RATIONALITY OR IRRATIONALITY OF PREFERENCES? QUANTITATIVE TESTS OF DECISION THEORIES

BY

YING GUO

DISSERTATION

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

Urbana, Illinois

Doctoral Committee:

Professor Michel Regenwetter, Chair
Professor Hua-Hua Chang
Associate Professor Aron Barbey
Associate Professor Daniel Newman
Associate Professor Hans-Friedrich Köhn

Abstract

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 is a very fundamental element of utility and plays an important role in many major contemporary theories of decision making under risk or uncertainty. One has to be very careful about claiming violations of transitivity of preferences. In my thesis, I present a comprehensive analysis of several decision heuristics that permit intransitive preferences: the lexicographic semiorder model (Tversky, 1969), the similarity model (Rubinstein, 1988), and perceived relative argument model (PRAM, Loomes, 2010a), as well as several transitive decision theories: the linear order model and 49 versions of Cumulative Prospect Theory (CPT, Tversky and Kahneman, 1992a). For each decision theory, I use two kinds of probabilistic specifications 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. I test these probabilistic models on data sets from different experiments, using both frequentist (Davis-Stober, 2009, Iverson and Falmagne, 1985, Silvapulle and Sen, 2005) and Bayesian (Myung et al., 2005) order-constrained, likelihood-based statistical inference methods. This thesis is one of the largest scale projects for a systematic evaluation of both transitive and intransitive decision theories. The quantitative analyses in this paper consumed about 822,000 CPU hours 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). Individual model selection using Bayes factors shown extensive heterogeneity across participants and stimulus sets. In general, the overall conclusion is that Cumulative Prospective Theory and Perceived Relative Argument Model was systematically violated, and the intransitive heuristics performed reasonable well.

To my parents,

Yumei Ding and Zhizhong Guo,

And my husband,

Weihua Zheng,

And my kids,

Tyler and Staci,

For making me who I am today.

Acknowledgments

I thank Dr. Michel Regenwetter for his insights and assistance in the preparation of this paper. I thank Dr. Barbey, Dr. Chang, Dr. Köhn, and Dr. Newman for being in my committee and for their valuable suggestions. 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 and 2011 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

2AFC	Two-alternative forced-choice.
BF	Bayes factor.
CPT	Cumulative Prospect Theory.
GBF	Group Bayes factor.
LO	The linear order model.
LSO-Diff	The lexicographic semiorder model using a linear utility function $u(x) = x$ for utility.
LSO-Ratio	The lexicographic semiorder model using a log utility 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.
PRAM	Perceived relative argument model.
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 a linear utility function $u(x) = x$ for utility.
SIM-Ratio	The similarity model using a log utility function $u(x) = log(x)$ for utility.

Chapter 1

A Literature Review of Intransitive Theories of Decision Making under Risk or Uncertainty

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), and Cumulative Prospect Theory (CPT, Tversky and Kahneman, 1992b). Transitivity of preferences is a very 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. However, in the past few decades, researchers have provided much empirical evidence and seem to have agreed that transitivity of preferences is violated in human and animal decision makers (see, e.g., Brandstätter et al., 2006, González-Vallejo, 2002, Loomes and Sugden, 1987, Tversky, 1969). Transitivity of preferences is very central to many prominent theories in psychology and economics, and we have to be very careful about claiming violations of transitivity of preferences.

Overall, this paper reviews several major intransitive theories of decision making under risk or uncertainty, summarizes the empirical studies testing these theories, and discusses some methodological problems with these studies. I start with the latter.

1.2 Methodological Problems

In this section, I will discuss some common methodological flaws in the studies testing transitivity and/or intransitivity of preferences. As I mention above, many scholars have claimed that transitivity of preferences is violated in human and animal decision makers. However, in the studies where researchers concluded violations of transitivity of preferences, there are pervasive methodological problems in collecting, modeling, and analyzing the empirical data. In fact, by default, it is easier to violate transitivity of preferences (i.e., satisfy intransitivity of preferences). For example, when people choose from all ten possible pairwise comparisons of five objects, there are only 120 transitive preferences whereas there are 904 intransitive preferences; there are much more intransitive preference patterns than transitive ones (Regenwetter et al., 2011a). When we collect no data, the "a priori" is that people have intransitive preferences. Therefore, any claims of empirical violations of transitivity of preferences require rigorous evidence.

Before I discuss the methodological problems, I will illustrate one example of rigorous testing of transitivity of preferences. Preference is defined as people's attitude towards a set of items and used by many theories in psychology and economics (Lichtenstein and Slovic, 2006); 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. Tversky modeled choices as probabilistic. We need to figure out how variable choices are related to the 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, a problem that was only solved in the past decade with a breakthrough in order-constrained, likelihood-based inference (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 challenges. 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: the *distance-based* probabilistic specification models preferences as deterministic and views the observed variation in binary choice as errors/trembles, which is also called an *error model*; the *mixture-based* probabilistic specification (i.e., *random preference*) models preferences as probabilistic and views the observed variation in binary choice as variation in preferences or a reflection of uncertain preferences. For the second part of Luce's challenge, Regenwetter et al. (2014) employed order-constrained, likelihood-based statistical tests, with both the frequentist and Bayesian likelihood-based statistical inference framework for binary choice data with order-constraints on each choice probability (Davis-Stober, 2009, Iverson and Falmagne, 1985, Myung et al., 2005, Silvapulle and Sen, 2005).

Now, I go back to the methodological problems in the existing literature about transitivity. One problematic approach some studies employed is *pattern counting*, which involves counting suspicious observations, such as *the number of cyclical choice triplets* among a group of participants — this number is used to represent the *degree of intransitivity* (Bradbury and Nelson, 1974, Chen and Corter, 2006, Gonzalez-Vallejo et al., 1996, Mellers et al., 1992, Riechard, 1991). Regenwetter et al. (2010, 2014) showed that the degree of intransitivity is not monotonically related to the goodness of fit of a probabilistic model of transitivity. In other words, a large number of cyclical choice patterns does not mean a significant violation of a model of transitivity. Thus, any conclusions derived from pattern counting as the data analysis method are in question.

Other studies combine pattern counting with hypothesis testing, in which the hypotheses are wrongly specified. For example, one commonly used hypothesis is that certain intransitive patterns occur significantly more often than expected by chance. Regenwetter et al. (2010) showed that this approach could lead to a conclusion that preferences are intransitive, and at the same time, to the conclusion that they are consistent with transitivity on the same data, which is paradoxical. Another commonly used hypothesis is that a predicted cyclical pattern occurs significantly more often than its reverse (Loomes et al., 1991). Regenwetter et al. (2010) pointed out that the probability of the predicted cyclical pattern could be close to zero, but as long as the predicted cycle occurs significantly more often than its reverse, researchers would conclude that there is evidence for violations of transitivity. However, such evidence does not seem to tell how well transitive theories can account for people's choices.

Another common problematic approach is to conduct multiple binomial tests. For example, one could use two separate binomial tests to see: whether the probability of x chosen over y is larger than $\frac{1}{2}$ and the probability of y chosen over z is larger than $\frac{1}{2}$. Such tests would inflate Type I error (McNamara and Diwadkar, 1997, Schuck-Paim and Kacelnik, 2002, Shafir, 1994, Waite, 2001); in other words, one may accumulate false significant results. One may use the Bonferroni correction to fix Type I error, but doing

so would reduce the power rapidly (Hays, 1988). The solution is to test all constraints simultaneously (Regenwetter et al., 2014).

