the binary preference relation predicted by a lexicographic semiorder model. Panel E shows an example of the binary preference relation predicted by a similarity model.

Panel A: Tversky's (1969) gambles

Lottery	Prob. of winning	Payoff (in $)	Expected value (in $)
a	7/24	5.00	1.46
b	8/24	4.75	1.58
c	9/24	4.50	1.69
d	10/24	4.25	1.77
e	11/24	4.00	1.83

Panel B: The probability of winning differences (column-row)

Lottery	a	b	c	d	e
a	-	1/24	2/24	3/24	4/24
b		-	1/24	2/24	3/24
c			-	1/24	2/24
d				-	1/24

Panel C: The payoff differences (row-column)

Lottery	a	b	c	d	e
a	-	$.25	$.50	$.75	$1
b		-	$.25	$.50	$.75
c			-	$.25	$.50
d				-	$.25

Panel D: A lexicographic semiorder[1]

Binary Preference Relation					
Lottery	a	b	c	d	e
a	-	$\sim$	$\succ$	$\succ$	$\prec$
b		-	$\sim$	$\succ$	$\succ$
c			-	$\sim$	$\succ$
d				-	$\sim$

Binary Choice Probabilities[3]					
Gamble	a	b	c	d	e
a	-	$\frac{1}{2}$	1	1	0
b		-	$\frac{1}{2}$	1	1
c			-	$\frac{1}{2}$	1
d				-	$\frac{1}{2}$

Panel E: A similarity model[2]

Preferences by Probability					
lottery	a	b	c	d	e
a	-	$\sim$	$\sim$	$\sim$	$\prec$
b		-	$\sim$	$\sim$	$\sim$
c			-	$\sim$	$\sim$
d				-	$\sim$

Preferences by Payoff					
lottery	a	b	c	d	e
a	-	$\sim$	$\succ$	$\succ$	$\succ$
b		-	$\sim$	$\succ$	$\succ$
c			-	$\sim$	$\succ$
d				-	$\sim$

Binary Preference Relation					
Lottery	a	b	c	d	e
a	-	$\sim$	$\succ$	$\succ$	$\sim$
b		-	$\sim$	$\succ$	$\succ$
c			-	$\sim$	$\succ$
d				-	$\sim$

Binary Choice Probabilities					
Gamble	a	b	c	d	e
a	-	$\frac{1}{2}$	1	1	$\frac{1}{2}$
b		-	$\frac{1}{2}$	1	1
c			-	$\frac{1}{2}$	1
d				-	$\frac{1}{2}$

1. It is the binary preference pattern predicted by a lexicographic semiorder model if a decision maker considers the probabilities before the payoffs and uses a probability threshold of $\frac{3.5}{24}$ and a payoff threshold of $0.35.
2. It is the binary preference pattern predicted by a similarity model if a decision maker uses a probability threshold of $\frac{3.5}{24}$ and a payoff threshold of $0.35.

Table 1.2: The 21 Preference patterns predicted by the LSO-Diff model for Tversky (1969)'s gambles.

	(a,b)	(a,c)	(a,d)	(a,e)	(b,c)	(b,d)	(b,e)	(c,d)	(c,e)	(d,e)
1	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$
2	$\prec$	$\prec$	$\prec$	$\succ$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$
3	$\prec$	$\prec$	$\succ$	$\succ$	$\prec$	$\prec$	$\succ$	$\prec$	$\prec$	$\prec$
4	$\prec$	$\succ$	$\succ$	$\succ$	$\prec$	$\succ$	$\succ$	$\prec$	$\succ$	$\prec$
5	$\sim$	$\prec$	$\prec$	$\prec$	$\sim$	$\prec$	$\prec$	$\sim$	$\prec$	$\sim$
6	$\sim$	$\prec$	$\prec$	$\succ$	$\sim$	$\prec$	$\prec$	$\sim$	$\prec$	$\sim$
7	$\sim$	$\prec$	$\succ$	$\succ$	$\sim$	$\prec$	$\succ$	$\sim$	$\prec$	$\sim$
8	$\sim$	$\sim$	$\prec$	$\prec$	$\sim$	$\sim$	$\prec$	$\sim$	$\sim$	$\sim$
9	$\sim$	$\sim$	$\prec$	$\succ$	$\sim$	$\sim$	$\prec$	$\sim$	$\sim$	$\sim$
10	$\sim$	$\sim$	$\sim$	$\prec$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$
11	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$
12	$\sim$	$\sim$	$\sim$	$\succ$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$
13	$\sim$	$\sim$	$\succ$	$\prec$	$\sim$	$\sim$	$\succ$	$\sim$	$\sim$	$\sim$
14	$\sim$	$\sim$	$\succ$	$\succ$	$\sim$	$\sim$	$\succ$	$\sim$	$\sim$	$\sim$
15	$\sim$	$\succ$	$\prec$	$\prec$	$\sim$	$\succ$	$\prec$	$\sim$	$\succ$	$\sim$
16	$\sim$	$\succ$	$\succ$	$\prec$	$\sim$	$\succ$	$\succ$	$\sim$	$\succ$	$\sim$
17	$\sim$	$\succ$	$\succ$	$\succ$	$\sim$	$\succ$	$\succ$	$\sim$	$\succ$	$\sim$
18	$\succ$	$\prec$	$\prec$	$\prec$	$\succ$	$\prec$	$\prec$	$\succ$	$\prec$	$\succ$
19	$\succ$	$\succ$	$\prec$	$\prec$	$\succ$	$\succ$	$\prec$	$\succ$	$\succ$	$\succ$
20	$\succ$	$\succ$	$\succ$	$\prec$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$
21	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$

Table 1.3: Summary of the models analyzed in this paper.

Model Name	Core Theory	Label for Core Theory	Utility Function	Preferences	Response Process
noisy-LSO-Diff	Lexicographic semiorder	LSO-Diff	$u(x) = x$	Deterministic	Probabilistic
noisy-LSO-Ratio	Lexicographic semiorder	LSO-Ratio	$u(x) = log(x)$	Deterministic	Probabilistic
noisy-SIM-Diff	Similarity	SIM-Diff	$u(x) = x$	Deterministic	Probabilistic
noisy-SIM-Ratio	Similarity	SIM-Ratio	$u(x) = log(x)$	Deterministic	Probabilistic
noisy-SIM-Ratio	Similarity	SIM-Ratio	$u(x) = log(x)$	Deterministic	Probabilistic
noisy-SIM-Ratio	Similarity	SIM-Ratio	$u(x) = log(x)$	Deterministic	Probabilistic
noisy-LO	Linear order	LO	-	Deterministic	Probabilistic
noisy-Payoff-only	Only consider payoff	Payoff-only	-	Deterministic	Probabilistic
noisy-Prob-only	Only consider probability	Prob-only	-	Deterministic	Probabilistic
random-LSO-Diff	Lexicographic semiorder	LSO-Diff	$u(x) = x$	Probabilistic	Deterministic
random-LSO-Ratio	Lexicographic semiorder	LSO-Ratio	$u(x) = log(x)$	Probabilistic	Deterministic
random-SIM-Diff	Similarity	SIM-Diff	$u(x) = x$	Probabilistic	Deterministic
random-SIM-Ratio	Similarity	SIM-Ratio	$u(x) = x$	Probabilistic	Deterministic
random-LO	Linear order	LO	-	Probabilistic	Deterministic
saturated	All binary preference patterns	saturated	-	-	-

Table 1.4: Cash I and Cash II stimuli in Regenwetter et al. (2011a).

Cash I			Cash II		
Gamble	Prob. of Winning	Payoff (in $)	Gamble	Prob. of Winning	Payoff (in $)
a	7/24	28	a	0.28	31.43
b	8/24	26.6	b	0.32	27.50
c	9/24	25.2	c	0.36	24.44
d	10/24	23.8	d	0.40	22
e	11/24	22.4	e	0.44	20

Table 1.5: The total number of people who are fit by the distance-based models using frequentist tests. I use a significance level of 0.05. The total number of participants is shown in parentheses in the header.

Model		Number of Fits						
Core Theory	τ	Tversky' set (8)	Cash I (18)	II (18)	Cash I & II (18)	Session I (67)	Session II (54)	Sessions I & II (54)
LSO-Diff	0.50	8	18	18	8	65	52	45
LSO-Diff	0.25	6	16	13	7	56	48	28
LSO-Diff	0.10	1	8	8	6	24	30	13
LSO-Ratio	0.50	8	18	18	5	65	52	45
LSO-Ratio	0.25	6	16	13	5	57	48	28
LSO-Ratio	0.10	1	10	8	5	26	34	13
SIM-Diff	0.50	8	18	18	8	66	52	46
SIM-Diff	0.25	8	17	15	7	59	48	30
SIM-Diff	0.10	1	9	9	6	24	30	13
SIM-Ratio	0.50	8	18	18	7	66	52	46
SIM-Ratio	0.25	8	17	15	5	59	48	30
SIM-Ratio	0.10	1	11	9	5	27	34	13
LO	0.50	5	17	17	5	66 (67)	54 (54)	40 (54)
LO	0.25	1	9	9	5	25 (32)	22 (25)	10(13)
LO	0.10	0	7	7	5	13 (17)	14 (19)	7(8)
Payoff-only	0.50	2	3	8	3	34	32	23
Payoff-only	0.25	0	2	3	2	14	11	6
Payoff-only	0.10	0	1	2	1	7	8	4
Prob-only	0.50	0	13	5	5	29	21	17
Prob-only	0.25	0	7	5	5	9	6	3
Prob-only	0.10	0	6	5	5	6	6	3

Table 1.6: The total number of people who are fit with substantial evidence by the distance-based models using Bayes factor analyses. The total number of participants is shown in parentheses in the header.

Core Theory	τ	Tversky' set (8)	Cash I (18)	Cash II (18)	Cash I & II (18)	Session I (67)	Session II (54)	Sessions I & II (54)
LSO-Diff	0.50	8	17	16	15	66	52	51
LSO-Diff	0.25	4	15	11	9	57	49	41
LSO-Diff	0.10	1	10	9	6	34	39	23
LSO-Ratio	0.50	8	17	14	13	66	52	51
LSO-Ratio	0.25	2	14	11	9	56	47	40
LSO-Ratio	0.10	0	10	8	6	34	39	23
SIM-Diff	0.50	8	17	16	15	64	52	49
SIM-Diff	0.25	8	16	12	11	60	51	46
SIM-Diff	0.10	1	10	9	6	35	38	23
SIM-Ratio	0.50	8	17	15	14	62	52	48
SIM-Ratio	0.25	4	15	13	12	58	50	44
SIM-Ratio	0.10	1	10	8	6	36	39	23
LO	0.50	1	12	12	9	45	36	30
LO	0.25	0	9	8	7	20	20	11
LO	0.10	0	8	7	6	16	16	9
Payoff-only	0.50	2	2	6	2	26	22	16
Payoff-only	0.25	0	2	3	2	17	16	8
Payoff-only	0.10	0	1	2	1	9	9	5
Prob-only	0.50	0	13	6	6	21	10	8
Prob-only	0.25	0	7	5	5	10	7	5
Prob-only	0.10	0	7	5	5	8	6	3

Table 1.7: The total number of people who are fit with decisive evidence by the distance-based models using Bayes factor analyses. The total number of participants is shown in parentheses in the header.

Core Theory	τ	Tversky' set (8)	Cash I (18)	Cash II (18)	Cash I & II (18)	Session I (67)	Session II (54)	Sessions I & II (54)
LSO-Diff	0.50	0	0	0	0	65	50	49
LSO-Diff	0.25	0	9	8	6	45	44	30
LSO-Diff	0.10	0	9	7	6	32	36	21
LSO-Ratio	0.50	0	0	0	0	62	49	46
LSO-Ratio	0.25	0	10	7	6	46	44	30
LSO-Ratio	0.10	0	7	7	6	30	36	19
SIM-Diff	0.50	0	0	0	0	62	50	47
SIM-Diff	0.25	0	10	9	7	51	45	34
SIM-Diff	0.10	0	9	8	6	33	36	21
SIM-Ratio	0.50	0	0	0	0	62	51	48
SIM-Ratio	0.25	0	10	8	6	52	45	37
SIM-Ratio	0.10	0	8	7	6	32	36	19
LO	0.50	0	0	0	0	0	0	0
LO	0.25	0	8	7	6	18	19	11
LO	0.10	0	7	7	6	15	14	7
Payoff-only	0.50	1	2	4	2	25	21	14
Payoff-only	0.25	0	2	3	2	16	15	7
Payoff-only	0.10	0	1	2	1	9	9	5
Prob-only	0.50	0	9	5	5	17	10	7
Prob-only	0.25	0	7	5	5	10	7	5
Prob-only	0.10	0	7	5	5	8	6	3

Table 1.8: The results for the mixture models of LSO-Diff, LSO-Ratio, SIM-Diff, SIM-Ratio, and LO using both frequentist and Bayesian methods. Each panel shows the number of permissible predicted patterns, the number of inequality constraints, whether a polytope is full dimensional, the number of participants who are successfully fit by the mixture models using frequentist tests (labeled "Freq Fits"), Bayes factor methods with substantial evidence (labeled "BF Fits (Substantial)"), and Bayes factor methods with decisive evidence (labeled "BF Fits (Decisive)"). Panel A shows results for Tversky's set, Panel B shows results for Cash I and Cash II in Regenwetter et al. (2011a), and Panel C shows results for Sessions I and II in the Guo and Regenwetter (2014) experiment. The maximum Bayes factor for the random-LO model for Tversky's set, Cash I, and Cash II is less than 100, so the Bayes factor analysis with decisive evidence is not applicable.

Panel A: Tversky's set, 8 participants.

	LSO-Diff	LSO-Ratio	SIM-Diff	SIM-Ratio	LO
Number of Patterns	21	111	21	101	120
Number of Constraints	18	24	30	36	40
Full Dimensional?	No	Yes	No	Yes	Yes
Freq Fits	-	5	-	7	6
BF Fits (Substantial)	8	5	8	6	2
BF Fits (Decisive)	3	0	3	1	-

Panel B: Cash I (C1) and Cash II (C2) in Regenwetter et al. (2011a), 18 participants.

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
Number of Patterns	21	51	111	111	21	51	101	111	120	
Number of Constraints	18	39	24	1956	30	37	36	2046	40	
Full Dimensional?	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	
Freq Fits	-	13	9	11	-	13	14	7	17	17
BF Fits (Substantial)	16	11	5	11	17	9	9	12	12	12
BF Fits (Decisive)	10	5	0	1	12	5	3	3	-	-
The number of participants who are simultaneously fit in both Cash I and Cash II										
Fits Freq	-		5		-		5		17	
BF Fits (Substantial)	10		4		9		6		8	
BF Fits (Decisive)	1		0		1		2		-	

Panel C: Session I (S1) and Session II (S2) in the Guo and Regenwetter (2014) experiment, 67 participants in S1 and 54 in S2.

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Number of Patterns	135		401		128		339		14400	
Number of Constraints	189(201)		32015		59(71)		625		80	
Full Dimensional?	No		Yes		No		Yes		Yes	
Freq Fits	-	-	46	30	-	-	30	18	64	54
BF Fits (Substantial)	54	37	47	33	56	47	49	46	62	51
BF Fits (Decisive)	42	24	35	22	49	36	48	35	34	33
The number of participants who are simultaneously fit in both Session I and Session II										
Fits Freq	-		22		-		9		51	
BF Fits (Substantial)	30		27		40		37		48	
BF Fits (Decisive)	14		13		28		28		22	

Table 1.9: The best model for each participant in Tversky's experiment and Regenwetter et al.'s experiment. The column labeled "BF" shows the Bayes factor of the substantive model against the saturated model. The column labeled "Best/Second" shows the Bayes factor of the best model against the second-best model.

Tversky (1969)				
ID	Core Theory	Stochastic Form & τ	BF	Best/Second
1	LSO-Diff	Random	1119	3
2	SIM-Ratio	Random	43	1
3	LSO-Ratio	Random	53	2
4	LSO-Diff	Random	60	1
5	SIM-Diff	Random	1042	2
6	LSO-Diff	Fixed-0.50	27	1
7	SIM-Ratio	Random	395	5
8	SIM-Diff	Random	706	3

Regenwetter et al. (2011)								
	Cash I				Cash II			
ID	Core Theory	Stochastic Form & τ	BF	Best/Second	Core Theory	Stochastic Form & τ	BF	Best/Second
1	SIM-Diff	Random	168	3	LSO-Diff	Random	220	8
2	Payoff-only	Fixed-0.25	1975	3	Payoff-only	Fixed-0.10	5659793	15
3	Prob-only	Fixed-0.10	1596997327	21	Prob-only	Fixed-0.10	5754379	11
4	-	-	-	-	Payoff-only	Fixed-0.50	187	7
5	Prob-only	Fixed-0.10	177548988	21	Prob-only	Fixed-0.25	305226	5
6	SIM-Diff	Fixed-0.25	723	1	LSO-Diff	Random	243	1
7	Prob-only	Fixed-0.25	439099	1	LO	Random	15	3
8	Prob-only	Fixed-0.10	382023676	21	Prob-only	Fixed-0.10	194383842	51
9	Prob-only	Fixed-0.50	30	1	LSO-Diff	Random	1843	1
10	Prob-only	Fixed-0.10	2234430	5	Prob-only	Fixed-0.10	6818583	13
11	Prob-only	Fixed-0.10	150818569	21	Prob-only	Fixed-0.10	248904793	51
12	LSO-Diff	Fixed-0.50	21	1	LO	Random	8	1
13	SIM-Diff	Fixed-0.25	49	1	SIM-Diff	Random	510	1
14	Payoff-only	Fixed-0.10	1596997327	21	Payoff-only	Fixed-0.10	3138587985	51
15	SIM-Diff	Random	1053	3	LSO-Diff	Random	163	6
16	Prob-only	Fixed-0.25	52974	10	LSO-Diff	Fixed-0.50	14	2
17	SIM-Ratio	Random	35	1	Payoff-only	Fixed-0.25	5659	6
18	SIM-Ratio	Random	449	1	SIM-Ratio	Random	166	1

Table 1.10: The best model for each participant in Session I and Session II of the Guo and Regenwetter (2014) experiment. The column labeled "BF" shows the Bayes factor of the substantive model against the saturated model. The column labeled "Best/Second" shows the Bayes factor of the best model against the second-best model.

	Session I				Session II			
ID	Core Theory	Stochastic Form & τ	BF	Best/Second	Core Theory	Stochastic Form & τ	BF	Best/Second
1	Payoff-only	Fixed-0.50	33559	2	SIM-Diff	Fixed-0.10	954836	1
2	SIM-Diff	Random	104719	2	SIM-Ratio	Random	85318	9
4	SIM-Ratio	Random	307735	1	LO	Random	180	2
5	LSO-Ratio	Random	416287	158	SIM-Diff	Fixed-0.25	5896	1
7	SIM-Ratio	Random	498911	2	LSO-Diff	Random	29465	1
9	LSO-Ratio	Random	60311	19	Prob-only	Fixed-0.10	5187202440167750	128
11	Payoff-only	Fixed-0.50	371948	4	SIM-Ratio	Random	10502329	27
12	Payoff-only	Fixed-0.25	3194624973	117	LO	Fixed-0.25	209489	121
13	Prob-only	Fixed-0.25	2233633096	125	Prob-only	Fixed-0.25	3261817	17
14	Payoff-only	Fixed-0.10	22486473040159900	128	Payoff-only	Fixed-0.10	3971999329511040	128
15	SIM-Ratio	Random	1964223	11	LSO-Ratio	Random	182223	8
16	SIM-Ratio	Random	111679	16	Payoff-only	Fixed-0.25	43831240414	48
17	LSO-Ratio	Random	356740	64	LSO-Ratio	Fixed-0.25	11670516	2
18	LSO-Ratio	Fixed-0.25	1010190	1	SIM-Diff	Fixed-0.10	21380272	1
19	SIM-Diff	Fixed-0.10	406820	1	SIM-Diff	Fixed-0.10	167806	1
20	Prob-only	Fixed-0.10	1819253918812050000	67	Prob-only	Fixed-0.10	2550400462237230000	128
21	LSO-Diff	Fixed-0.50	158	3	LO	Random	154	24
22	SIM-Ratio	Random	7663	1	SIM-Ratio	Random	474121	50
23	SIM-Ratio	Fixed-0.50	2338	1	SIM-Diff	Fixed-0.50	1300	1
24	Prob-only	Fixed-0.10	239193190081860	128	Prob-only	Fixed-0.10	7806193011064110	128
25	LSO-Ratio	Fixed-0.25	3281898	1	SIM-Ratio	Fixed-0.25	287554	1
26	Payoff-only	Fixed-0.10	1411523045980350000	128	Payoff-only	Fixed-0.10	1978805152871400000	128
27	Payoff-only	Fixed-0.50	579932	3	SIM-Diff	Fixed-0.50	8894	1
28	SIM-Ratio	Random	2250479	1	SIM-Ratio	Random	11890370	6
29	SIM-Ratio	Random	54455	10	SIM-Ratio	Random	70266	8
30	SIM-Diff	Random	4154974	1	SIM-Ratio	Random	154099	8
31	LO	Random	362	9	LSO-Diff	Random	3472	1
32	Prob-only	Fixed-0.10	1815686353759780	126	Prob-only	Fixed-0.10	5012316622465820000	128
33	-	-	-	-	LO	Fixed-0.25	37438	14
34	SIM-Ratio	Random	8475	1	SIM-Diff	Fixed-0.10	1092834	1
35	Payoff-only	Fixed-0.25	10989669301	6	Payoff-only	Fixed-0.25	12404659195	128
36	Payoff-only	Fixed-0.10	2540397851211	16	Prob-only	Fixed-0.25	5203891737	125
37	Prob-only	Fixed-0.10	4945357811010200	128	SIM-Ratio	Random	679857	4
38	Payoff-only	Fixed-0.50	27197	2	LSO-Ratio	Fixed-0.25	371586	2
39	SIM-Ratio	Fixed-0.50	8223	1	SIM-Diff	Fixed-0.10	75025	1
41	SIM-Ratio	Fixed-0.25	38815	1	Payoff-only	Fixed-0.50	126289	4
42	SIM-Ratio	Fixed-0.10	1124012	1	Payoff-only	Fixed-0.25	4863514511	50
43	Prob-only	Fixed-0.50	37800	3	SIM-Diff	Fixed-0.25	3753	1
44	LSO-Diff	Fixed-0.50	1664	1	SIM-Ratio	Fixed-0.10	199433150	1
46	LSO-Diff	Random	360432	2	Payoff-only	Fixed-0.25	25146569614	20
47	Payoff-only	Fixed-0.10	1952370581570670	128	Payoff-only	Fixed-0.50	389013	3

Continued on next page

Table 1.10 – continued from previous page

	Session I				Session II			
ID	Core Theory	Stochastic Form & τ	BF	Best/Second	Core Theory	Stochastic Form & τ	BF	Best/Second
48	SIM-Ratio	Random	67949	6	SIM-Diff	Fixed-0.25	10922	1
49	SIM-Ratio	Random	49544	5	SIM-Ratio	Random	202855	1
50	SIM-Ratio	Random	56663	2	SIM-Diff	Fixed-0.10	220737	1
52	SIM-Diff	Fixed-0.10	69822	1	SIM-Diff	Fixed-0.10	11376108	1
53	SIM-Ratio	Random	18652262	52	SIM-Diff	Fixed-0.25	12973	1
55	LSO-Diff	Fixed-0.50	110	2	Payoff-only	Fixed-0.10	7026734695266530000	128
56	SIM-Ratio	Random	305422	1	SIM-Diff	Fixed-0.10	485887	1
58	Payoff-only	Fixed-0.10	939305724364552	128	Payoff-only	Fixed-0.10	365449806241175000	128
59	SIM-Diff	Random	693086	1	SIM-Diff	Fixed-0.25	10115	1
61	Payoff-only	Fixed-0.25	2756504	3	SIM-Ratio	Random	4463528	2
65	SIM-Diff	Random	1949120	10	Prob-only	Fixed-0.10	48383069955994500	128
66	Prob-only	Fixed-0.10	283545259034153000	128	SIM-Ratio	Fixed-0.10	1297921170	1
67	Payoff-only	Fixed-0.10	293521003096709000	128	Payoff-only	Fixed-0.10	34512654355618000	128
3	Payoff-only	Fixed-0.50	17521	6				
6	Prob-only	Fixed-0.10	32568721096512	125				
8	Prob-only	Fixed-0.25	188290960	11				
10	SIM-Ratio	Random	9104	2				
40	SIM-Ratio	Random	5632975	13				
45	Payoff-only	Fixed-0.50	441298	2				
51	SIM-Diff	Fixed-0.10	1039566	1				
54	Prob-only	Fixed-0.50	413070	10				
57	SIM-Ratio	Random	2896719	3				
60	SIM-Diff	Fixed-0.10	1972426	1				
62	Prob-only	Fixed-0.50	38564	1				
63	Payoff-only	Fixed-0.10	22633597538109900	128				
64	SIM-Diff	Random	7756096	4				

Table 1.11: Ranking of each model from best (highest GBF) to worst (lowest GBF) in each stimulus set. Rankings in parentheses are worse than the saturated model on the same stimulus set. The first three best models are marked in boldfaced font.

Model Name	τ	Tversky	Cash I	Cash II	Session I	Session II
noisy-LSO-Diff	0.50	4	12	7	6	10
noisy-LSO-Diff	0.25	11	4	**3**	**3**	**2**
noisy-LSO-Diff	0.10	(18)	6	17	15	6
noisy-LSO-Ratio	0.50	7	14	11	8	12
noisy-LSO-Ratio	0.25	(13)	8	5	4	4
noisy-LSO-Ratio	0.10	(19)	11	(18)	14	8
noisy-SIM-Diff	0.50	**3**	10	6	5	9
noisy-SIM-Diff	0.25	6	**2**	**1**	**1**	**1**
noisy-SIM-Diff	0.10	(15)	5	4	11	5
noisy-SIM-Ratio	0.50	5	13	10	7	11
noisy-SIM-Ratio	0.25	9	7	**2**	**2**	**3**
noisy-SIM-Ratio	0.10	(17)	9	8	12	7
noisy-LO	0.50	(14)	15	15	16	17
noisy-LO	0.25	(20)	(18)	(19)	(18)	(19)
noisy-LO	0.10	(23)	(21)	(20)	(21)	(22)
noisy-Payoff-only	0.50	(16)	(23)	(22)	(19)	(20)
noisy-Payoff-only	0.25	(22)	(25)	(24)	(22)	(23)
noisy-Payoff-only	0.10	(24)	(26)	(26)	(24)	(25)
noisy-Prob-only	0.50	(21)	(20)	(21)	(20)	(21)
noisy-Prob-only	0.25	(25)	(22)	(23)	(23)	(24)
noisy-Prob-only	0.10	(26)	(24)	(25)	(25)	(26)
random-LSO-Diff	-	**1**	**3**	9	10	16
random-LSO-Ratio	-	10	(19)	14	17	(18)
random-SIM-Diff	-	**2**	**1**	12	9	13
random-SIM-Ratio	-	8	17	16	13	14
random-LO	-	(12)	16	13	(26)	15

1.9 Supplement Materials

The tables in the Supplement Materials report individual frequentist p-value and Bayes factors in each stimulus set.

Table 1.12: The frequentist and Bayes factor results for the distance-based models of LO, LSO-Diff, LSO-Ratio, SIM-Diff, and SIM-Ratio for Tversky (1969) data.

Panel A: The frequentist results.

	LSO-Diff			LSO-Ratio			SIM-Diff			SIM-Ratio			LO		
	$\tau =$			$\tau =$			$\tau =$			$\tau =$			$\tau =$		
	0.50	0.25	0.10	0.50	0.25	0.10	0.50	0.25	0.10	0.50	0.25	0.10	0.50	0.25	0.10
1	$\surd$	$\star$	$\star$	$\surd$	$\star$	$\star$	0.28	0.08	$\star$	0.34	0.08	$\star$	$\star$	$\star$	$\star$
2	$\surd$	0.34	$\star$	$\surd$	0.34	$\star$	$\surd$	0.34	$\star$	$\surd$	0.59	$\star$	0.13	$\star$	$\star$
3	0.35	$\star$	$\star$	0.54	$\star$	$\star$	$\surd$	0.36	$\star$	$\surd$	0.46	$\star$	$\star$	$\star$	$\star$
4	$\surd$	0.08	$\star$	$\surd$	0.08	$\star$	0.14	0.08	$\star$	0.28	0.08	$\star$	0.2	$\star$	$\star$
5	0.51	0.26	$\star$	0.51	0.26	$\star$	$\surd$	0.64	$\star$	$\surd$	0.64	$\star$	0.11	$\star$	$\star$
6	$\surd$	0.14	$\star$	$\surd$	0.14	$\star$	0.52	0.28	$\star$	0.52	0.28	$\star$	$\star$	$\star$	$\star$
7	$\surd$	0.36	0.15	$\surd$	0.36	0.15	$\surd$	0.36	0.15	$\surd$	0.75	0.15	0.55	$\star$	$\star$
8	$\surd$	0.77	$\star$	$\surd$	0.77	$\star$	$\surd$	0.77	$\star$	$\surd$	0.77	$\star$	$\surd$	0.09	$\star$
Fits	8	6	1	8	6	1	8	8	1	8	8	1	5	1	0

Panel B: The Bayes factors.