Another mistake that many empirical decision studies have made is that researchers often use between-participant modal choice (i.e., which choice alternative do most participants choose?). The *Condorcet Paradox* of social choice theory shows that even when each individual in a group has transitive preferences, the aggregated preferences by the majority rule can be intransitive and, therefore, the majority choices can be cyclical (Condorcet, 1785). In other words, a group can choose A over B, B over C and C over A, even though no individuals would make these choices. Hence, any conclusions derived from aggregating data among participants are not trustworthy. Theories of individual decision making would best be tested separately for each individual. Overall, scholars who doubt that preferences are transitive should formulate parsimonious alternatives and test those rigorously. We now proceed to consider such theories.

There are six major prominent intransitive theories of decision making under risk and uncertainty in the literature. The rest of the paper reviews each one of them: Section 1.3 reviews the additive difference model (Tversky, 1969) and its special cases; Section 1.4 discusses Regret Theory (Loomes and Sugden, 1982, 1987); Section 1.5 reviews three different heuristic models: lexicographic semiorder models (Tversky, 1969), similarity models (Leland, 1994, Rubinstein, 1988), and the priority heuristic (Brandstätter et al., 2006); Section 1.6 discusses the perceived relative argument model (Loomes, 2010b).

1.3 Additive Difference Model

Tversky (1969) introduced an *additive difference model* (also proposed by Morrison, 1962) to describe people's preferences when making decisions under uncertainty. The additive difference model predicts intransitive preferences for alternatives with multiple attributes/dimensions. The model is compensatory in the sense that the attractive attributes of an alternative can compensate for the less attractive ones.

Let x and y be two alternatives with elements of the form $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, where x_i $(i = 1, \ldots, n)$ is the value of alternative x on dimension i and y_i is the value of alternative y on dimension i. For example, a *dimension/attribute* of a gamble can be a monetary outcome or a probability of a monetary outcome. Suppose that u_i is a real-valued *utility* function, ϕ_i is a real-valued, increasing, and continuous *difference* function, and $x \succ y$ means that the decision maker prefers x to y. The *additive difference* model predicts the following:

$$x \succ y \Leftrightarrow \sum_{i=1}^{n} \phi_i\Big(u_i(x_i) - u_i(y_i)\Big) > 0, \tag{1.1}$$

$$\text{where } \phi_i(-r) = -\phi_i(r), \text{ for all i.}$$

In the additive difference model, the alternatives are first processed by making within-attribute evaluations. Next, the results of these within-attribute comparisons are added up to determine the preference.

Let $\epsilon > 0$ be a threshold value. One example of step functions works as follows:

$$f(r) = \begin{cases} 1, & \text{when } r > \epsilon; \\ 0, & \text{when } r \leq \epsilon. \end{cases}$$

When one or more of the difference functions ϕ_i in Display 1.1 are step functions, the *similarity model* (which will be discussed in Section 1.5) is a special case of the additive difference model.

Tversky (1969) stated the following theorems to indicate when transitivity holds for the additive difference model.

- For $n \geq 3$, transitivity holds if and only if all the difference functions ϕ_i in Display 1.1 are linear (i.e., $\phi_i(r) = t_i r$, for some positive t_i and for all i).
- For $n = 2$, transitivity holds if and only if the difference functions ϕ_1 and ϕ_2 applied to the two dimensions are identical except for a change of unit (i.e., $\phi_1(r) = \phi_2(sr)$).
- For $n = 1$, transitivity is always satisfied.

In other words, if none of the three conditions specified above holds, transitivity must be violated somewhere for the additive difference model.

Both Luce (1978) and Fishburn (1980) proposed the *lexicographic additive difference model* for alternatives with two attributes, because they did not think that the additive difference model captures the lexicographic character of some choices. The lexicographic additive difference model imposes a lexicographic ordering on the attributes while also using the additive difference model. When the within-attribute difference does not exceed the threshold on that attribute, the decision maker uses an additive difference structure. Suppose $x = (x_1, x_2)$ and $y = (y_1, y_2)$, where x_1 and y_1 are the *dominant attribute* (the first considered attribute by a lexicographic ordering) of x and y, respectively. Luce (1978) used within-threshold additivity for the following representation, in which δ is a real-valued function of dominant attribute x_1 and y_1, and the difference function ϕ in Display 1.1 is linear:

$$x \succ y \Leftrightarrow \begin{cases} u_1(x_1) > u_1(y_1) + \delta(y_1), \\ \text{ or} \\ u_1(x_1) \leq u_1(y_1) + \delta(y_1) \ \& \ u_1(y_1) \leq u_1(x_1) + \delta(x_1) \ \& \ u_1(x_1) + u_2(x_2) > u_1(y_1) + u_2(y_2). \end{cases}$$

Fishburn (1980) used the within-threshold additive differences for the following representation, in which ϵ is a positive threshold value.

$$x \succ y \Leftrightarrow \begin{cases} u_1(x_1) > u_1(y_1) + \epsilon, \\ \text{ or} \\ |u_1(x_1) - u_1(y_1)| \leq \epsilon \ \& \ \phi_1(u_1(x_1) - u_1(y_1)) + \phi_2(u_2(x_2) - u_2(y_2)) > 0. \end{cases}$$

Luce (1978) used a "variable" lexicographic threshold (i.e., a function δ of x_1 and y_1) and Fishburn (1980) used a "constant" lexicographic threshold (i.e., a fixed threshold ϵ).

Bouyssou and Vansnick (1986), Fishburn (1990, 1991), and Vind (1991) considered the mathematical properties of the earlier models and showed that those models can be generalized to a more general class. The *nontransitive additive skew symmetric model* is obtained by relaxing the subtractivity requirements in Display 1.1. Suppose w_i is a real-valued function on attribute i and skew symmetric (i.e., $w_i(x_i, y_i) + w_i(y_i, x_i) = 0$). The nontransitive additive skew symmetric model predicts:

$$x \succ y \Leftrightarrow \sum_{i=1}^{n} w_i(x_i, y_i) > 0. \tag{1.2}$$

Fishburn (1992) added a positive weight π_i for each dimension where $\sum_1^n \pi_i = 1$, and assumed the same utility function u for all the attributes. Hence, Display 1.1 becomes the following:

$$x \succ y \Leftrightarrow \sum_{i=1}^{n} \pi_i \phi_i \Big(u(x_i) - u(y_i) \Big) > 0. \tag{1.3}$$

Fishburn explained a topological approach to derive Display 1.3 from Display 1.2.