	LSO-Diff			LSO-Ratio			SIM-Diff			SIM-Ratio			LO		
	$\tau =$			$\tau =$			$\tau =$			$\tau =$			$\tau =$		
	0.50	0.25	0.10	0.50	0.25	0.10	0.50	0.25	0.10	0.50	0.25	0.10	0.50	0.25	0.10
1	21	0	0	9	0	0	12	5	0	8	2	0	0	0	0
2	18	6	0	7	1	0	23	6	0	32	8	0	0	0	0
3	13	0	0	5	0	0	30	19	0	25	24	0	0	0	0
4	18	9	1	25	2	0	8	9	1	4	2	0	0	0	0
5	15	1	0	5	0	0	46	5	0	33	2	0	0	0	0
6	27	1	0	6	0	0	24	12	0	7	3	0	0	0	0
7	41	27	16	21	5	3	61	28	16	77	49	5	2	0	0
8	45	95	0	23	43	0	54	97	0	29	48	0	7	0	0
Fits	8	4	1	8	2	0	8	8	1	8	4	1	1	0	0

Table 1.13: The frequentist results for the distance-based models for LSO-Diff, LSO-Ratio, SIM-Diff, and SIM-Ratio with $\tau = 0.50$, 0.25, and 0.10. There are 18 participants (# is the participant id). Rejections at a 0.05 level are marked $\star$. Perfect fits are checkmarks ($\surd$). Nonsignificant violations have their p-values listed. "Consistent Fits" are marked in `typewriter`.

Panel A: LSO-Diff and LSO-Ratio

	LSO-Diff						LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	$\surd$	0.45	0.27	$\star$	$\star$	$\star$	$\surd$	0.45	0.27	$\star$	0.14	$\star$
2	$\surd$	$\surd$	`0.22`	$\surd$	$\star$	0.55	$\surd$	$\surd$	0.22	$\surd$	$\star$	0.55
3	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.73`	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.73`
4	0.1	$\surd$	$\star$	$\star$	$\star$	$\star$	0.1	$\surd$	$\star$	$\star$	$\star$	$\star$
5	$\surd$	$\surd$	$\surd$	$\surd$	`0.99`	`0.57`	$\surd$	$\surd$	$\surd$	$\surd$	`0.99`	`0.57`
6	$\surd$	0.35	0.66	$\star$	0.13	$\star$	$\surd$	0.35	0.66	$\star$	0.13	$\star$
7	$\surd$	0.17	$\surd$	$\star$	0.81	$\star$	$\surd$	0.17	$\surd$	$\star$	0.81	$\star$
8	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	`0.90`	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	`0.90`
9	0.27	$\surd$	0.11	0.22	$\star$	$\star$	0.27	$\surd$	0.11	0.22	$\star$	$\star$
10	$\surd$	$\surd$	$\surd$	$\surd$	`0.72`	`0.30`	$\surd$	$\surd$	$\surd$	$\surd$	`0.72`	`0.30`
11	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	$\surd$
12	$\surd$	0.52	0.07	$\star$	$\star$	$\star$	$\surd$	0.52	0.07	$\star$	$\star$	$\star$
13	$\surd$	0.62	0.18	0.1	$\star$	0.16	$\surd$	0.62	0.18	0.1	$\star$	0.16
14	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
15	$\surd$	0.66	0.25	0.09	$\star$	$\star$	$\surd$	0.66	0.5	0.09	$\star$	$\star$
16	$\surd$	$\surd$	0.89	0.11	$\star$	$\star$	$\surd$	$\surd$	0.89	0.11	$\star$	$\star$
17	`0.45`	$\surd$	$\star$	0.45	$\star$	$\star$	0.55	$\surd$	$\star$	0.45	$\star$	$\star$
18	$\surd$	$\surd$	0.17	0.34	$\star$	$\star$	$\surd$	$\surd$	0.66	0.34	0.09	$\star$
Fits	18	18	16	13	8	8	18	18	16	13	10	8

Panel B: SIM-Diff and SIM-Ratio.

	LSO-Diff						LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	$\surd$	0.47	0.27	$\star$	$\star$	$\star$	$\surd$	0.52	0.27	$\star$	0.14	$\star$
2	$\surd$	$\surd$	`0.22`	$\surd$	$\star$	0.55	$\surd$	$\surd$	0.22	$\surd$	$\star$	0.55
3	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.73`	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.73`
4	0.1	$\surd$	$\star$	$\star$	$\star$	$\star$	0.1	$\surd$	$\star$	$\star$	$\star$	$\star$
5	$\surd$	$\surd$	$\surd$	$\surd$	`0.99`	`0.57`	$\surd$	$\surd$	$\surd$	$\surd$	`0.99`	`0.57`
6	$\surd$	$\surd$	0.66	0.13	0.13	0.08	$\surd$	$\surd$	0.66	0.13	0.13	0.08
7	$\surd$	0.17	$\surd$	$\star$	0.81	$\star$	$\surd$	0.17	$\surd$	$\star$	0.81	$\star$
8	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	`0.90`	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	`0.90`
9	0.27	$\surd$	0.12	0.22	$\star$	$\star$	0.36	$\surd$	0.12	0.22	$\star$	$\star$
10	$\surd$	$\surd$	$\surd$	$\surd$	`0.72`	`0.30`	$\surd$	$\surd$	$\surd$	$\surd$	`0.72`	`0.30`
11	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	$\surd$
12	$\star$	0.67	$\star$	$\star$	$\star$	$\star$	0.51	0.82	0.07	0.20	$\star$	$\star$
13	$\surd$	$\surd$	0.18	0.41	$\star$	0.51	$\surd$	$\surd$	0.18	0.41	$\star$	0.51
14	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
15	$\surd$	0.61	0.25	$\star$	$\star$	$\star$	$\surd$	0.61	0.50	0.09	$\star$	$\star$
16	$\surd$	0.78	0.89	0.14	0.09	$\star$	$\surd$	0.78	0.89	0.14	0.09	$\star$
17	$\surd$	$\surd$	$\star$	$\star$	$\star$	$\star$	$\surd$	$\surd$	0.5	0.45	$\star$	$\star$
18	$\surd$	$\surd$	0.17	0.34	$\star$	$\star$	$\surd$	$\surd$	0.66	0.61	0.09	$\star$
Fits	18	18	17	15	9	9	18	18	17	15	11	9

Table 1.14: The Bayes factors for the distance-based models for lexicographic semiorder model with $\tau = 0.50$, 0.25, and 0.10. There are 18 participants (# is the participant id).

Panel A: Lexicographic Semiorder Model.

	LSO-Diff						LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	32	5	52	0	3	0	17	3	38	0	3	0
2	38	31	94	9783	0	381290	10	14	19	4495	0	175187
3	49	20	47143	9827	76047492	112831	9	10	8919	4538	14387419	51921
4	0	19	0	0	0	0	0	9	0	0	0	0
5	49	20	39894	5985	8454714	1297	9	10	7553	2764	1599698	597
6	27	4	723	0	348	0	8	2	138	0	66	0
7	49	2	20909	0	15147	0	10	1	3976	0	2870	0
8	49	20	38846	15466	18191604	3811448	9	9	7350	7111	3441668	1751379
9	12	44	0	43	0	1	5	21	0	20	0	1
10	49	20	22875	10497	106401	133698	11	13	4434	5277	20534	75274
11	49	20	36037	16706	7181837	4880486	9	9	6818	7676	1358731	2242394
12	21	2	6	0	0	0	14	1	1	0	0	0
13	19	21	48	32	1	26	12	12	10	15	0	12
14	49	20	47143	20078	76047492	61541181	9	9	8925	9225	14388785	28275678
15	39	19	31	25	0	5	38	10	40	12	0	2
16	47	14	2523	0	175	0	9	7	477	0	33	0
17	4	28	0	125	0	0	4	13	0	57	0	0
18	39	19	43	2	1	0	46	13	115	1	1	0
Fits	17	16	15	11	10	9	17	14	14	11	10	8

Panel B: Similarity Model.

	LSO-Diff						LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	33	2	52	0	3	0	17	1	42	0	4	0
2	51	31	103	9783	0	381291	14	14	22	4495	0	175188
3	49	20	47146	9827	76047789	112831	10	10	9803	4538	15811978	51922
4	0	27	0	2	0	0	0	13	0	1	0	0
5	49	20	39896	5989	8454747	1297	10	10	8301	2766	1758091	597
6	26	29	723	13	348	1	8	20	152	6	72	1
7	50	1	20924	0	15149	0	11	0	4372	0	3155	0
8	49	20	38848	15467	18191675	3811463	10	9	8078	7111	3782442	1751386
9	22	36	0	43	0	1	15	17	0	20	0	1
10	50	20	22891	10505	106412	133711	13	13	4877	5281	22569	75282
11	49	20	36039	16718	7181865	4880968	10	9	7494	7682	1493265	2242616
12	16	6	6	1	0	0	6	5	1	1	0	0
13	22	63	49	294	1	141	10	61	10	142	0	65
14	49	20	47146	20079	76047789	61541421	10	9	9810	9225	15813479	28275788
15	42	13	31	24	0	5	33	6	44	11	0	2
16	80	4	5185	2	4630	0	17	2	1078	1	963	0
17	24	29	3	125	0	0	29	13	12	57	0	0
18	42	28	43	3	1	0	36	29	126	5	1	0
Fits	17	16	16	12	10	9	17	15	15	13	10	8

Table 1.15: The frequentist and Bayes factor results for the distance-based models for the linear order model with $\tau = 0.50$, 0.25, and 0.10. There are 18 participants (# is the participant id). Rejections at a 0.05 level are marked $\star$. Perfect fits are checkmarks ($\surd$). Nonsignificant violations have their p-values listed. "Consistent Fits" are marked in `typewriter`.

Panel A: The frequentist results.

	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	$\surd$	$\star$	$\star$	$\star$	$\star$	$\star$
2	$\surd$	$\surd$	0.22	$\surd$	$\star$	0.51
3	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.73`
4	$\star$	$\surd$	$\star$	$\star$	$\star$	$\star$
5	$\surd$	$\surd$	$\surd$	$\surd$	`0.99`	`0.57`
6	0.81	0.44	$\star$	$\star$	$\star$	$\star$
7	$\surd$	$\surd$	$\surd$	0.06	0.81	$\star$
8	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	`0.90`
9	$\surd$	$\surd$	$\star$	$\star$	$\star$	$\star$
10	$\surd$	$\surd$	$\surd$	$\surd$	`0.72`	`0.30`
11	$\surd$	$\surd$	$\surd$	$\surd$	`0.95`	$\surd$
12	0.63	0.33	$\star$	$\star$	$\star$	$\star$
13	0.67	$\surd$	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
15	$\surd$	$\surd$	$\star$	$\star$	$\star$	$\star$
16	$\surd$	0.28	0.89	$\star$	$\star$	$\star$
17	0.31	$\surd$	$\star$	0.45	$\star$	$\star$
18	$\surd$	0.23	$\star$	$\star$	$\star$	$\star$
Fits	17	17	9	9	7	7

Panel B: The Bayes factors.

	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	4	0	0	0	0	0
2	8	9	17	3137	0	47165
3	9	9	8250	4176	13308311	47953
4	0	4	0	0	0	0
5	9	9	6982	2544	1479575	551
6	2	1	0	0	0	0
7	9	6	3659	0	2651	0
8	9	9	6798	6573	3183531	1619865
9	3	7	0	0	0	0
10	9	9	4003	4461	18620	56822
11	9	9	6306	7100	1256821	2074207
12	1	3	0	0	0	0
13	3	1	0	0	0	0
14	9	9	8250	8533	13308311	26154900
15	7	5	0	0	0	0
16	8	0	441	0	31	0
17	0	8	0	47	0	0
18	5	0	0	0	0	0
Fits	12	12	9	8	8	7

Table 1.16: The frequentist results for the distance-based models for LSO-Diff and LSO-Ratio with $\tau = 0.50$, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2). Rejections at a 0.05 level are marked $\star$. Perfect fits are checkmarks ($\surd$). Nonsignificant violations have their p-values listed. Nonsignificant violations where Session II replicates Session I are marked in `typewriter`.

	LSO-Diff						LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1	√	√	0.09	0.15	⋆	⋆	`0.97`	√	⋆	0.44	⋆	⋆
2	√	√	`0.30`	`0.30`	⋆	⋆	√	√	`0.30`	`0.30`	⋆	⋆
4	√	`0.35`	0.11	⋆	⋆	⋆	√	`0.46`	0.42	⋆	⋆	⋆
5	`0.79`	√	⋆	0.49	⋆	⋆	`0.81`	√	⋆	0.49	⋆	⋆
7	√	√	0.11	0.06	⋆	⋆	√	`0.88`	0.14	0.07	⋆	⋆
9	0.31	√	⋆	√	⋆	0.65	√	√	0.17	√	⋆	0.65
11	√	√	0.08	0.59	⋆	0.12	√	√	0.08	0.59	⋆	0.12
12	√	⋆	0.59	⋆	⋆	⋆	√	⋆	0.59	⋆	⋆	⋆
13	√	`0.50`	0.28	⋆	⋆	⋆	√	`0.50`	0.28	⋆	⋆	⋆
14	√	√	√	√	√	`0.94`	√	√	√	√	√	`0.94`
15	`0.85`	`0.50`	⋆	0.11	⋆	⋆	`0.87`	√	⋆	0.16	⋆	⋆
16	√	√	⋆	√	⋆	0.15	√	√	⋆	√	⋆	0.15
17	`0.57`	`0.83`	⋆	0.11	⋆	⋆	`0.70`	√	0.13	0.96	⋆	0.09
18	`0.52`	√	`0.07`	`0.56`	⋆	0.34	√	√	`0.78`	`0.93`	⋆	0.34
19	√	√	`0.79`	`0.52`	`0.67`	`0.29`	√	√	`0.79`	`0.52`	`0.67`	`0.29`
20	√	√	√	√	√	√	√	√	√	√	√	√
21	0.71	⋆	⋆	⋆	⋆	⋆	0.59	⋆	⋆	⋆	⋆	⋆
22	`0.91`	`0.90`	`0.37`	`0.06`	0.08	⋆	`0.91`	`0.91`	`0.37`	`0.06`	0.08	⋆
23	`0.83`	`0.68`	⋆	⋆	⋆	⋆	`0.83`	`0.68`	⋆	⋆	⋆	⋆
24	√	√	√	√	`0.51`	`0.97`	√	√	√	√	`0.51`	`0.97`
25	`0.70`	`0.96`	`0.06`	`0.29`	⋆	⋆	√	√	`0.73`	`0.55`	0.21	0.13
26	√	√	√	√	√	√	√	√	√	√	√	√
27	√	√	0.38	0.38	⋆	⋆	√	√	0.38	0.45	⋆	⋆
28	√	√	`0.66`	`0.76`	`0.67`	`0.19`	√	√	`0.66`	`0.76`	`0.67`	`0.19`
29	`0.71`	√	`0.16`	`0.25`	⋆	⋆	`0.71`	√	`0.16`	`0.25`	⋆	⋆
30	√	√	`0.32`	`0.60`	`0.18`	`0.23`	√	√	`0.32`	`0.60`	`0.18`	`0.23`
31	⋆	0.78	⋆	0.18	⋆	⋆	⋆	0.78	⋆	0.28	⋆	⋆
32	√	√	√	√	`0.94`	√	√	√	√	√	`0.94`	√
33	⋆	0.47	⋆	⋆	⋆	⋆	⋆	0.23	⋆	⋆	⋆	⋆
34	`0.98`	√	`0.10`	`0.77`	⋆	0.62	`0.98`	√	`0.10`	`0.77`	⋆	0.62
35	√	√	`0.97`	√	⋆	0.09	√	√	√	√	0.36	0.09
36	√	√	√	0.31	0.62	0.12	√	√	√	0.31	0.62	0.12
37	√	`0.90`	√	0.07	0.98	⋆	√	`0.90`	√	0.07	0.98	⋆
38	√	√	`0.36`	`0.17`	⋆	⋆	√	√	`0.36`	`0.73`	⋆	⋆
39	√	`0.81`	`0.28`	`0.38`	⋆	0.26	√	`0.81`	`0.28`	`0.38`	⋆	0.26
41	`0.55`	√	0.09	⋆	⋆	⋆	√	√	0.57	⋆	⋆	⋆
42	√	√	⋆	0.76	⋆	⋆	√	√	`0.44`	`0.93`	⋆	0.29
43	`0.71`	`0.71`	⋆	0.13	⋆	⋆	`0.71`	√	⋆	0.16	⋆	⋆