Butler (1998) adapted the additive difference model to the case of gambles with state-contingent consequences. This special case of the additive difference model works in the following way. First, Butler defined each attribute i as a "state of the world" and assumed that each attribute i has a probability π_i. Second, Butler interpreted each ϕ_i in Display 1.1 as a product of the form $\phi_i = \pi_i \phi$, using a fixed ϕ. The difference utility function ϕ is constant across attributes and is assumed to take the form of a power function, such as $\phi(r) = r^\beta$, where $\beta > 0$. Third, Butler assumed the subjective utility function u to be linear. Fourth,

Butler normalized the magnitudes of the differences of two gambles on an attribute as fractions of the largest attribute in the gamble pairs (i.e., $max = max\Big\{\cup_{i=1}^{n}\{x_i, y_i\}\Big\}$). Still using the two alternatives x and y, the special case predicts:

$$x \succ y \Leftrightarrow \sum_{i:x_i>y_i} \pi_i \left(\frac{x_i - y_i}{max}\right)^{\beta} > \sum_{i:y_i>x_i} \pi_i \left(\frac{y_i - x_i}{max}\right)^{\beta}. \tag{1.4}$$

Butler (1998) derived a n-act version of Display 1.4 and demonstrated that the n-act version is the choice-rule equivalent of a generalized form of "Regret Theory" (for details of "Regret Theory," please refer to Section 1.4).

Bouyssou and Pirlot (2002, 2004) studied the additive difference model while replacing additivity and subtractivity by decomposability requirements. Suppose F is a function that increases in all its arguments. The *decomposable model* predicts the following:

$$x \succ y \Leftrightarrow F\Big(u_1(x_1), \ldots, u_n(x_n)\Big) > F\Big(u_1(y_1), \ldots, u_n(y_n)\Big).$$

Bouyssou and Pirlot (2002, 2004) proposed the *nontransitive decomposable model*, where w_i is a real-valued function on the attribute i and skew symmetric. The model works as follows:

$$x \succ y \Leftrightarrow F\Big(w_1(x_1, y_1), \ldots, w_n(x_n, y_n)\Big) > 0.$$

The nontransitive decomposable model could be viewed both as a generalization of the decomposable model by dropping transitivity and as a generalization of the nontransitive additive skew symmetric model by dropping additivity. Bouyssou and Pirlot (2002, 2004) stated several axioms for this model.

In sum, this section introduced the additive difference model and its variations. The family of additive difference models only models preferences and is purely algebraic and deterministic. I have not found any studies testing these models empirically; therefore, no data have been collected to test these models. One paper, Tversky (1969), specified a probabilistic specification for the additive difference model. Let F be any normal or logistic distribution function, and the probabilistic model works as follows:

$$P(x, y) = F\left(\sum_{i=1}^{n} \phi_i[u_i(x_i) - u_i(y_i)]\right).$$

However, Tversky did not collect any data or perform any statistical tests for this model. All of these papers in this section focus on theoretical aspects, mathematical structure, and properties of the family of the additive difference models. Because no data have been ever collected and no statistical analysis has ever been performed, we have yet to investigate whether these models can help better understand human

behavior.

1.4 Regret Theory

Loomes and Sugden (1982) started from the premise that a decision maker feels disappointed and experiences regret when he makes a choice and finds out that the other choice would have led to a better outcome. They tried to model this idea and named their theory *Regret Theory*. Almost at the same time, Bell (1982, 1983) independently suggested that incorporating regret into decision theory would help predict a decision maker's behavior better than expected utility theory. Both Loomes and Sugden (1982) and Bell (1982, 1983) developed Regret Theory to explain some paradoxes and effects that cannot be explained by prospect theory (Kahneman and Tversky, 1979), such as the Allais paradox. In the Allais paradox, there are two gamble pairs:

Gamble $1A$: (\$1 million, 100%) vs. Gamble $1B$: (\$5 million, 11%; \$1 million, 89%; \$0, 1%)

Gamble $2A$: (\$1 million, 11%; \$0, 89%) vs. Gamble $2B$: (\$5 million, 10%; \$0, 90%)

Allais paradox within a person is defined as choosing $1A$ and $2B$ or choosing $1B$ and $2A$. Most people choose Gamble $1A$ over Gamble $1B$, whereas a majority of people select Gamble $2B$ over Gamble $2A$. Loomes and Sugden (1982) and Bell (1982) used Regret Theory to explain the Allais paradox. They stated that in the first choice, people may feel angry or regret if they take the gamble over the sure \$1 million; however, in the second choice, regret plays little or no role. Please note that in published analyses of the Allais paradox, researchers typically use the aggregated data among a group of participants. This is susceptible to aggregation artifacts. We do not know how many people actually make individual choices like those in the Allais paradox.

In Regret Theory, preferences are defined over actions. An *action* is an n-tuple of consequences (one consequence for each state of the world) and decision makers know all possible consequences. Regret theory applies to a very specific type of decision: no matter what decision people make, they find out the consequence of their choice as well as those of the other choice options that they have not picked. For example, people decide whether to take an umbrella before they go outside. In either case, they find out what the result would have been otherwise. For instance, if a person does not take an umbrella and it rains, he gets wet and regrets not having the umbrella. Often in our daily life, we find out the consequences of both the choices we made and the choices we could have made.

Suppose there are a finite number n states of the world. Each state i has probability π_i, where $0 < \pi_i \leq 1$ and $\sum_1^n \pi_i = 1$. These probabilities may be subjective probabilities that represent the decision maker's

confidence in the occurrence of the states or objective probabilities which the decision maker knows about. I write z_{ji} for the consequence of action j in the event that state i occurs. Suppose a decision maker faces the choice between two actions $A_1 = (z_{11}, \pi_1; \ldots; z_{1n}, \pi_n)$ and $A_2 = (z_{21}, \pi_1; \ldots; z_{2n}, \pi_n)$. If he chooses A_1 and state i occurs, then he receives the consequence z_{1i}. The decision maker experiences z_{1i} and knows that he has missed out on z_{2i} (i.e., *the forgone act*) because of his decision. If he had chosen A_2, he would have received z_{2i}. The decision maker experiences regret or rejoicing based on the comparison of z_{1i} and z_{2i}. If z_{1i} is smaller than z_{2i} , he experiences the unpleasant sensation of regret due to the feeling of a loss; otherwise, he experiences the pleasant sensation of rejoicing when he has done better than he might have otherwise. Regret theory can predict some effects/paradoxes that could not be explained by prospect theory, such as the Allais paradox, the certainty effect (Tversky and Kahneman, 1986), and certain kinds of intransitive preferences.

Suppose there is a difference function ψ, and Z represents a set of consequences. For all r and $s \in Z$, $r \succeq s \Leftrightarrow \psi(r, s) \geq 0$. Regret theory (Loomes and Sugden, 1982) predicts that

$$A_1 \succeq A_2 \Leftrightarrow \sum_{i=1}^{n} \pi_i \psi(z_{1i}, z_{2i}) \geq 0. \tag{1.5}$$

If ψ is linear, Regret Theory is equivalent to expected utility theory. The function ψ also satisfies three restrictions.

- ψ is strictly increasing in its first argument and non-decreasing in its second: for all r, s and $t \in Z$, if $\psi(r, s) \geq 0$, then $\psi(r, t) \geq \psi(s, t)$. In other words, $\psi(s, t)$ increases when the consequence s is substituted with a prefered consequence r.
- ψ is skewed symmetrically: for all r and $s \in Z$, $\psi(r, s) = -\psi(s, r)$.
- ψ is a convex function: for all r, s, and $t \in Z$, if $\psi(r, s) > 0$, $\psi(r, t) > 0$, and $\psi(s, t) > 0$, then $\psi(r, t) > \psi(r, s) + \phi(s, t)$. This property is also referred to as *regret aversion* and it postulates a disproportionate aversion to large regrets.

Let $g, h \in (1, \ldots, n)$. Suppose a_g and b_h are consequences of gambles, where a_g and $b_h \in Z$. Let p_g be the probability of the consequence a_g, and q_h be the probability of the consequence b_h. Now, a decision maker has to choose between two gambles, $G = (a_1, p_1; \ldots; a_n, p_n)$ and $H = (b_1, q_1; \ldots; b_n, q_n)$. Using the function ψ in Display 1.5, the skew-symmetric bilinear function (Fishburn, 1991) is defined as follows:

$$G \succeq H \Leftrightarrow \sum_{g=1}^{n} \sum_{h=1}^{n} p_g q_h \psi(a_g, b_h) \geq 0. \tag{1.6}$$

When a decision maker chooses between two gambles, G and H, one needs to define the matrix of state-contingent consequences in order to use Regret Theory to predict his choice. Please recall that the probability π_i in Display 1.5 represents the probability for each state i. If the probability distributions of the consequences of Gambles G and H — $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$ — are statistically independent, one can define $n \times n$ states of the world. For a state where Gamble G yields a_g and Gamble H yields b_h, the state occurs with the probability $p_g q_h$. Putting together all the information in Display 1.5,

$$G \succeq H \Leftrightarrow \sum_{g=1}^{n} \sum_{h=1}^{n} p_g q_h \psi(a_g, b_h) \geq 0.$$

This formula is exactly the same as Display 1.6. The skew-symmetric bilinear function is equivalent to Regret Theory when the probability distributions of the consequences of a gamble pair are statistically independent (Loomes and Sugden, 1987).

Loomes et al. (1991), Loomes and Taylor (1992), Starmer and Sugden (1998), and Humphrey (2001) reported experimental results supporting Regret Theory. They tested Regret Theory against empirical data by providing evidence that the regret cycle predicted by Regret Theory occurs more often than its reverse among all the participants. As mentioned in Section 1.2, this hypothesis is problematic and cannot tell whether Regret Theory accounts for those participants' choices. Also, these studies used the aggregated data from a group of participants. Pooled data are susceptible to aggregation artifacts and do not tell us about the individual performance. Therefore, we cannot conclude anything about whether Regret Theory can explain people's behavior.

The key component of Regret Theory is that it assumes that regret and rejoicing can influence the satisfaction a decision maker experiences from his choice. Researchers have suggested that when participants learn about a forgone act, they feel better if their outcome is better and they feel worse if their outcome is worse. This claim has been supported using psychological (Inman et al., 1997, Mellers et al., 1999), physiological (Camille et al., 2004), and neurophysiological evidence (Coricelli et al., 2005).

Humphrey (2004) provided a modified version of Regret Theory, *the feedback-conditional Regret Theory*. Image a decision maker still faces the option of two actions, $A_1 = (z_{11}, \pi_1; \ldots; z_{1n}, \pi_n)$ and $A_2 = (z_{21}, \pi_1; \ldots; z_{2n}, \pi_n)$. There are two different utility functions in feedback-conditional Regret Theory, $m(\cdot, \cdot)$ and $o(\cdot, \cdot)$. The function $m(\cdot, \cdot)$ shows the anticipated utility when the chosen option fully reveals the state of the world, which is what Regret Theory models. The function $o(\cdot, \cdot)$ describes the anticipated utility when the chosen option does not fully reveal the state of the world. Three restrictions are imposed on these two utility functions.

1. For all r, $s \in Z$ (where Z represents a set of consequences), if $r > s$, then $m(r,s) > o(r,s)$. This property means that people experience more rejoicing when the state of the world is fully revealed.

2. For all r, $s \in Z$, if $r > s$, then $m(s,r) < o(s,r)$. This property means that people experience more regret when the state of the world is not fully revealed.

3. For all r, $s \in Z$, if $r > s$, then $o(s,r) - m(s,r) > m(r,s) - o(r,s)$. This property means that revealing the state of the world has greater impact on the anticipated regret than rejoicing.

The feedback-conditional Regret Theory describes a modified function $M(\cdot,\cdot)$, which describes the modified anticipated utility of having chosen z_{1i} and having missed out on z_{2i} as the forgone act, when state i occurs. Humphrey (2004) described the difference function ψ in Display 1.5 as $\psi(z_{1k}, z_{2k}) = M(z_{1k}, z_{2k}) - M(z_{2k}, z_{1k})$. The modified function $M(\cdot,\cdot)$ in feedback-conditional Regret Theory is written as $M(\cdot,\cdot) = m(\cdot,\cdot) + o(\cdot,\cdot)$. Humphrey reported the empirical implications of feedback-conditional Regret Theory but did not provide any statistical tests of this theory.

Birnbaum and Schmidt (2008) tested Regret Theory and a special case of Regret Theory. They considered the case when the difference function ψ in Display 1.5 becomes

$$\psi(z_{1i}, z_{2i}) = \begin{cases} 1, & \text{when } z_{1i} > z_{2i}, \\ 0, & \text{when } z_{1i} = z_{2i}, \\ -1, & \text{when } z_{1i} < z_{2i}. \end{cases}$$

This special case of Regret Theory is called *the majority rule* or *the most probable winner model*. Birnbaum and Schmidt employed a model to test both Regret Theory and the majority rule, called a "true and error model." A *true and error model* assumes that each choice has a different but fixed error rate and each person has a different but fixed true preference pattern. The error rate for a choice is estimated from preference reversals between repeated presentations of the same choice using a chi-square statistical test. Birnbaum and Schmidt used 15 gamble pairs in their experiment and repeated each pair twice. They concluded that the occasional errors that occurred during the participants' decision processes could explain the cyclical choices of the participants. Birnbaum and Schmidt found that few participants showed a repeated intransitive pattern. They concluded that their data did not support Regret Theory. However, Birnbaum and Schmidt used a true and error model to analyze his data, which requires knowing which observations to glue together to make a pattern. In other words, the true and error model requires artificial "blocking." The analysis results can change with different ways of blocking (Cha et al., 2013). Because this true and error model is

problematic, we do not know much about the performance of Regret Theory and majority rule from the study by Birnbaum and Schmidt (2008).

Raeva et al. (2010) studied how regret and rejoicing impact people's decisions for intertemporal choices. They asked participants to make a risky decision prior to choosing between an intertemporal pair. They provided feedback on the risky decision, triggering the decision maker to experience regret or rejoicing before choosing between these intertemporal choice options. Raeva et al. (2010) concluded that regret and rejoicing experienced prior to an intertemporal choice influence the way people relate to the future; the experience of regret makes people more unwilling to wait, whereas the experience of rejoicing makes people more willing to wait.

Baillon et al. (2015) investigated Regret Theory using a true and error model (Birnbaum and Schmidt, 2008) on both pooled and individual data.. They found that the pooled data showed regret aversion, whereas the individual data showed both regret aversion and rejoicing seeking. Baillon et al. found no evidence that a regret cycle occurred more often at either the pooled or the individual level. They reported that there was no correlation between the number of regret cycles and regret aversion, and concluded there was little evidence of intransitive choices when a true and error model was used. The conclusions in the studies of Baillon et al. (2015) are not trustworthy, as they employed some problematic test methods, such as pattern counting with a wrong hypothesis that regret cycles occur more often than its reverse, and a true and error model with arbitrary blocking.