Continued on next page

Table 1.16 – continued from previous page

	LSO-Diff						LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
44	0.74	√	0.34	0.25	⋆	⋆	0.74	√	0.34	0.77	⋆	0.29
46	√	√	0.07	√	⋆	0.09	√	√	⋆	√	⋆	0.09
47	√	√	√	0.26	0.98	⋆	√	√	√	0.26	0.98	⋆
48	√	0.88	0.36	0.19	⋆	0.12	√	0.89	0.36	0.19	⋆	0.12
49	0.74	√	⋆	0.6	⋆	0.24	0.74	√	⋆	0.6	⋆	0.24
50	0.84	√	0.16	0.30	⋆	0.28	√	√	0.29	0.30	⋆	0.28
52	√	√	0.26	0.42	0.22	0.79	√	√	0.26	0.42	0.22	0.79
53	√	√	0.22	0.50	⋆	0.15	√	√	0.22	0.50	⋆	0.15
55	0.23	√	⋆	√	⋆	√	0.23	√	⋆	√	⋆	√
56	√	√	0.92	0.79	0.34	0.59	√	√	0.92	0.79	0.34	0.59
58	√	√	√	√	0.90	√	√	√	√	√	0.90	√
59	√	√	0.53	0.52	0.26	0.16	√	√	0.53	0.52	0.26	0.16
61	√	√	0.12	0.8	⋆	0.8	√	√	0.66	0.8	⋆	0.8
65	√	√	0.6	√	⋆	√	√	√	0.6	√	⋆	√
66	√	√	√	0.43	√	0.1	√	√	√	0.89	√	0.28
67	√	√	√	√	0.98	0.99	√	√	√	√	0.98	0.99
3	√		0.16		⋆		0.9		0.16		⋆	
6	√		√		0.85		√		√		0.85	
8	√		0.37		⋆		√		0.37		⋆	
10	0.7		0.37		⋆		√		0.87		⋆	
40	√		0.75		0.49		√		0.75		0.49	
45	√		0.18		⋆		√		0.69		⋆	
51	√		0.44		0.5		√		0.44		0.5	
54	√		0.21		⋆		√		0.21		⋆	
57	√		0.65		0.69		√		0.65		0.69	
60	√		0.23		0.75		√		0.23		0.75	
62	√		0.09		⋆		√		0.09		⋆	
63	√		√		0.97		√		√		0.97	
64	0.99		0.11		⋆		0.99		0.11		⋆	
Fits	65	52	56	48	24	30	65	52	57	48	26	34

Table 1.17: The frequentist results for the distance-based models for SIM-Diff and SIM-Ratio with $\tau = 0.50$, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2). Rejections at a 0.05 level are marked $\star$. Perfect fits are checkmarks ($\surd$). Nonsignificant violations have their p-values listed. Nonsignificant violations where Session II replicates Session I are marked in `typewriter`.

	SIM-Diff						SIM-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1	√	√	0.09	0.15	⋆	⋆	`0.97`	√	⋆	0.44	⋆	⋆
2	√	√	`0.43`	`0.30`	⋆	⋆	√	√	`0.50`	`0.30`	⋆	⋆
4	√	`0.38`	0.11	⋆	⋆	⋆	√	`0.46`	0.42	⋆	⋆	⋆
5	`0.19`	√	⋆	0.49	⋆	⋆	`0.19`	√	⋆	0.49	⋆	⋆
7	√	√	`0.48`	`0.22`	⋆	⋆	√	`0.88`	`0.48`	`0.27`	⋆	⋆
9	√	√	0.17	√	⋆	0.65	√	√	0.28	√	⋆	0.65
11	√	√	0.08	0.59	⋆	0.12	√	√	0.08	0.69	⋆	0.12
12	√	⋆	0.59	⋆	⋆	⋆	√	⋆	0.59	⋆	⋆	⋆
13	√	`0.50`	0.28	⋆	⋆	⋆	√	`0.50`	0.28	⋆	⋆	⋆
14	√	√	√	√	√	`0.94`	√	√	√	√	√	`0.94`
15	√	`0.50`	0.53	0.11	⋆	⋆	√	`0.50`	0.53	0.16	⋆	⋆
16	√	√	0.17	√	⋆	0.15	√	√	0.32	√	⋆	0.15
17	`0.57`	`0.83`	⋆	0.11	⋆	⋆	`0.65`	√	0.13	0.8	⋆	0.09
18	`0.52`	√	`0.07`	`0.56`	⋆	0.34	√	√	`0.78`	`0.93`	⋆	0.34
19	√	√	`0.79`	`0.52`	`0.67`	`0.29`	√	√	`0.79`	`0.52`	`0.67`	`0.29`
20	√	√	√	√	√	√	√	√	√	√	√	√
21	0.21	⋆	⋆	⋆	⋆	⋆	0.18	⋆	⋆	⋆	⋆	⋆
22	`0.85`	`0.90`	`0.37`	`0.08`	0.08	⋆	√	`0.90`	`0.37`	`0.14`	0.08	⋆
23	√	`0.68`	⋆	⋆	⋆	⋆	√	`0.68`	⋆	⋆	⋆	⋆
24	√	√	√	√	`0.51`	`0.97`	√	√	√	√	`0.51`	`0.97`
25	`0.70`	`0.96`	`0.06`	`0.29`	⋆	⋆	√	√	`0.73`	`0.55`	0.21	0.13
26	√	√	√	√	√	√	√	√	√	√	√	√
27	√	√	0.38	0.45	⋆	⋆	√	√	0.38	0.6	⋆	⋆
28	√	√	`0.66`	`0.76`	`0.67`	`0.19`	√	√	`0.66`	`0.76`	`0.67`	`0.19`
29	`0.80`	√	`0.16`	`0.25`	⋆	⋆	√	√	`0.34`	`0.27`	0.1	⋆
30	√	√	`0.32`	`0.60`	`0.18`	`0.23`	√	√	`0.32`	`0.60`	`0.18`	`0.23`
31	`0.08`	`0.85`	⋆	0.13	⋆	⋆	`0.06`	`0.85`	⋆	0.13	⋆	⋆
32	√	√	√	√	`0.94`	√	√	√	√	√	`0.94`	√
33	0.12	0.47	⋆	⋆	⋆	⋆	0.13	0.23	⋆	⋆	⋆	⋆
34	√	√	`0.20`	`0.77`	⋆	0.62	√	√	`0.20`	`0.77`	⋆	0.62
35	√	√	`0.97`	√	⋆	0.09	√	√	√	√	0.36	0.09
36	√	√	√	0.31	0.62	0.12	√	√	√	0.31	0.62	0.12
37	√	√	√	0.19	0.98	⋆	√	√	√	0.37	0.98	⋆
38	√	√	`0.36`	`0.16`	⋆	⋆	√	√	`0.36`	`0.68`	⋆	⋆
39	√	`0.83`	`0.28`	`0.38`	⋆	0.26	√	√	`0.29`	`0.38`	⋆	0.26
41	`0.55`	√	0.09	⋆	⋆	⋆	√	√	0.57	⋆	⋆	⋆
42	√	√	⋆	0.76	⋆	⋆	√	√	`0.44`	`0.93`	⋆	0.29
43	`0.71`	`0.71`	⋆	0.22	⋆	⋆	`0.71`	`0.71`	⋆	0.22	⋆	⋆

Continued on next page

Table 1.17 – continued from previous page

	SIM-Diff						SIM-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
44	0.53	$\surd$	0.34	0.25	$\star$	$\star$	0.53	$\surd$	0.34	0.77	$\star$	0.29
46	$\surd$	$\surd$	0.07	$\surd$	$\star$	0.09	$\surd$	$\surd$	$\star$	$\surd$	$\star$	0.09
47	$\surd$	$\surd$	$\surd$	0.39	0.98	$\star$	$\surd$	$\surd$	$\surd$	0.39	0.98	$\star$
48	$\surd$	0.87	0.36	0.19	$\star$	0.12	$\surd$	0.87	0.36	0.19	$\star$	0.12
49	0.79	$\surd$	$\star$	0.6	$\star$	0.24	0.80	$\surd$	$\star$	0.6	$\star$	0.24
50	0.84	$\surd$	0.16	0.30	$\star$	0.28	$\surd$	$\surd$	0.29	0.30	$\star$	0.28
52	$\surd$	$\surd$	0.26	0.42	0.22	0.79	$\surd$	$\surd$	0.26	0.42	0.22	0.79
53	$\surd$	$\surd$	0.40	0.50	$\star$	0.15	$\surd$	$\surd$	0.68	0.50	$\star$	0.15
55	$\star$	$\surd$	$\star$	$\surd$	$\star$	$\surd$	$\star$	$\surd$	$\star$	$\surd$	$\star$	$\surd$
56	$\surd$	$\surd$	0.92	0.79	0.34	0.59	$\surd$	$\surd$	0.93	0.98	0.34	0.59
58	$\surd$	$\surd$	$\surd$	$\surd$	0.90	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.90	$\surd$
59	$\surd$	$\surd$	0.53	0.52	0.26	0.16	$\surd$	$\surd$	0.53	0.52	0.26	0.16
61	$\surd$	$\surd$	0.12	0.8	$\star$	0.8	$\surd$	$\surd$	0.66	0.8	$\star$	0.8
65	$\surd$	$\surd$	0.83	$\surd$	$\star$	$\surd$	$\surd$	$\surd$	0.83	$\surd$	$\star$	$\surd$
66	$\surd$	$\surd$	$\surd$	0.43	$\surd$	0.1	$\surd$	$\surd$	$\surd$	0.89	$\surd$	0.28
67	$\surd$	$\surd$	$\surd$	$\surd$	0.98	0.99	$\surd$	$\surd$	$\surd$	$\surd$	0.98	0.99
3	$\surd$		0.17		$\star$		$\surd$		0.2		$\star$	
6	$\surd$		$\surd$		0.85		$\surd$		$\surd$		0.85	
8	$\surd$		0.6		$\star$		$\surd$		0.6		$\star$	
10	0.7		0.37		$\star$		$\surd$		0.87		$\star$	
40	$\surd$		0.75		0.49		$\surd$		0.75		0.49	
45	$\surd$		0.18		$\star$		$\surd$		0.69		$\star$	
51	$\surd$		0.44		0.5		$\surd$		0.44		0.5	
54	$\surd$		0.6		$\star$		$\surd$		0.73		$\star$	
57	$\surd$		0.65		0.69		$\surd$		0.73		0.69	
60	$\surd$		0.23		0.75		$\surd$		0.23		0.75	
62	$\surd$		0.13		$\star$		$\surd$		0.13		$\star$	
63	$\surd$		$\surd$		0.97		$\surd$		$\surd$		0.97	
64	0.88		0.28		$\star$		0.88		0.19		$\star$	
Fits	66	52	59	48	24	30	66	52	59	48	27	34

Table 1.18: The frequentist results for the distance-based models for linear order model with $\tau = 0.50, 0.25$, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2). Rejections at a 0.05 level are marked $\star$. Perfect fits are checkmarks ($\surd$). Nonsignificant violations have their p-values listed. Nonsignificant violations where Session II replicates Session I are marked in `typewriter`. Frequentist p-values are computed only for vertices whose The Bayes factors are larger than 3.2.