In summary, most of the papers studying Regret Theory tested the hypothesis that the regret cycle occurs more often than its reverse, and used data pooled across participants. As discussed in Section 1.2, the hypothesis that the regret cycle occurs more often than its reverse is problematic. The probability of the predicted cyclical pattern could be close to zero, but as long as the predicted cyclical pattern occurs significantly more often than its reverse cycle, researchers would conclude that there is evidence to support Regret Theory. Aggregating data among participants is also problematic because of the potential for aggregation artifacts. Thus, we still do not know much about whether Regret Theory can explain human behavior.

Regret theory models deterministic hypothetical constructs. Birnbaum and Schmidt (2008) and Baillon et al. (2015) specified a probabilistic model of Regret Theory and tested the probabilistic specification using a true and error model (Birnbaum, 2004, Birnbaum and Gutierrez, 2007). However, the assumption of blocking in the true and error model has been questioned (Cha et al., 2013, Regenwetter et al., 2011b).

Even though Regret Theory models decisions under uncertainty and is specified in terms of the state of the world, one could still use gamble pairs to test Regret Theory. I have shown that when the probability distributions of the consequences of a gamble pair are independent, then the skew-symmetric bilinear function

is equivalent to Regret Theory (Loomes and Sugden, 1987). In general, one could design an experiment with appropriate stimuli, collect repeated choices for each participant, specify different probabilistic models of Regret Theory, and test the probabilistic models of Regret Theory against laboratory data using order-constrained, likelihood-based inference methods (Regenwetter et al., 2014).

1.5 Heuristic Models

In this section, I review three intransitive heuristic models, including lexicographic semiorder models (Tversky, 1969), similarity models (Leland, 1994, Rubinstein, 1988), and the priority heuristic (Brandstätter et al., 2006). These three intransitive heuristic models are illustrated using Tversky's (1969) stimulus set (for details, see Panel A of Table 1.1). Tversky's stimulus set comprises five different gambles: a, b, c, d, and e. Alternative x is written as $x = (x_1, \ldots, x_n)$, where x_i $(i = 1, \ldots, n)$ is the value of alternative x on dimension i. In Tversky's stimulus set, there are four dimensions $(n = 4)$ in each gamble—(x_1, x_2, x_3, x_4)—in which x_1 is the maximum gain, x_2 is the probability of the maximum gain, x_3 is the minimum gain ($x_3 = \$0$ in Tversky's gambles), and x_4 is the probability of minimum gain $(x_4 = 1 - x_2)$. 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 winning \$5 and a $\frac{17}{24}$ chance winning nothing. The gambles are designed such that the expected values increase in the probabilities of maximum gains, whereas they decrease in the maximum gains. The probability of maximum gain of each gamble increases in equal steps $\left(\frac{1}{24}\right)$, whereas the maximum gain of the corresponding gamble 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.

1.5.1 Lexicographic Semiorder Model

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 are intransitive heuristic decision models.

In this paragraph, I will explain how a lexicographic semiorder works. 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, $x \prec_i y$ to denote that the decision maker prefers y to x on attribute i, and $x \sim_i y$ to denote that the decision maker is indifferent between x and y on attribute i. A lexicographic semiorder model works as follows:

1. The decision maker considers gamble attributes sequentially, for example, first the maximum gain and then the probability of maximum gain, or first the probability of maximum gain and then the maximum

Table 1.1: Tversky's (1969) gambles. Panel A shows the probabilities of maximum gains, maximum gains, and expected values for each of the five gambles. Panel B shows the differences in the probabilities of maximum gains among pairs. Panel C shows the differences of the maximum gains 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 gain	Gain (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 maximum gain 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 maximum gain 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	-	$\succsim$	$\succ$	$\succ$	$\succ$
b		-	$\succsim$	$\succ$	$\succ$
c			-	$\succsim$	$\succ$
d				-	$\succsim$

Panel E: A similarity model[2]

Preferences by probability					
lottery	a	b	c	d	e
a	-	$\succsim$	$\succsim$	$\succsim$	$\succ$
b		-	$\succsim$	$\succsim$	$\succsim$
c			-	$\succsim$	$\succsim$
d				-	$\succsim$

Preferences by gain					
lottery	a	b	c	d	e
a	-	$\succsim$	$\succ$	$\succ$	$\succ$
b		-	$\succsim$	$\succ$	$\succ$
c			-	$\succsim$	$\succ$
d				-	$\succsim$

Binary preference relation					
Lottery	a	b	c	d	e
a	-	$\succsim$	$\succ$	$\succ$	$\succsim$
b		-	$\succsim$	$\succ$	$\succ$
c			-	$\succsim$	$\succ$
d				-	$\succsim$

1. It is the binary preference pattern predicted by a lexicographic semiorder model if a decision maker considers the probabilities before the maximum gains, and uses a probability threshold of $\frac{3.5}{24}$ and a gain threshold of $.3.

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 gain threshold of $.3.

gain. For each attribute i, the decision maker uses a threshold $\epsilon_i > 0$.

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 reach a decision 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 comprised all possible pairwise combinations in 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 maximum gain and in which the value of maximum gain was shown on top of the shaded area. Because the probabilities were not displayed in numerical form, it was not possible for decision makers to calculate the exact expected values. Tversky (1969) predicted that for the "adjacent pairs", that is, for pairs (a, b), (b, c), (c, d), and (d, e), decision makers would prefer gambles with higher maximum gains, because the probabilities of maximum gains were visually very similar. In other words, the differences in the probabilities of maximum gain 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 maximum gain, because the difference in the probabilities would be large enough to exceed the corresponding threshold and the decision maker would determine his preference before 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 the probabilities of maximum gain and then the maximum gain for the ten gamble pairs in Tversky's stimulus set. Suppose that he uses a linear utility function for all the attribute values, such as for example, $u(r) = r$, where $r \in \mathbb{R}$, that he uses $\frac{3.5}{24}$ as the threshold for the probability of maximum gain for all pairs. Panel B in Table 1.1 shows the differences of probabilities of maximum gain in all ten pairs from Tversky (1969). It shows that the decision maker prefers e to a for pair (a,e) based on the probability of maximum gain, because the probability difference is $\frac{4}{24}$, larger than the threshold. For the remaining pairs, he does not have a preference, moves on to the next attribute, the maximum gain, and uses \$0.35 as the threshold for all ten pairs. Panel C in Table 1.1 shows the maximum gain difference in each pair. It shows that for pairs (a, c), (a, d), (b, d), (b, e), and (c, e), the differences between the maximum gains exceed \$0.35; therefore, he prefers the gambles with higher maximum gains

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, 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 maximum gain before considering the maximum gain, uses a probability threshold of $\frac{3.5}{24}$, and a maximum gain threshold of \$.3. 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.