	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
1	$\surd$	$\surd$	0.18	0.11		
2	`0.57`	$\surd$				
4	$\surd$	$\surd$	0.11			
5	0.11	0.36				
7	$\surd$	$\surd$	$\star$			
9	$\star$	$\surd$		$\surd$		0.65
11	$\surd$	0.99	0.08			
12	$\surd$	$\surd$	0.59	0.95	$\star$	$\star$
13	$\surd$	$\surd$	`0.28`	`0.31`	$\star$	$\star$
14	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	`0.94`
15	0.55	0.84				
16	0.21	$\surd$		$\surd$		0.15
17	$\surd$	$\surd$	0.07	0.86		$\star$
18	`0.45`	$\surd$	$\star$			
19	`0.88`	`0.94`				
20	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
21	`0.95`	`0.88`				
22	`0.69`	`0.94`				
23	`0.37`	`0.85`				
24	$\surd$	$\surd$	$\surd$	$\surd$	`0.51`	`0.97`
25	$\surd$	`0.95`	$\star$			
26	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
27	$\surd$	$\surd$	0.38			
28	`0.88`	$\surd$				
29	0.69	0.66				
30	`0.89`	`0.97`				
31	$\surd$	`0.97`				
32	$\surd$	$\surd$	$\surd$	$\surd$	`0.94`	$\surd$
33	0.15	$\surd$		0.76		$\star$
34	0.27	0.93				
35	$\surd$	$\surd$	`0.97`	$\surd$	$\star$	0.09
36	$\surd$	$\surd$	$\surd$	0.31	0.62	0.12
37	$\surd$	`0.56`	$\surd$		0.99	
38	$\surd$	$\surd$		$\star$		
39	$\surd$	`0.96`				
41	`0.67`	$\surd$		$\star$		
42	$\surd$	$\surd$	$\star$	0.76		$\star$
43	$\surd$	`0.97`				
44	`0.39`	$\surd$		0.07		

Continued on next page

Table 1.18 – continued from previous page

	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
46	√	√	$\star$	√		0.09
47	√	√	√	0.26	0.98	
48	√	0.95				
49	0.87	0.97				
50	√	0.98				
52	0.84	√				
53	√	0.82				
55	√	√	$\star$	√		√
56	0.97	0.89				
58	√	√	√	√	0.90	√
59	√	0.57				
61	√	0.95	0.12			
65	0.18	√	$\star$	√		√
66	√	√	√	$\star$	√	
67	√	√	√	√	0.98	0.99
3	√					
6	√		√		0.85	
8	√		0.37		$\star$	
10	0.83					
40	√					
45	√		0.18			
51	0.86					
54	√		0.21			
57	0.98					
60	0.99					
62	0.19					
63	√		√		0.97	
64	0.70					
Fits	66	54	25	22	13	14

Table 1.19: The Bayes factors for the distance-based models for LSO-Diff with τ = 0.50, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	LSO-Diff					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
1	1610	2831	27	29774	0	905322
2	2198	3274	41	8481	1	1114
4	3127	11	67	0	0	0
5	495	3200	0	5745	0	7
7	3076	1561	560	0	0	0
9	395	7462	0	1760182509	0	38423721779020
11	3480	5878	589	15751	0	17363
12	6224	0	23666734	0	6440	0
13	4980	1964	16545431	85782	98140	324
14	7763	7732	5112328732	3210053087	166566466964153	29422217255642
15	522	1216	43	22	6	0
16	650	7563	1	324675856	0	6827952
17	150	285	56	3287	1	2
18	932	4381	5962	12550819	0	20271587
19	4518	5009	165901	89746	385725	159105
20	7767	7767	7139015495	7260177363	1347595495416320	1889185527583140
21	158	1	0	0	0	0
22	1846	1424	5521	423	3993	125
23	314	193	5	7	0	1
24	7661	7732	2238292336	3319938407	1771801409029	57823651933818
25	2646	1000	48930	37429	1088	276
26	7766	7766	6849450022	6965697447	10455726266521210	14657815947195730
27	5415	2087	1825	39	0	2
28	17584	14845	327136	431368	669816	133171
29	1523	1674	523	165	100	4
30	9457	6456	31533	18926	28538	14300
31	16	1428	0	307	0	1
32	7754	7767	3406191657	7508705001	13449528546369	37128271277524150
33	0	68	0	1146	0	7
34	196	6037	3	336698	0	1036167
35	6875	7469	81406683	91886485	5612044	4917
36	7644	7344	1201461621	38547349	18817761875	2
37	7758	796	4317354679	0	36632280081557	0
38	5886	4902	107	26616	0	3
39	2557	4276	546	43602	130	71135
41	1075	1178	196	44	0	0
42	1271	6687	8469	36030731	606421	38499
43	582	2152	0	944	0	0
44	1664	3652	393	175094	0	10936312
46	3325	7594	2549	186315530	0	9503281
47	7757	3107	4005136235	47	14462004307998	0
48	3295	2176	1079	10328	311	9833
49	1685	7179	1	26922	0	18764
50	1460	4185	11	100547	0	209291
52	3318	11089	41794	1708981	66201	10786236
53	1839	4593	44	11551	4	4706
55	110	7767	68	7636141162	0	520498866316045
56	9403	4948	81867	186621	103874	460692
58	7736	7766	3187567485	6403539275	6957820180480	2707035601786514
59	10716	2925	81304	9590	32406	4615
61	3976	16381	40422	740836	0	2191633
65	5743	7760	379	4977987334	0	358393110785141
66	7765	3736	6143805440	1957413	2100335252104813	682524055
67	7739	7759	4761179599	4894911917	2174229652568243	255649291523146
3	1207		1		0	
6	7720		1922363623		241249785900	
8	6079		1403979		3	
10	929		23		0	
40	23582		301759		251198	
45	3282		1710		0	
51	7506		309349		985662	
54	4741		104		0	
57	11001		304292		872552	
60	6853		473816		1870149	
62	4353		59		0	
63	7760		4932983840		167656278060078	
64	2307		18		0	
Fits	66	52	57	49	34	39

Table 1.20: The Bayes factors for the distance-based models for LSO-Ratio with $\tau = 0.50$, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	LSO-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
1	367	4235	1	94296	0	385476
2	1081	1516	14	2860	0	375
4	4257	40	1349	0	0	0
5	801	2269	0	2870	0	18
7	3255	418	202	0	0	0
9	2788	2518	17	592610856	0	12935717429928
11	1882	3364	144	5313	0	5846
12	3206	0	8788265	0	2720	0
13	1804	825	5597928	29149	33091	109
14	2823	3651	1729725686	1182856254	56162644326473	12141021059537
15	250	3434	14	115	2	0
16	545	3348	0	112796422	0	2454094
17	1677	9012	5549	11670516	92	1150137
18	10095	9917	1010190	18481880	15	9628327
19	1818	2000	55852	30215	129858	53564
20	2665	2665	2405132964	2445951209	4537241111735642	6360729308037096
21	10	2	0	0	0	0
22	913	717	1859	142	1344	42
23	132	87	2	2	0	0
24	2786	2798	757312753	1123253730	597413478031	19496850457761
25	3269	3262	3281898	286263	237655	226145
26	2621	2621	2306044072	2345181767	3520021336981464	4934695455148333
27	3709	1748	1371	20	0	1
28	7105	7448	110205	161231	225499	44835
29	1433	937	176	65	34	1
30	3872	2665	10610	6372	9608	4814
31	7	864	0	220	0	0
32	2617	2621	1146782344	2527999270	4527913821222	1249959149376280
33	0	7	0	0	0	0
34	93	2334	1	113353	0	348834
35	6215	3230	148831040	31751180	1686321050	1696
36	3412	2991	417461995	13291239	6768849571	1
37	2619	582	1453549216	0	12332611336333	0
38	3444	3138	99	371586	0	4
39	2674	2724	211	14673	44	23948
41	3866	5734	38737	33748	1	0
42	5566	6912	402375	96578909	1118401	89322090
43	319	1290	0	578	0	0
44	1602	6865	361	4625889	0	198418867
46	1299	4346	358	71167527	0	9783316
47	2666	1954	1349333874	22	4869236375834	0
48	3599	857	383	3477	105	3310
49	1248	2942	0	9065	0	6317
50	3183	2084	29	33853	0	70460
52	1121	4489	14070	575347	22287	3631276
53	1402	1594	15	3889	1	1584
55	40	2621	22	2570903942	0	1.752309756973707e+16
56	3553	1725	27562	62828	34970	155096
58	3593	2665	1174137566	2157351166	2870367432697	911436203603761
59	5341	1119	32054	3229	10910	1554
61	6010	7358	1051610	249589	1	737832
65	3007	2619	141	1675967932	0	120656505852415
66	2664	14100	2069845899	26929803	707165273061241	1291442707
67	2655	2619	1604039738	1647998447	732044898268241	86066805723560
3	814		0		0	
6	3109		662768293		82814940101	
8	2597		485114		1	
10	4912		345		0	
40	10271		101925		84568	
45	5634		226423		0	
51	2573		104145		331831	
54	2198		38		0	
57	4961		102447		293752	
60	2843		159515		629601	
62	2622		29		0	
63	3117		1700705396		57552047888802	
64	997		6		0	
Fits	66	52	56	47	34	39

Table 1.21: The Bayes factors for the distance-based models for SIM-Diff with τ = 0.50, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	SIM-Diff					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
1	1096	2886	24	31385	0	954836
2	5305	5497	183	9104	1	1175
4	4901	23	80	0	0	0
5	1	1484	0	5896	0	7
7	9488	2627	25726	13	0	0
9	1577	8020	2312	1857773071	0	40529023106031
11	7381	18191	1390	21642	0	18313
12	9211	0	27321874	0	8326	0
13	6274	2112	17870580	90538	105541	342
14	8207	8312	5392188638	3388031065	175676256409322	31034311070190
15	4736	1240	18906	31	10	0
16	3704	8571	45	344137688	0	7212466
17	166	301	59	3467	1	2
18	973	3769	6288	13229642	0	21380272
19	12963	8349	181353	95420	406820	167806
20	8211	8211	7529820866	7657615137	14212976728830790	19925081392276660
21	3	0	0	0	0	0
22	4709	5170	5840	888	4211	132
23	2139	1300	6	10	0	1
24	8099	8174	2360821369	3501678999	1868704094092	60986120968964
25	1468	1048	39920	39478	1145	291
26	8210	8210	7224403696	7347014755	11027566844762140	15459475605581420
27	7665	8894	2106	80	0	2
28	32618	24099	354071	456942	706447	140454
29	2368	3439	553	184	106	4
30	27734	9550	36603	20054	30099	15082
31	8	1233	0	194	0	1
32	9744	8211	3678941408	7919747714	14463800575590	39158876476553460
33	1	34	0	105	0	1
34	5490	11590	18	364900	0	1092834
35	7268	7898	85863060	96916574	5918976	5186
36	8081	9514	1267232222	41657522	19846935703	2
37	8202	6761	4553696146	6	38635758734288	0
38	4204	2513	113	14709	0	3
39	7320	9669	624	46166	137	75025
41	1247	1341	208	47	0	0
42	1330	7070	8931	38003140	639587	40605
43	1571	2603	4	3753	0	0
44	873	3818	416	184795	0	11535531
46	3567	8024	2970	196514844	0	10023031
47	8340	7327	4227196753	296	15254402361779	0
48	7730	3531	1187	10922	328	10371
49	3782	16395	2	30771	0	19791
50	3121	6080	15	106290	0	220737
52	7031	14513	48056	1802480	69822	11376108
53	15401	6670	274	12973	4	4964
55	0	8211	0	8054160005	0	54896579105191460
56	17614	9365	89000	196855	109555	485887
58	8768	8210	3378637755	6754082832	7349648130325	2855087756594834
59	17480	3482	89026	10115	34179	4867
61	4273	28822	42663	799464	0	2311491
65	11873	8795	7086	5276379569	0	378575931581731
66	8346	3790	6484439107	2064487	2215416207012147	719852399
67	8182	8203	5021816844	5162869957	2293141788374076	269631164705841
3	3102		1		0	
6	8749		2037594606		254835708870	
8	11385		17127535		8444	
10	1484		35		0	
40	36536		320827		264935	
45	3745		1812		0	
51	12559		326321		1039566	
54	13441		16432		0	
57	20159		321863		920270	
60	17681		514000		1972426	
62	4447		207		0	
63	8204		5203026198		176825671102010	
64	5378		322		0	
Fits	64	52	60	51	35	38

Table 1.22: The Bayes factors for the distance-based models for SIM-Ratio with $\tau = 0.50$, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	SIM-Ratio					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
1	345	4042	1	94759	0	387410
2	3715	6847	168	3099	0	377
4	6648	113	1525	0	0	0
5	1	1095	0	2816	0	18
7	7781	815	8767	20	0	0
9	2444	2579	3208	596008212	0	13001842628587
11	3921	23178	324	91650	0	5877
12	4512	0	9667695	0	3351	0
13	2166	844	5761500	29317	33910	110
14	2844	3730	1738486079	1189637765	56444381681181	12203084019224
15	8958	1533	9451	162	13	0
16	5615	3613	94	113926627	0	2470201
17	576	4392	5418	5540871	92	1155327
18	4715	7493	1006865	18564433	15	9676626
19	13412	11500	58794	31837	130509	53835
20	2685	2684	2417314119	2458339009	4560001977408838	6392637617746312
21	1	0	0	0	0	0
22	5930	8314	1947	867	1351	42
23	2338	922	5	3	0	0
24	2807	2818	761148276	1128942577	600410375835	19594655522657
25	1941	2867	2551704	287554	238298	227281
26	2640	2640	2317723297	2357059209	3537679364161072	4959450130796080
27	5067	8261	1508	63	0	1
28	23140	19973	142827	164813	227500	45108
29	5410	3773	1619	103	185	1
30	18331	11930	13958	10435	9693	4859
31	3	689	0	65	0	0
32	3133	2640	1180272984	2540802612	4640034339692	125622951268380
33	0	7	0	0	0	0
34	6219	14697	12	155713	0	351946
35	5875	3251	149565687	31911995	1694780397	1704
36	3434	3698	419576273	13687158	6802805198	1
37	2638	8408	1460910933	16	12394477325624	0
38	2954	1680	104	192952	0	4
39	8223	10615	369	15389	44	24069
41	3183	4505	38815	34063	1	0
42	4042	6535	402940	97054942	1124012	89770170
43	874	992	1	1641	0	0
44	899	5751	365	4650583	0	199433150
46	1348	4355	397	71527941	0	9832393
47	2731	4342	1357070063	130	4894127134158	0
48	11320	3304	588	3604	105	3327
49	3259	12451	1	12013	0	6351
50	5370	4604	42	34601	0	70813
52	7994	12289	17113	588063	22408	3649506
53	29799	2861	1157	4227	2	1592
55	0	2641	0	2583924580	0	1.761100139290613e+16
56	15838	8990	122698	84947	49709	156479
58	3870	2685	1185902077	2168277318	2889206071463	916008381927407
59	13226	1890	34926	3248	10971	1561
61	5044	30267	1049212	272045	1	741563
65	6767	2828	3616	1692760927	0	121448392043037
66	2728	10273	2081712038	27049856	710780182278217	1297921170
67	2675	2638	1612163581	1656344924	735717168255501	86498555945857
3	2002		1		0	
6	3356		669409143		83358466592	
8	4645		5640120		2773	
10	3350		540		0	
40	24754		108954		85026	
45	5284		228560		0	
51	8971		114550		333622	
54	7912		19653		0	
57	23891		229564		306883	
60	17585		170211		632782	
62	2552		96		0	
63	3138		1709318855		57840755144812	
64	3297		69		0	
Fits	62	52	58	50	36	39