Ever since the lexicographic semiorder model was proposed by Tversky (1969), the lexicographic heuristic has been discussed in many studies of decision making across a number of different fields: in psychology (Birnbaum, 2010, Tversky, 1972), in marketing science (Kohli and Jedidi, 2007, Yee et al., 2007), and in economics (Fishburn, 1980, Luce, 1978, Manzini and Mariotti, 2007, 2012). For example, in the area of consumer research, Yee et al. (2007) reported in their study that about two-thirds of the participants used the lexicographic rules for evaluating smart phones. Kohli and Jedidi (2007) studied whether people used the lexicographic rules on personal computer references and reported that the lexicographic rules were widely used by consumers. Both studies used the aggregated data among a group of participants. This is susceptible to aggregation artifacts.

Birnbaum (2010) described three implications of the family of lexicographic semiorder models as follows: *priority dominance* is the property that when a person prefers a choice option based on a dimension with priority, variations of other attributes should never reverse that preference; *attribute integration* is the property that when two changes in attributes, independently, are too small to reverse a preference, combining them cannot reverse a preference; *attribute interaction* is the property that when an attribute is the same in both gambles, changing its value should not change the preference between the two gambles. Birnbaum tested these three properties, as well as transitivity, with four different experiments using a true and error model. He reported violations of priority dominance, violations of attribute integration, violations of attribute interaction, and acceptance of transitivity. Therefore, Birnbaum concluded that the family of lexicographic semiorders does a poor job of describing how people make decisions. Birnbaum used a true and error model, results of which may vary with different artificial blockings. Thus, the conclusions in Birnbaum (2010) are questionable.

Davis-Stober (2012) described the convex polytope of simple lexicographic semiorder models (i.e., the convex hull of all simple lexicographic semiorders). He demonstrated that this polytope is equivalent to a 'mixture model' of probabilistic choices, which greatly constrains the set of permissible ternary choice

probabilities. This paper focuses on the mathematical properties of the mixture model of lexicographic semiorders; it does not perform any statistical analysis on the mixture model of lexicographic semiorders.

Regenwetter et al. (2011b) tested the mixture model of the lexicographic semiorder models using individual data sets, which are two-alternative forced choices, from Tversky (1969) and Regenwetter et al. (2011a). They found that the lexicographic semiorder model was rejected by about half of the participants. They suspected model mimicry between the lexicographic semiorders and linear orders.

Lexicographic semiorder models are purely algebraic. Tversky (1969), Birnbaum (2010), and Regenwetter et al. (2011b) all recast lexicographic semiorder models as probabilistic models and tested the resulting probabilistic models using different statistical methods. Tversky (1969) tested weak stochastic transitivity (Luce and Suppes, 1965) with likelihood ratio statistical tests (Mood, 1950). Birnbaum (2010) used a true and error model (Birnbaum, 2004, Birnbaum and Gutierrez, 2007) Regenwetter et al. (2011b) specified a mixture model and used order-constrained, likelihood-based statistical methods.

In my dissertation, I plan to test lexicographic semiorder models using rigorous statistical tests on individual data. I will specify distance-based and mixture-based probabilistic models of lexicographic semiorders, and test them with order-constrained, likelihood-based statistical tests using both the frequentist and Bayesian methods (Davis-Stober et al., 2015, Regenwetter and Davis-Stober, 2012, Regenwetter et al., 2014).

1.5.2 Similarity Model

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 imposes a lexicographic order on gamble attributes, a similarity model assumes that the decision maker considers all attributes simultaneously.

Rubinstein (1988) defines two types of similarity, the ϵ-difference similarity and λ-ratio similarity. Suppose that $\epsilon > 0$ is the threshold value. For any m, $n \in R$, Rubinstein defined the difference similarity by $m \sim n$ if $|m - n| \leq \epsilon$, and the ratio similarity by $m \sim n$ if $1/\lambda \leq m/n \leq \lambda$. 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 maximum gains or the probabilities of maximum gains.

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).

Extending Rubinstein's model, Leland (1994) proposed a new decision process for a more generalized case (i.e., gambles with more than two outcomes). He assumed that the decision maker employs the following three-step decision procedure. Suppose a decision maker has to choose between two gambles x and y. The expected utility value of a gamble z is written as $EU(z)$, and the threshold value for the expected utility value is written as ϵ_{EU}.

1. If $|EU(x) - EU(y)| > \epsilon_{EU}$, then the decision maker prefers the gamble with higher expected utility value. Otherwise, the decision maker goes to Step 2.

2. Compare gains and probabilities in terms of their equality and inequality. The decision maker compares each pair of gains and their corresponding probabilities. He decides one or more of the following: (a) whether each comparison of gains and their corresponding probabilities "favor" one gamble over the other (e.g., one gamble has a larger gain at a higher probability for a given comparison); (b) whether the comparison is "inconclusive" (e.g., one gamble has a larger gain but a lower probability); and (c) whether the comparison is "inconsequential" (e.g., the two gambles have identical gains and probabilities). After considering all pairwise comparisons for the two gambles, the decision maker prefers the gamble that is favored in one or more comparisons and inconsequential in the rest. Otherwise, he proceeds to Step 3.

3. Compare gains and probabilities in terms of their similarity and dissimilarity. Again, the decision maker repeats the set of comparisons in Step 2 in terms of similarity/dissimilarity. He prefers the gamble when it is favored in one or more paired comparisons (e.g., when one gamble has a higher and dissimilar gain at a similar probability with the other gamble) and inconclusive or inconsequential in the remaining comparisons. Otherwise, the decision maker does not form a preference and, instead, chooses at random.

Leland (1994) reported that the similarity model specified above violates transitivity of preferences. He compared the similarity model to Regret Theory, and concluded that both models are good alternatives to expected utility theory. Leland (2002) used a similarity model on intertemporal choice and concluded that the similarity model was able to explain some violations of the standard discounting utility model in the intertemporal study. However, both papers used pattern counting as the data analysis method, which is problematic.

Vilà (1998) generalized the similarity model by Rubinstein (1988) and applied it to alternatives with three attributes. Vilà used λ-ratio similarity and studied the mathematical properties of the model and

reported that this similarity model could predict intransitivity for alternatives with three attributes. This paper is purely theoretical and provides no empirical evidence.

Buschena and Zilberman (1999) studied the effects of similarity on risky choices by fitting two probit regression models of the observed choices between risky pairs on each of the perceived and objective similarity rankings. They used ϵ-difference similarity and reported that the more similar a gamble pair is, the more likely a decision maker will make a riskier choice; the more dissimilar a gamble pair is, the more likely he will make a safer choice.