Table 1.23: The Bayes factors for the distance-based models for the linear order model with $\tau = 0.50$, 0.25, and 0.10. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	LO					
	$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.10$	
	S1	S2	S1	S2	S1	S2
1	58	72	1	15	0	0
2	2	3	0	0	0	0
4	64	15	0	0	0	0
5	0	0	0	0	0	0
7	34	30	0	0	0	0
9	0	70	0	16501711	0	360222391678
11	53	2	3	0	0	0
12	72	66	223317	209489	59	1728
13	73	72	167805	34451	929	0
14	73	73	47928082	30094254	1561560627789	275833286772
15	0	9	0	0	0	0
16	0	72	0	3043846	0	64012
17	35	69	0	39507	0	0
18	12	35	0	0	0	0
19	0	1	0	0	0	0
20	73	73	66928270	68064163	126337077695290	177111143210931
21	4	6	0	0	0	0
22	0	3	0	0	0	0
23	0	1	0	0	0	0
24	73	73	20984051	31124429	16610638202	542096736880
25	63	24	0	0	0	0
26	73	73	64213594	65303414	98022433748644	137417024504971
27	68	14	13	0	0	0
28	1	1	0	0	0	0
29	0	0	0	0	0	0
30	2	1	0	0	0	0
31	40	5	0	0	0	0
32	73	73	31933047	70394109	126089330122	348077543226816
33	0	72	0	37438	0	9
34	0	1	0	0	0	0
35	73	72	763574	861440	52544	46
36	73	69	11263735	361381	176416517	0
37	73	4	40475201	0	343427625765	0
38	41	59	0	0	0	0
39	12	1	0	0	0	0
41	10	54	0	0	0	0
42	64	73	1	337936	0	354
43	11	3	0	0	0	0
44	0	64	0	3	0	0
46	43	73	0	1746305	0	86812
47	73	29	37548153	0	135581290387	0
48	10	1	0	0	0	0
49	12	3	0	0	0	0
50	24	1	0	0	0	0
52	0	1	0	0	0	0
53	6	1	0	0	0	0
55	42	73	0	71588823	0	487967687171330
56	1	1	0	0	0	0
58	73	73	29883451	60033181	65229564192	25378458766751
59	2	0	0	0	0	0
61	69	1	383	0	0	0
65	11	73	0	46668632	0	3359935413611
66	73	54	57598176	0	19690642988484	0
67	73	73	44636067	45889800	20383402992845	2396712108029
3	17		0		0	
6	73		18022163		2261716743	
8	46		13076		0	
10	1		0		0	
40	5		0		0	
45	61		17		0	
51	0		0		0	
54	36		1		0	
57	1		0		0	
60	1		0		0	
62	4		0		0	
63	73		46246724		1571777606813	
64	3		0		0	
Fits	45	36	20	20	16	16

Table 1.24: The frequentist and Bayes factor results for the mixture models for Tversky (1969) data.

Panel A: The frequentist results for the mixture models.

	LSO-Diff	LSO-Ratio	SIM-Diff	SIM-Ratio	LO
1	$\surd$	0.68	0.14	0.40	0.34
2	0.28	0.13	0.36	0.22	0.63
3	0.62	0.31	0.06	0.14	$\star$
4	0.91	0.44	$\star$	0.10	0.30
5	0.70	$\star$	0.81	0.73	0.20
6	0.45	$\star$	0.47	$\star$	$\star$
7	0.20	0.10	0.20	0.10	$\surd$
8	0.67	$\star$	0.67	0.22	$\surd$
Fits	8	5	7	7	6

Panel B: The Bayes factors for the mixture models.

	LSO-Diff	LSO-Ratio	SIM-Diff	SIM-Ratio	LO
1	1119	11	333	6	0
2	7	6	19	43	3
3	27	53	14	22	0
4	60	42	21	1	1
5	588	2	1042	20	2
6	25	0	8	0	0
7	18	23	57	395	16
8	226	3	706	48	18
Fits	8	5	8	6	2

Table 1.25: The frequentist and Bayes factor results for the mixture models for Cash I and Cash II from Regenwetter et al. (2011a.)

Panel A: The frequentist results for the mixture models.

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	0.09	0.86	⋆	0.64	0.09	0.21	0.11	⋆	√	0.30
2	⋆	0.77	⋆	0.26	⋆	0.53	⋆	0.13	√	√
3	√	0.85	0.93	0.48	√	0.64	0.89	0.19	√	√
4	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	0.10	0.76
5	√	0.32	⋆	0.08	√	0.32	0.08	⋆	√	√
6	0.50	0.39	⋆	0.21	0.50	0.12	0.15	⋆	0.64	0.38
7	√	⋆	0.53	⋆	√	⋆	0.53	⋆	√	√
8	√	0.22	0.51	⋆	√	0.19	0.51	⋆	√	√
9	0.16	√	⋆	√	0.16	0.31	⋆	0.09	√	√
10	√	0.31	√	0.24	√	0.14	0.98	0.24	√	0.54
11	√	0.11	0.71	⋆	√	0.10	0.61	⋆	√	0.58
12	0.17	⋆	⋆	⋆	0.17	⋆	0.07	⋆	√	√
13	0.19	0.41	⋆	√	0.19	0.64	0.07	0.50	√	√
14	√	√	0.92	√	√	√	0.92	√	√	√
15	0.54	⋆	0.41	⋆	0.54	⋆	0.41	⋆	√	√
16	0.08	⋆	⋆	⋆	0.08	⋆	⋆	⋆	⋆	⋆
17	0.11	0.43	0.09	0.23	0.11	0.17	0.09	⋆	0.17	√
18	0.64	0.60	0.79	√	0.64	0.74	0.88	0.36	√	0.45
Fits	16	13	9	11	16	13	14	7	17	17

Panel B: The Bayes factors for the mixture model analysis.

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II	Cash I	Cash II
1	62	220	0	26	168	1	4	1	13	0
2	9	62	0	8	29	477	0	26	13	28
3	712711	3	0	12	1963269	1	18	15	1	11
4	0	0	0	0	0	0	0	0	0	4
5	15073	7	0	5	35234	8	1	13	3	10
6	57	243	0	12	106	219	0	20	5	4
7	7077	0	1	0	17366	0	17	0	8	15
8	83525	1	0	1	242219	0	3	4	2	5
9	2	1843	0	75	7	1255	0	43	9	20
10	6330	1	47	13	18985	2	733	142	10	2
11	84610	3	0	1	245758	1	14	6	5	2
12	9	1	0	2	16	0	0	0	4	8
13	11	138	0	93	30	510	1	356	13	13
14	707556	0	18	0	1916996	20	97	11	1	0
15	336	163	34	3	1053	4	280	0	19	17
16	315	7	0	0	966	0	1	0	0	0
17	7	69	4	9	22	117	35	8	1	13
18	135	21	98	130	415	30	449	166	19	3
Fits	16	11	5	11	17	9	9	12	12	12

Table 1.26: The Bayes factors for the mixture models for the 2012 experiment data. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1	20201	928	120	3	94	32	3	14122	338	348
2	3633	95	1197	2344	104719	4960	50237	85318	117	202
4	20991	0	47008	1	19788	0	307735	9	340	180
5	2638	418	416287	2282	0	13	0	93	0	27
7	24651	29465	23898	72	151502	24731	498911	246	69	282
9	145	0	60311	0	28	27	1724	7	0	0
11	89922	3561	941	118783	46379	382272	3699	10502329	254	268
12	7200912	0	593	0	753417	0	3131	0	51	2
13	0	244	0	0	38	4021	0	6	21	97
14	11518	1002	0	0	291	10	0	87	112	23
15	10224	14	47678	182223	173201	130	1964223	24245	25	48
16	276	1	5547	1	6796	364	111679	3	4	103
17	0	0	356740	24951	0	0	229	32944	180	247
18	0	179	2341	32400	0	54	624	9240	86	254
19	319	34	33	445	12959	2236	23113	90967	207	188
20	41938662431182	0	0	0	27094855541435520	501929	211119372	21636497	9	10
21	12	0	24	0	0	0	0	0	62	154
22	12	166	22	2688	779	9473	7663	474121	90	167
23	5	1	6	0	417	44	452	13	28	130
24	0	0	0	0	390813	20	320	0	47	65
25	9207	1	30888	30	296	0	2163	95	373	98
26	5	46	0	0	2	257	0	0	19	8
27	1376	28	709	11	10241	2992	25741	3624	232	233
28	32600	64527	20468	223921	1808685	2027060	2250479	11890370	234	246
29	1	35	1143	8963	47	1569	54455	70266	80	39
30	74879	2	26381	1635	4154974	282	3480295	154099	272	185
31	13	3472	0	1638	0	2864	0	843	362	284
32	310	7308	0	0	3556	2418123	1042	1297195723	3	5
33	0	0	0	0	0	0	0	0	1	194

Continued on next page

Table 1.26 – continued from previous page

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
34	155	21	48	20	8282	979	8475	13282	25	237
35	4570	391	0	74	1722	484	0	23786	204	221
36	658	8320	0	161	130	684967	0	22912	127	32
37	0	2452	0	4511	773	161310	0	679857	64	87
38	14198	48050	175	25866	9999	7764	1360	6329	337	251
39	31	53	23	106	2722	5150	7010	38266	315	239
41	0	0	918	569	3	3	5133	23434	126	301
42	3	1795	425	534	0	911	9040	1498	344	135
43	893	268	4	30	12016	1818	105	91	32	144
44	180	34	105	8	34	12	132	27901	7	226
46	360432	59860	2191	11	208084	41715	5688	561865	267	267
47	368971	13551	0	1268	1210534	128527	0	7068	40	70
48	54	1	339	0	3394	210	67949	183	329	176
49	87	2037	361	1291	8756	185948	49544	202855	116	264
50	35	2	1177	1	2444	97	56663	342	307	244
52	53	15	6	44	2398	2594	2883	18880	143	309
53	4479	6	64798	0	357859	315	18652262	19	121	99
55	2	12376	0	0	0	2095	0	0	7	5
56	5213	67	1714	75	220339	3242	305422	15106	222	235
58	5896	6	0	0	787853	777	0	0	19	14
59	29259	0	11607	0	693086	1	527322	3	241	76
61	8	13862	71	18837	77	894791	9544	4463528	154	294
65	113177	0	4432	0	1949120	150686	186862	200177	7	1
66	0	3205	0	1425	106	48	0	432691	3	291
67	284	69313	0	0	74	37820	0	0	20	55
3	22		1		702		317		195	
6	0		0		205		0		76	
8	153		26		7603		14832		1	
10	212		1149		1321		9104		49	
40	3196		40394		439280		5632975		358	
45	0		0		0		1		146	

Continued on next page

Table 1.26 – continued from previous page

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
51	326		9		17024		4047		218	
54	1332		7		41221		1240		5	
57	854		11792		56195		2896719		241	
60	5929		747		211015		278255		306	
62	36916		508		19321		579		10	
63	133		0		2333		0		25	
64	948125		535909		7756096		2190597		115	
Fits	54	37	47	33	56	47	49	46	62	51

Table 1.27: The frequentist analysis results for the mixture models for the 2012 experiment data. There are 67 participants in Session I (S1) and of which, 54 returned for Session II (S2).

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1	0.23	0.07	⋆	0.07	⋆	0.06	⋆	⋆	√	√
2	0.53	0.10	⋆	⋆	0.43	0.12	⋆	⋆	√	√
4	0.91	⋆	0.40	⋆	0.63	⋆	⋆	⋆	√	√
5	0.58	0.37	0.14	0.09	⋆	0.09	⋆	⋆	⋆	0.32
7	0.47	0.33	0.49	0.17	0.38	⋆	0.38	⋆	0.27	√
9	⋆	0.07	⋆	⋆	⋆	⋆	⋆	⋆	⋆	0.26
11	0.46	0.44	0.21	0.76	0.24	0.44	⋆	0.73	√	√
12	0.34	⋆	0.11	⋆	0.18	⋆	⋆	⋆	0.99	0.97
13	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	0.74	0.70
14	0.63	0.24	0.41	0.82	0.78	0.38	0.21	0.73	√	0.69
15	0.37	0.07	0.23	0.28	0.51	⋆	0.19	⋆	0.68	0.32
16	0.13	0.37	0.20	0.09	0.16	0.48	0.09	⋆	0.12	√
17	⋆	⋆	0.22	0.13	⋆	⋆	⋆	⋆	√	√
18	⋆	0.20	0.15	0.30	⋆	0.22	⋆	0.09	0.37	√
19	0.62	0.28	0.13	⋆	0.55	0.15	0.13	⋆	√	√
20	0.93	0.40	0.93	0.17	0.91	0.34	0.65	0.19	0.78	√
21	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	0.69	√
22	0.35	0.26	0.12	⋆	0.37	0.32	0.09	0.06	√	√
23	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	0.38	√
24	0.32	⋆	0.14	⋆	0.38	⋆	⋆	⋆	√	√
25	0.26	⋆	0.40	⋆	0.16	⋆	⋆	⋆	√	√
26	0.30	0.29	0.21	0.19	0.44	0.16	0.06	⋆	√	0.99
27	0.64	0.52	0.28	0.08	0.74	0.50	0.21	⋆	√	√
28	0.69	0.49	0.67	0.64	0.78	0.57	0.60	0.40	√	√
29	0.06	0.08	0.21	⋆	0.08	0.08	0.09	⋆	√	0.13
30	0.60	0.12	0.36	0.07	0.63	0.12	0.17	⋆	√	√
31	⋆	0.19	⋆	0.11	⋆	⋆	⋆	⋆	√	√
32	0.20	0.91	⋆	0.41	0.27	0.92	⋆	⋆	0.57	√
33	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	0.15	√
34	0.19	0.29	⋆	0.15	0.19	0.36	⋆	0.11	0.29	√
35	⋆	0.84	0.06	⋆	0.19	0.51	⋆	⋆	√	√
36	0.16	0.92	0.06	0.83	0.09	0.95	⋆	0.74	√	0.36
37	0.10	0.58	⋆	0.38	0.06	0.49	⋆	0.23	√	√
38	0.38	0.84	0.28	0.22	0.12	0.30	0.13	0.07	√	√
39	0.17	0.24	0.12	0.13	0.17	0.43	0.06	0.08	√	√
41	⋆	⋆	0.06	0.14	⋆	⋆	⋆	⋆	0.39	√
42	0.10	⋆	0.11	0.11	0.09	⋆	⋆	⋆	√	√
43	0.29	0.09	0.07	⋆	0.26	0.09	0.07	⋆	0.57	√
44	0.49	0.08	0.08	⋆	0.28	0.13	⋆	⋆	0.15	√
46	0.69	0.41	0.18	0.49	0.39	0.29	0.06	0.48	√	√
47	0.18	0.39	0.19	0.32	0.27	0.39	0.07	0.26	√	√