Lorentziadis (2013) extended Rubinstein's similarity model and proposed a model named *the indistinguishable probability model*. In that model, for gamble probabilities, a decision maker has a clearly defined partition $\varsigma = (I_j, j = 1, \ldots, J)$ of $[0, 1]$, where the I_j are disjoint intervals of the form $[f_{j-1}, f_j)$ with $0 \leq f_{j-1} \leq f_j \leq 1$ and $I_J = 1$. The probabilities belonging to the same interval of partition ς are indistinguishable (i.e., within each given interval, the decision maker is unable to discern the differences in the probabilities and treat them as equal). Lorentziadis (2013) defined a representative point via a transformation to represent indistinguishable probabilities in a unique manner:

$$B(p) = \begin{cases} 0, & \text{for } p \text{ in } I_1, \\ f_{j-1}, & \text{for } p \text{ in } I_j, \\ 1, & \text{for } p \text{ in } I_J. \end{cases}$$

The model assumes that the decision maker views the probability in each probability interval as the minimum of the probability interval he employs (i.e. he transforms the probability to the lower endpoint of the interval to which the probability belongs). For example, for gamble $v = (v_1, p_1; \ldots; v_n, p_n)$, the decision maker views it as $\Big(v_1, B(p_1); \ldots; v_n, B(p_n)\Big)$, if he uses the indistinguishable probability model. Then the decision maker computes an expected utility value for the gamble after the transformation of the probabilities. In sum, Lorentziadis (2013) used both the ϵ-difference similarity and λ-ratio similarity, but did not provide any empirical tests of the indistinguishable probability model.

In my dissertation, for lack of empirical tests of similarity models, I plan to test similarity models against laboratory individual data with order-constrained, likelihood-based statistical tests (Regenwetter et al., 2014). I use the two types of similarity defined by Rubinstein (1988): difference similarity and ratio similarity. The similarity models I will test work as follows: the decision maker picks a threshold for each attribute/dimension of a gamble pair and forms a preference for that attribute. The decision maker derives his final preferences from integrating all the preferences on all the 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 satisfies a similarity model with difference similarity. He uses $\frac{3.5}{24}$ as the threshold of the probability of maximum gain, and \$.35 as the threshold of the maximum gain for all ten pairs. He forms preferences for the ten gamble pairs (all pairwise combinations of the five gambles in Tversky's stimulus set) in terms of both the probabilities of maximum gains and the maximum gains, as shown in the top two tables of Panel E in Table 1.1. When considering the probabilities of maximum gains, he prefers e over a, and he is indifferent about the remaining pairs. When considering the maximum gains, he is indifferent about the adjacent pairs, and prefers gambles with higher maximum gains for the other pairs. After integrating his preferences on both attributes, the decision maker derives his final preference, which is 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 maximum gains. For pair (a, e), the decision maker prefers e to a ($e \succ a$) based on the probabilities of maximum gains and prefers a to e ($a \succ e$) based on the maximum gains. Thus, after integrating preferences on 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 an intransitive preference.

I plan to test the similarity models specified above using a rigorous quantitative framework for testing theories of binary choice proposed by Regenwetter et al. (2014). I consider two different probabilistic specifications, including the distance-based and mixture-based probabilistic specifications. I will test these probabilistic specifications against laboratory data at the individual level, with order-constrained statistical inferences that use both frequentist and Bayesian methods (Davis-Stober, 2009, Myung et al., 2005, Regenwetter et al., 2014, Silvapulle and Sen, 2005). The stimulus sets I plan to use are from Experiment I in Tversky (1969), Cash I and Cash II in Regenwetter et al. (2011a), and an experiment that I conducted in 2012.

1.5.3 Priority Heuristic

Brandstätter et al. (2006) proposed a theory of decision making under risk, *the priority heuristic*, which models decision makers' cognitive processes and predicts intransitivity. The priority heuristic consists of

three parts. The first part is *the priority rule*. A decision maker considers first the minimum gains, then the probabilities of minimum gains, and lastly, the maximum gains. The second part is *the stopping rule*. The decision maker stops to make a decision when the minimum gains differ by 1/10 (or more) of the maximum gains, or when the probabilities of minimum gains differ by 1/10 (or more) of the probability scale. The third part is *the decision rule*. The decision maker prefers the gamble with the more attractive gains or probabilities.

The priority heuristic was motivated by the stimulus set in Tversky (1969) to explain alleged intransitive choices in that experiment. The priority heuristic is similar to a lexicographic semiorder model in that both models impose a lexicographic decision rule on the gamble attributes and have thresholds associated with all of the gamble attributes. There are two differences between a lexicographic semiorder model and the priority heuristic. The first difference is that the priority heuristic imposes a specific lexicographic rule on the gamble attributes, whereas a lexicographic semiorder model allows all possible lexicographic rules. The second difference is that the priority heuristic specifies the thresholds as 1/10 of the maximum values of the two alternatives for a given attribute, whereas a lexicographic semiorder model allows the thresholds to be any value.

Suppose the decision maker uses the priority heuristic to form preferences on the ten gamble pairs in Tversky's (1969) stimulus set. He first considers the minimum gains, which are all zeros. Then he moves to the probabilities of minimum gains with the threshold value 0.1, or $\frac{2.4}{24}$. The differences between the probabilities of the minimum gains for pairs (a, d), (a, e), and (b, e) exceed $\frac{2.4}{24}$. Thus, the decision maker prefers the gambles with lower probabilities of minimum gains in these three pairs, $a \prec d$, $a \prec e$, and $b \prec e$. For the remaining pairs, he proceeds to the next attribute, the maximum gains. He prefers the gamble with higher maximum gains. Now, the decision maker has derived the one and only preference pattern predicted by the priority heuristic on the ten gamble pairs from Tversky (1969), which is, $a \succ b$, $a \succ c$, $a \prec d$, $a \prec e$, $b \succ c$, $b \succ d$, $b \prec e$, $c \succ d$, $c \succ e$, and $d \succ e$. Here, $a \succ b$, $b \succ d$, and $a \prec d$ form an intransitive preference.

Brandstätter et al. (2006) reported that the priority heuristic could account for some violations of rationality (e.g., violations of transitivity) that expected utility theory cannot explain, and concluded that the priority heuristic has descriptive accuracy. Brandstätter et al. (2006) also reported that the priority heuristic outperformed other simple heuristics, such as minimax, maximal, tallying, and most-likely, in that the priority heuristic correctly predicted the participants' modal choices most of the time. In other words, the priority heuristic correctly predicted which choice the majority of participants preferred more often than the other heuristics. However, using modal choice as an indicator for model performance is susceptible to aggregation artifacts, and as stated in Section 1.2, it does not tell anything about individual performance.

Glöckner and Betsch (2008) tested the priority heuristic and Cumulative Prospect Theory (CPT) empirically by analyzing individual choice patterns, decision times, and information search parameters in diagnostic decision tasks. They used a Bayesian strategy classification method to classify individual choice patterns to see which theories could explain the choice patterns better, the priority heuristic or CPT. Glöckner and Betsch (2008) reported that the data are better explained by CPT than the priority heuristic. They suggested that the decision tasks in the previous study by Brandstätter et al. (2006), which supported the priority heuristic over CPT might not have provided the diagnostic tasks to compare the priority heuristic with CPT. Glöckner and Herbold (2011) also studied the priority heuristic, CPT, and decision field theory using eye-tracking methods. They concluded that individuals used compensatory strategies (e.g., CPT) rather than a non-compensatory heuristic model (e.g., the priority heuristic).

Johnson et al. (2008) used a computer-based information-search tracing tool, MouselabWeb, to trace how participants searched for information when presented with a gamble pair on a computer screen. Johnson et al. have specific expectations about how different model types relate to tracing patterns: according to a process decision model, such as the priority heuristic, the decision maker compares the gains and probabilities between the two gambles; that is, he searches for information across gambles. According to an integrative model, such as CPT, the decision maker looks for information within gambles. Johnson et al. reported that participants mostly searched for information within gambles, and compared information between gambles less often. Johnson et al. (2008) concluded that this result serves as evidence supporting integrative decision models, rather than the priority heuristic.