Continued on next page

Table 1.27 – continued from previous page

	LSO-Diff		LSO-Ratio		SIM-Diff		SIM-Ratio		LO	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
48	0.09	0.14	0.11	⋆	0.10	0.17	0.06	⋆	√	√
49	0.10	0.35	0.09	⋆	0.10	0.37	⋆	⋆	⋆	√
50	0.10	0.21	0.18	⋆	0.12	0.26	0.11	⋆	√	√
52	⋆	0.20	⋆	0.21	⋆	0.23	⋆	0.07	√	√
53	0.72	⋆	0.81	⋆	0.78	⋆	0.65	⋆	√	0.29
55	⋆	√	⋆	0.67	⋆	√	⋆	0.37	0.17	√
56	0.28	0.09	0.26	⋆	0.35	0.18	0.21	⋆	√	√
58	0.20	⋆	0.10	⋆	0.12	0.07	⋆	⋆	0.74	0.91
59	0.66	⋆	0.17	⋆	0.57	⋆	0.08	⋆	√	0.39
61	0.28	0.45	0.24	0.56	0.12	0.43	0.21	0.55	√	√
65	0.72	0.32	0.81	⋆	0.61	0.29	0.78	⋆	0.40	0.67
66	0.30	0.69	⋆	0.88	0.57	0.91	⋆	0.88	0.32	√
67	0.26	0.26	0.20	0.29	0.13	0.19	⋆	0.07	√	√
3	0.11		0.16		0.14		0.08		√	⋆
6	⋆		⋆		⋆		⋆		√	⋆
8	0.12		⋆		⋆		⋆		0.69	⋆
10	⋆		0.24		⋆		0.14		0.32	⋆
40	0.34		0.19		0.34		0.36		√	⋆
45	⋆		⋆		⋆		⋆		√	⋆
51	0.28		⋆		0.45		⋆		√	⋆
54	0.23		⋆		0.15		⋆		0.73	⋆
57	0.69		0.43		0.66		0.40		√	⋆
60	0.66		0.12		0.45		0.08		√	⋆
62	⋆		⋆		⋆		⋆		0.41	⋆
63	0.06		⋆		0.25		⋆		√	⋆
64	0.46		0.44		0.58		0.16		√	⋆
Fits	51	40	46	30	49	37	30	18	64	54

Chapter 2

Parsimonious Testing of Transitive or Intransitive Preferences: Reply to Birnbaum (2011)[1]

Michel Regenwetter, University of Illinois at Urbana-Champaign, USA
Jason Dana, University of Pennsylvania, USA
Clintin P. Davis-Stober, University of Missouri, USA
Ying Guo, University of Illinois at Urbana-Champaign, USA

[1]The analysis results for the random-LSO-Ratio model for Tversky (1969)'s set and Cash I and Cash II in Regenwetter et al. (2011a) are reported in the published paper, under Section "Alternative Intransitive Models."

Abstract

Birnbaum (2011) raises important challenges to testing transitivity. We summarize why an approach based on counting response patterns does not solve these challenges. Foremost, we show why parsimonious tests of transitivity require at least five choice alternatives. While Regenwetter et al.'s approach still achieves high power with modest sample sizes for five alternatives, pattern-counting approaches face the difficulty of combinatoric explosion in permissible response patterns. Even for fewer than five alternatives, if the choice of how to "block" individual responses into response patterns is slightly mistaken then intransitive preferences can mimic transitive ones. Meanwhile, statistical tests on proportions of response patterns rely on similar "independent and identically distributed" (iid) sampling assumptions as tests based on response proportions. For example, the hypothetical data of Birnbaum (2011, Tables 2 and 3) hinge on the assumption that response patterns are properly blocked, as well as sampled independently and with a stationary distribution. We test an intransitive lexicographic semiorder model on Tversky's (1969) and Regenwetter et al.'s (2011) data and, consistent with Birnbaum's (2011) concern, we find evidence for model mimicry in some cases.

Regenwetter, Dana, and Davis-Stober (2011), henceforth RDDS, investigated transitivity of preferences through powerful and parsimonious quantitative tests. Regenwetter et al. (2010, 2011) made extensive efforts to spell out and eliminate unnecessary and, in many cases, unwanted assumptions in the literature. To protect against serious aggregation paradoxes that create the false appearance of intransitivity, they moved from aggregation across people to individual choice data. By collecting repeated choices from the same individual, they avoided the assumption, implicit in single observations, that preferences are fixed. These repeated choices were interspersed with rich and similar looking distractors to keep respondents from recognizing choice alternatives, in an effort to approximate iid sampling.

Birnbaum (2011) describes an alternative quantitative approach to testing transitivity on within-subject data. He agrees with RDDS's substantive conclusion that evidence for intransitivity is lacking and also with their criticism of weak stochastic transitivity. He argues, however, that RDDS do not go far enough in criticizing past approaches, particularly because they analyze proportions of binary responses. He contrasts their approach with that of Birnbaum and Gutierrez (2007), who, instead, analyze proportions of binary response patterns. A *response pattern* is the series of responses that a respondent makes across a complete repetition of all unique gamble pairs. Using a hypothetical example, Birnbaum shows how the RDDS approach could conclude that choices are transitive when in fact the decision maker has intransitive preferences, a phenomenon we call *model mimicry*. His example shows how analyzing patterns, as Birnbaum and Gutierrez (2007) do, could diagnose this true intransitivity, because their approach identifies a preference distribution while RDDS do not. Birnbaum's comment also questions the untested RDDS assumption that a respondent's choices form iid draws from a probability distribution over preference orders, finding it "empirically doubtful" that responses to the same gamble pair or to related gamble pairs by the same respondent are statistically independent.

Our reply focusses on a small number of key points. We start and end with the central question: *How can we test transitivity of preferences in a parsimonious and statistically powerful fashion?*

The importance of considering at least five choice alternatives.

If one is going to draw conclusions from failing to reject transitivity, as both Birnbaum (2011) and RDDS do, it is crucial that transitivity be a strong hypothesis that we would expect to overturn if untrue. Looking at Birnbaum's Table 2, one can see that with 3 gambles, there are 8 possible response patterns. Six of these 8 patterns (75%) are transitive. We can frame this problem in terms of the RDDS approach by imagining a cube (see Regenwetter et al., 2010, for a visualization) in which one's probability of choosing A over B, from 0 to 1, is one dimension, and the probabilities of choosing B over C and A over C, from 0 to 1, are the other two dimensions. Inside this unit cube, 67% of the space satisfies the triangle inequalities that RDDS use to test transitivity. Retaining transitivity with three gambles is is not very informative because most conceivable data sets will support transitivity.

On the other hand, if one uses 5 gambles, as RDDS did, there are 10 unique gamble pairs and $2^{10} = 1{,}024$ possible response patterns. Of these, only 120 patterns (12%) are transitive. In terms of the RDDS tests, the 10 binomial choice probabilities create a 10-dimensional unit hypercube, inside of which only 5% of the space satisfies the triangle inequalities (see Regenwetter et al., 2010). Thus, in either approach, moving from 3 gambles to 5 gambles transforms transitivity from an almost meaninglessly lax hypothesis to a strong hypothesis with serious potential for rejection. For this reason, it is crucial that any approach that retains transitivity be able to do so with at least 5 choice alternatives.

Because there are 1,024 possible response patterns for 5 gambles, combinatoric explosion will pose a formidable problem for any pattern counting approach. Consider again Birnbaum's (2011) Table 2. The example data use 200 repetitions so that there are 25 observations for each of the 8 possible response patterns. To obtain an average of 25 observations per pattern with 5 gambles, one would now need 1,024 patterns times 10 decisions (there are 10 gamble pairs per pattern) times 25 observations per pattern = 256,000 decisions in this hypothetical experiment, not including any filler choices between blocks. The RDDS approach estimates 10 binomials for the 10 unique gamble pairs, and thus requires only 250 decisions for an experiment with a comparable number of 25 observations per cell. A similar combinatoric explosion occurs when respondents are allowed to express indifference, because then there are even many more permissible patterns (see Table 2.1).

Since a strong test of transitivity requires 5 gambles, the RDDS approach has a major advantage over pattern counting approaches in that it scales comfortably to that many choice alternatives. It does so, however, because it makes certain iid sampling assumptions that Birnbaum (2011) questions, especially because these assumptions are not tested. If pattern counting will prove difficult in parsimonious testing environments, does it at least free us of such assumptions? To answer this question, let us explicate what each approach assumes.

What does each approach assume about iid sampling?

Consider Table 2 of Birnbaum (2011). Model 1 tests the iid assumptions of RDDS on hypothetical data. The table summarizes information about 200 observed response patterns, with each pattern consisting of three decisions, for a total of 600 decisions. Since we could assign a 0 or 1 to each item (as Birnbaum does for patters of three in his Table 2) and since all sequences of 600 responses are allowable, there are 2^{600} degrees of freedom in the data, representing all possible temporal series of responses in the experiment.

Birnbaum's chi squared test, his Eq. (3), for Model 1 has 4 degrees of freedom. RDDS's goodness-of-fit test would assume 3 degrees of freedom for these data. How do both approaches reduce the degrees of freedom so dramatically?

Birnbaum's test in Table 2 uses a *blocking assumption* that classifies decisions as response patterns using the temporal sequencing of the data: Responses to the 3 unique choice pairs constitute a block and the response made on the first replicate of a choice, e.g., between A and B, cannot be swapped with the response made on the second replicate. The chi-squared test does not, however, consider the temporal sequence in which the 200 patterns were observed, but simply counts how often each of the 8 kinds of patterns occur, reducing the data to 7 degrees of freedom (the number is 7 because once 7 pattern frequencies are observed, the 8th is determined since we know the total number of patterns observed). The chi squared test, then, assumes that these 200 *response patterns* are iid draws from a distribution over 8 binary relations. The 3 choice probabilities in Model 1 (the probabilities of choosing A over B, B over C, and A over C) are free parameters consuming 3 more degrees of freedom, leaving 7-3 = 4 degrees of freedom in the chi squared test. For brevity we skip similar calculations for other tests in Birnbaum's (2011) Tables 2 and 3.

RDDS differ in that they do not preserve any temporal information about the sequence of these decisions. They assume that the 600 *individual responses* are iid draws from a probability distribution over preference rankings. The 3 binomial probabilities of choosing A over B, B over C, and A over C are the only things to be estimated and hence, RDDS reduce the data complexity from 2^{600} to 3 degrees of freedom. RDDS's iid assumption is stronger than the one used in Birnbaum's Table 2 because iid sampling of 600 responses implies iid sampling of 200 response patterns, but not vice versa.

Birnbaum's Model 1 uses the assumptions of blocking and iid sampling of patterns to show how one would test and reject iid sampling of preferences underlying individual decisions. If applied to real data, this would imply a significant rejection of RDDS's iid assumption, but it would not evaluate the blocking and iid pattern assumptions that it uses. Our Table 2.1 summarizes these and other insights. Pattern counting approaches like Birnbaum's (2011), then, necessarily require their own iid assumption. We are unsure how these assumptions would be tested and Birnbaum (2011) does not appear to provide suggestions.

If pattern counting approaches also involve untested iid, as well as blocking, assumptions, do they at least free us from model mimicry because they actually identify preference states? This is the question we consider next.

Does analyzing response patterns solve the problem of model mimicry?

While RDDS estimate binary choice probabilities and test transitivity, they do not estimate the unique distribution of preferences that their model assumes exists. Birnbaum (2011) gives a hypothetical mixture of intransitive states that RDDS would falsely diagnose as supporting transitivity, while an analysis of response patterns detects intransitivity.

What if the data are incorrectly blocked? Much like Birnbaum's (2011) thought experiment showed potential model mimicry in the RDDS approach, we give a simple example using 3 choice alternatives where pattern counting is vulnerable to model mimicry. Imagine a decision maker had only intransitive true preferences, which with three gambles means either: $a \succ b, b \succ c, c \succ a$ coded by Birnbaum (2011) as 001, or its reverse, $b \succ a, c \succ b, a \succ c$ coded as 110. This decision maker is presented the following sequence of paired comparisons: $(a,b)_1, (b,c)_2, (a,c)_3, (a,b)_4, (b,c)_5, (a,c)_6, (a,b)_7, (b,c)_8, (a,c)_9$, where the subscript denotes the trial number. According to the blocking assumption, this decision maker remains in a fixed preference state throughout each complete replication of all unique choice pairs, i.e., the trial intervals 1-3, 4-6, and 7-9. But imagine that the blocking assumption is slightly incorrect in that the first block is shortened by one single trial. Thus, the decision maker is in a fixed preference state, say, 001 for trials 1-2 and 6-8, but 110 for trials 3-5, and 9. The sequence of 9 responses, 000111000, when blocked, will appear as the follows:

Block 1: trials 1-3 = 000, i.e., $a \succ b, b \succ c$, and $a \succ c$.

Block 2: trials 4-6 = 111, i.e., $b \succ a, c \succ b$, and $c \succ a$.

Block 3: trials 7-9 = 000, i.e., $a \succ b, b \succ c$, and $a \succ c$.

An analysis of response patterns would mistakenly conclude that this decision maker is transitive and makes no errors. Hence, an intransitive process would have mimicked a transitive one. The problem is not attributable to the simplicity of this example. For five gambles there are 1,024 possible patterns. If preferences switch at times other than between blocks, then the real preference patterns may be unrecoverable. If the decision maker's true preference states do not last equally long, nearly any response pattern (transitive or not) is mathematically possible, even if the decision maker expresses her true preference with no error and has only a few true preference states. Birnbaum (2011, p. 7 in page proofs) raised the possibility of a pattern counting approach in which the blocks and their lengths are estimated from the data. Such an approach, however, would introduce a great deal of model complexity, as each change in true preference is a parameter to estimate and each additional observation provides one more possible transition between preference states.

Birnbaum (2011) has hit upon an important problem in model mimicry that we agree warrants investigation. But pattern counting approaches do not solve the problem. The choice of how to block data into patterns always creates the possibility of model mimicry. Detecting and accommodating violations of the blocking assumption seems to us a major challenge. Within the approach of RDDS, we now show how one can test for specific intransitive processes to try to identify model mimicry.

Alternative intransitive models.

We ask whether certain alternative models may provide an alternative account for Tversky's (1969) and

RDDS's data on five choice alternatives. We formalize Tversky's (1969) idea of lexicographic semiorders. We focus on one probabilistic heuristic model for choices among two outcome cash gambles with one nonzero positive outcome and one zero outcome, such as RDDS's Cash I and Cash II gambles and Tversky's (1969) gambles (see Table 2.2).

Attribute Order. The decision maker sequentially considers the attributes: With some unknown probability, she first considers the chance of winning, otherwise payoff.

Threshold of discrimination. Each attribute has a threshold. If two gambles differ by a factor greater than the threshold on the attribute under consideration, then the decision maker (DM) chooses the option that is 'better' on that attribute. Otherwise, he moves to the next attribute. We allow the two thresholds to be random variables with any joint distribution whatsoever, hence permitting many preference states.

Indifference. If the DM has considered both attributes without a conclusion, then we assume, for simplicity, that he chooses either alternative with probability one half.

Table 2.2 shows Tversky's (1969) gambles and the ratios for each attribute. If the DM always considers payoff before chance, with fixed payoff and chance thresholds of 1.18 and 1.2, then, writing $\succ$ for strict preference and $\sim$ for indifference, she has the preferences on the left of Panel 3 (from top). Notice the intransitive cycle $a \succ e, e \succ c, c \succ a$. The right side of Panel 3 shows the choice probabilities for a DM with just that one preference.

For brevity, we only sketch the model and its test. The lexicographic semiorder given in Panel 3 of Table 2.2 is but one of 111 such preferences one can derive for Tversky's gambles as one varies the sequence of attributes and the threshold values. Likewise, for RDDS's Cash I and Cash II gambles, there are similar collections of 111 distinct lexicographic semiorders. The model states that the probability of choosing i over j equals the probability that she currently strictly prefers i to j plus $\frac{1}{2}$ times the probability that she is indifferent between i and j. This mixture model is similar to that of RDDS, with two main differences: 1) Instead of 120 linear orders, we consider 111 lexicographic semiorders. 2) This model does not force "complete" preferences, rather it permits indifference among choice alternatives.

Just as the linear ordering model translates geometrically into a convex polytope, so do these lexicographic semiorder models translate into polytopes. We leave a formal discussion for elsewhere. Table 2.2 summarizes a number of interesting findings. For Tversky's data, we found the model to be rejected for three out of eight participants, whereas we found it rejected in nine out of 18 participants in RDDS's Cash I replication of Tversky (1969) and in seven out of the same 18 in Cash II. This speaks directly to Birnbaum's (2011) concern about model mimicry: Several participants are fit by both the linear ordering model and the lexicographic semiorder model. Is there an explanation for this finding? It is important to realize that many lexicographic semiorders are transitive, and some are linear orders. We therefore determined the collection of binomial distributions that form the overlap between the linear ordering model and the lexicographic semiorder model and tested those intersections, too. We rejected that overlapping model on only three out of eight participants for Tversky's data and on only nine out of 18 participants for Cash I as well as only six out of 18 participants for Cash II. Hence, we agree with Birnbaum's concern about model mimicry: Parts of the lexicographic semiorder model can mimic parts of the linear order model, and, indeed, both models fit a large proportion of the participants.

If we give positive probability only to the 104 intransitive cases among the 111 lexicographic semiorders, then we reject the model on 15 out of 18 participants in both Cash I and Cash II. Incidentally, the priority heuristic (Brandstätter et al., 2006) is one of the 104 intransitive preference states in this model for each gamble set. We thus reject a broad generalization of that intransitive heuristic in which the order of the "reasons" and the thresholds may, but need not, vary on 15 out of 18 participants. This analysis also addresses Birnbaum's (2011) concern about the stationarity component of RDDS's iid sampling assumption. If the binomial probabilities change over time but always satisfy a given mixture model, then the average binary choice probabilities will also satisfy that model because mixture models form convex polytopes. Hence, we expect that a false fit of the linear order model caused by nonstationary probabilities in the lexicographic semiorder model requires that the latter model also fit. For a pattern-counting approach, protection against violations of its stationarity assumptions appears to us more complex, due to the complicated interplay among blocking, iid sampling, many degrees of freedom, and limitations in the amount of data one individual can provide.

How can we achieve parsimonious testing of transitivity?

Tables 2.1 and 2.2 summarize our findings. Both Birnbaum Birnbaum (2011) and Regenwetter et al. (2010, 2011) deliberately eliminate common, and often undesirable, assumptions in the literature. Both make related iid sampling assumptions to reduce the complexity inherent in a binary sequence of hundreds or thousands of decisions in an experiment, so as to achieve statistical testability. Birnbaum also makes blocking and independent error assumptions that RDDS do not make. RDDS could enlarge their polytopes to allow additional errors. Such extensions would reduce the parsimony of their test, making transitivity easier to fit.

Much of Regenwetter et al. (2010, 2011) aims at classifying and dissecting the implicit or explicit assumptions made in various approaches and developing parsimonious quantitative tests. Since every test makes some assumptions, testing these assumptions is valuable. Even more valuable is to use assumptions that need to hold only approximately for the substantive conclusions to be valid. We have provided some evidence that RDDS's conclusions are somewhat robust to possible violations of stationarity. More work is needed to evaluate the robustness of either approach to violations of all their respective assumptions, such as the independent sampling assumption in each approach. Another avenue to enhance parsimonious testing is methodological innovation. Sophisticated statistical methods may help pattern counting approaches overcome some of the formidable challenges posed by combinatoric explosion. Within the RDDS approach, where limits on mathematical knowledge pose a greater obstacle than attainable sample size, novel efforts are under way to test polytopes without having to fully characterize their mathematical properties.

ACKNOWLEDGMENTS: M. Regenwetter is supported by *DRMS* Award No. SES #08-20009 of the National Science Foundation. Any opinions expressed in this publication are those of the authors and do not necessarily reflect the views of universities or the funding agency.

Table 2.1: Communalities (centered) and differences (split) betweeen Birnbaum's (2011) and Regenwetter et al.'s (2010, 2011) approaches to testing transitivity of preferences, as well as key strengths and weaknesses. Birnbaum (1984) used similar calculations to count patterns.

<table>
<tr><th>Birnbaum (2011)</th><th>(Regenwetter et al., 2010, 2011)</th></tr>
<tr><td>Uses blocking assumptions in order to
group observed binary responses into patterns.</td><td>Does not make blocking assumptions.</td></tr>
<tr><td colspan="2">For the hypothetical data of Birnbaum's Tables 2 & 3:
Reduces 2^{600} degrees of freedom in an observed ordered sequence of 600 binary data to</td></tr>
<tr><td>7 degrees of freedom
for $2^{\binom{3}{2}} = 8$ pattern proportions,
by assuming that
the observed ordered sequence of 200 patterns
originates form iid sampling
of 200 binary relations with no indifference.</td><td>3 degrees of freedom
for $\binom{3}{2} = 3$ binary choice proportions,
by assuming that
the observed ordered sequence of 600 responses
originates from iid sampling
of 600 strict linear orders.</td></tr>
<tr><td>Can identify a unique preference distribution
from pattern frequencies.</td><td>Cannot identify a unique preference distribution
from response frequencies.</td></tr>
<tr><td colspan="2">Tests transitivity under assumption that
the decision makers are never indifferent between any two prospects.
(Tests "strict linear orders.")</td></tr>
<tr><td colspan="2">Can be extended to a more direct test of transitivity ("strict weak orders")
by permitting additional "no preference" response category ("ternary choice")</td></tr>
<tr><td>at cost of combinatoric explosion:
3 prospects: $3^{\binom{3}{2}} - 1 = 26$ degrees of freedom,
5 prospects: $3^{\binom{5}{2}} - 1 = 59,049$ degrees of freedom.</td><td>without combinatoric explosion:
3 prospects: $\binom{3}{2} \times 2 = 6$ degrees of freedom,
5 prospects: $\binom{5}{2} \times 2 = 20$ degrees of freedom.</td></tr>
<tr><td colspan="2">An experiment with 5 choice prospects and 20 observations per empirical cell corresponds to</td></tr>
<tr><td>$20 \times \binom{5}{2} \times 2^{\binom{5}{2}} = 204,800$ binary choices or
$20 \times \binom{5}{2} \times 3^{\binom{5}{2}} = 11,809,800$ ternary choices,
plus fillers between blocks.</td><td>$20 \times \binom{5}{2} = 200$ binary choices or
$20 \times \binom{5}{2} = 200$ ternary choices,
plus distractors between choices.</td></tr>
<tr><td>Assumes each observed pattern is
composed of a preference relation
and independent errors.</td><td>Does not assume errors, but could enlarge
their model (the "polytope") to accommodate
interdependent errors, with risk of overfitting.</td></tr>
<tr><td colspan="2">Can fall victim to model mimicry.</td></tr>
<tr><td colspan="2">Avoids aggregation across individuals.</td></tr>
<tr><td colspan="2">Avoids descriptive modal choice analysis.</td></tr>
<tr><td colspan="2">Uses quantitative goodness-of-fit methodology.</td></tr>
<tr><td colspan="2">Does not assume that preferences are induced by independent random utilities.</td></tr>
<tr><td colspan="2">Concludes overall that:
Existing evidence for intransitivity of preferences is not compelling.</td></tr>
</table>

Table 2.2: Tversky's (1969) gambles. From top to bottom, the first panel shows the chance of winning and payoff value for each of the five gambles. The second panel shows the chance ratios (left), and the payoff ratios (right) as decimals. The third panel provides one of the 111 lexicographic semiorders one can obtain this way (left) and the choice probabilities of a DM who has only that one single preference. The bottom panel shows the result of testing the lexicographic semiorder mixture model on Tversky's 8 participants and on RDDS's 18 participants for Cash I and Cash II (with $\alpha = 0.05$). The last column of that panel shows the number of simultaneous inequality constraints tested. PH denotes the Priority Heuristic. RDDS = Regenwetter, Dana, & Davis-Stober (2011).

Gamble	Chance of winning	Payoff
a	7/24	\$5.00
b	8/24	\$4.75
c	9/24	\$4.50
d	10/24	\$4.25
e	11/24	\$4.00

Chance Ratios (column/row)					Payoff Ratios (row/column)				
Gamble	b	c	d	e	Gamble	b	c	d	e
a	1.143	1.286	1.429	1.571	a	1.053	1.111	1.176	1.250
b	-	1.125	1.250	1.375	b	-	1.056	1.118	1.188
c		-	1.111	1.222	c		-	1.059	1.125
d			-	1.100	d			-	1.063

DM with fixed thresholds (payoff: 1.18; chance: 1.2), who considers payoff before chance of winning (Comparing row gambles to column gambles)									
Binary Preference					Choice probabilitiy				
Gamble	b	c	d	e	Gamble	b	c	d	e
a	$\sim$	$\prec$	$\prec$	$\succ$	a	$\frac{1}{2}$	0	0	1
b	-	$\sim$	$\prec$	$\succ$	b	-	$\frac{1}{2}$	0	1
c		-	$\sim$	$\prec$	c		-	$\frac{1}{2}$	0
d			-	$\sim$	d			-	$\frac{1}{2}$

Data Set	Number of distinct lexicographic semiorders	Number of Rejections	Number of Constraints
Tversky	111 (incl. PH)	3 of 8 participants	24
Cash I	111 (incl. PH)	9 of 18 participants	24
Cash II	111 (incl. PH)	7 of 18 participants	1956

References

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological review*, 77(3):153.

Bar-Hillel, M. and Margalit, A. (1988). How vicious are cycles of intransitive choice? *Theory and Decision*, 24(2):119–145.

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):22–36.

Birnbaum, M. H. (1984). Transitivity in the big ten. *Bulletin of the Psychonomic Society*, 22(4):351–353.

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comment on Regenwetter, Dana, and Davis-stober (2011). *Psychological Review*, 118:674–682.

Birnbaum, M. H. and Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organizational Behavior and Human Decision Processes*, 104(1):96–112.

Block, W. E., Barnett, I., et al. (2012). Transitivity and the money pump. *Quarterly Journal of Austrian Economics*, 15(2):237–251.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2):409–432.

Busemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3):432–459.

Carbone, E. and Hey, J. D. (2000). Which error story is best? *Journal of Risk and Uncertainty*, 20(2):161–76.

Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology*, 53(1):1–13.

Davis-Stober, C. P., Brown, N., and Cavagnaro, D. R. (2015). Individual differences in the algebraic structure of preferences. *Journal of Mathematical Psychology*, 66:70 – 82.

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

Guo, Y. (2018). A review of intransitive theories of decision making under risk or uncertainty. Qualifying exam paper.

Guo, Y. and Regenwetter, M. (2014). Quantitative tests of the Perceived Relative Argument Model: comment on loomes (2010). *Psychological Review*, 121(4):696–705.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–89.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, 39(3-4):633–640.

Hey, J. D. (2005). Why we should not be silent about noise. *Experimental Economics*, 8(4):325–345.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326.

Iverson, G. J. and Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences*, 10(2):131–153.

Jeffreys, H. (1998). *The Theory of Probability, the 3rd Edition.* Oxford University Press.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Klugkist, I. and Hoijtink, H. (2007). The bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, 51(12):6367–6379.

Leland, J. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Unvertainty*, 9(2):151–172.

Lichtenstein, S. and Slovic, P. (2006). *The construction of preference.* Cambridge University Press, New York.

Loomes, G., Moffatt, P. G., and Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, 24(2):103–130.

Loomes, G. and Sugden, R. (1987). Some implications of a more general form of regret theory. *Journal of Economic Theory*, 41(2):270–287.

Loomes, G. and Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39(3-4):641–648.

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24(2):178–191.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* John Wiley, New York.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, 46(1):1–26.

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology*, 41(1):79–87.

Myung, J., Karabatsos, G., and Iverson, G. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, 49(3):205–225.

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., and Stevens, J. R. (2017). Heterogeneity and parsimony in intertemporal choice. *Decision*, 5(2):63–94. http://dx.doi.org/10.1037/dec0000069.

Regenwetter, M., Dana, J., and Davis-Stober, C. P. (2011a). Transitivity of preferences. *Psychological Review*, 118(1):42–56.

Regenwetter, M., Dana, J., Davis-Stober, C. P., et al. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Psychology*, 1:148.

Regenwetter, M., Dana, J., Davis-Stober, C. P., and Guo, Y. (2011b). Parsimonious testing of transitive or intransitive preferences: Reply to Birnbaum (2011). *Psychological Review*, 118(4):684–668.

Regenwetter, M. and Davis-Stober, C. P. (2012). Choice variability versus structural inconsistency of preferences. *Psychological Review*, 119(2):408–416.

Regenwetter, M., Stober, C., Lim, S., Cha, Y.-C., Cavagnaro, D., Guo, Y., Popova, A., and Zwilling, C. (2014). QTEST: Quantitative testing of theories of binary choice. *Decision*, 1(1):2–34.

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, 46(1):145–153.

Silvapulle, M. and Sen, P. (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. John Wiley & Sons, New York: New York.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38(2):332–382.

Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., and Friston, K. J. (2007). Comparing hemodynamic models with dcm. *NeuroImage*, 38(3):387 – 401.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., et al. (2014). Xsede: accelerating scientific discovery. *Computing in Science & Engineering*, 16(5):62–74.

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

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.