Messner and Regenwetter (2009) submitted the priority heuristic for testing on 267 different individual data sets, of which 22 data sets were cherry-picked for generating intransitive choices. As I mentioned earlier, the priority heuristic is motivated to account for data from Tversky (1969). All the 22 individual data sets are either the original data from Tversky (1969) or from studies using the gambles in Tversky (1969). Messner and Regenwetter (2009) used a probabilistic specification to model variability in the participants' choices, the modal choice probabilistic specification, which assumes that a decision maker has a fixed preference and makes errors up to 50% of the time. They tested the modal choice probabilistic specification of the priority heuristics using order-constrained inference statistics (Davis-Stober, 2009) at individual level. The modal choice model is a very lenient model, but the priority heuristic reportedly accounted for 10 of the 22 cherry-picked participants (about 45%) and only 20 of the 245 non-cherry-picked participants (about 8%). The priority heuristic does a fair job to explain the data that it is designed for (i.e., to explain data using gambles from Tversky, 1969), and does a poor job to explain the other data. The lack of consistency to fit the data with the priority heuristic suggests that the priority heuristic is not a good model to describe

people's decision processes.

Birnbaum and Bahra (2007), Birnbaum (2008a), Birnbaum and LaCroix (2008), and Birnbaum (2010) reported evidence against the priority heuristic. Again, all four papers used a true and error model as the statistical model, which uses an artificial "blocking" assumption.

Rieger and Wang (2008) tested the validity of the priority heuristic on the empirical data from Tversky and Kahneman (1992a) by computing the average deviation between the estimated certainty equivalent and the measured one. They showed that the priority heuristic did not perform well at all on these data. Rieskamp (2008) provided a probabilistic generalization for the priority heuristic, CPT, and decision field theory; he added an error term in each gamble pair to explain the participants' choice variabilities; and then he tested these models by using a maximum likelihood ratio test. He concluded that the probabilistic versions for all three models were always better than the deterministic versions in describing decisions under risk. When comparing all three probabilistic models, decision field theory did the best and the priority heuristic did the worst. However, both papers used pooled data rather than individual data in their analysis. Hence, any conclusions from these two papers do not inform us of individuals' performances.

Arló-Costa and Pedersen (2013) proposed a modification of the priority heuristic by extending it from choices under risk to uncertainty. The central change was to use upper and lower probabilities with uncertainty. The priority rule was modified and assumed that a decision maker goes through a reasoning process in the following order: minimum gain, upper probability of minimum gain, maximum gain, and lower probability of maximum gain. The stopping rule and decision rule remained the same. The authors did not provide any empirical tes of the modified priority heuristic; instead, the study focused on the descriptive accuracy of the new model in explaining, for example, the Allais paradox (Allais, 1953) and the Ellsberg paradox (Ellsberg, 1961).

Brandstätter and Gussmack (2013) tested expected utility theory and the priority heuristic by applying a new process-tracing method: the predict-aloud protocols. They concluded that when a task was difficult, the decision maker used the priority heuristic the most, but when the task was easy, he tended to use the similarity rule instead. The study also revealed that the decision maker tended to compare between gambles rather than within gambles—which is opposite of what Johnson et al. (2008) reported.

Pachur et al. (2013) used Mouselab to measure the direction of the participants' information searching (i.e., within gambles or between gambles) and acquisition frequencies (i.e., how frequently individual reasons are looked up). Pachur et al. compared the priority heuristic with expectation models (e.g., expected value theory, expected utility theory, and prospect theory). They reported that the priority heuristic predicted the direction of information searching better than the expectation model. However, only the similarity rule

could account for the number of times that the participants attempted to acquire information. Pachur et al. also concluded that participants were more likely to use the priority heuristic when the task was difficult, which is consistent with what Brandstätter and Gussmack (2013) reported. Brandstätter and Gussmack (2013), Johnson et al. (2008), and Pachur et al. (2013) did not provide any statistical tests for how well the priority heuristic explains the participants' choices.

The priority heuristic is a deterministic heuristic theory. Many papers have specified probabilistic specifications of the priority heuristic and tested the probabilistic specifications using some statistical tests, including Birnbaum (2008a), Birnbaum (2010), Birnbaum and Bahra (2007), Birnbaum and LaCroix (2008), Glöckner and Betsch (2008), Messner and Regenwetter (2009), and Rieskamp (2008). Again, most studies suffer from some methodological problems.

I plan to test the priority heuristic using the statistical methods discussed in Regenwetter et al. (2014) on the following stimulus sets: gambles from Experiment I in Tversky (1969), Cash I and Cash II in Regenwetter et al. (2011a), and an experiment I conducted in 2012.

1.6 Perceived Relative Argument Model (PRAM)

Loomes (2010, Psychological Review) developed a descriptive model of individual decision making under risk, the *Perceived Relative Argument Model* (PRAM). PRAM describes how decision makers choose among lotteries in which one can win various gains with various probabilities. According to PRAM, the decision maker compares the perceived argument favoring one lottery based on probabilities with the perceived argument favoring the other lottery based on gains. She prefers one lottery over the other based on the perceived relative argument in its favor. PRAM violates several key axioms of rational behavior, e.g., independence, betweenness, and transitivity. PRAM applies to a specific domain. It models pairwise preference among two lotteries S, R of a particular form. Writing x_i for the ith monetary outcome and p_i, q_i for the probability of the ith monetary outcome in S and R, respectively, the 'safer' lottery S and the 'riskier' lottery R must satisfy the following properties:

$$\begin{array}{lll} S = (x_3, p_3; x_2, p_2; x_1, p_1) & & x_3 > x_2 > x_1 \geq 0; \\ R = (x_3, q_3; x_2, q_2; x_1, q_1) & \text{with} & q_3 > p_3; \;\; q_2 < p_2; \;\; q_1 > p_1. \end{array} \tag{1.7}$$

According to PRAM, a decision maker faced with the choice between S and R evaluates the relative argument in favor of S in terms of probabilities using a *perception function for probabilities*, $\phi(b_S, b_R)$, via:

$$\phi(b_S, b_R) = \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha}, \qquad \text{where} \qquad b_S = q_1 - p_1; b_R = q_3 - p_3. \tag{1.8}$$

This function has a real-valued free parameter α. The decision maker also relies on a *perception function*

for outcomes, $\xi(y_R, y_S)$, to determine the relative argument in favor of R regarding the outcomes:

$$\xi(y_R, y_S) = \left(\frac{c_S}{c_R}\right)^{\delta}, \qquad \text{where} \qquad c_S = c(x_3) - c(x_2); c_R = c(x_2) - c(x_1). \tag{1.9}$$

Here, $\delta \geq 1$ is a free parameter and $c(x)$ is the utility for money.

Let $S \prec R$ denote that R is preferred to S, $S \sim R$ denote that a person is indifferent between S and R, and $S \succ R$ denote that S is preferred to R. PRAM compares the perception of probabilities with the perception of gains and makes the following predictions:

$$\left\langle \quad S \quad \begin{matrix} \prec \\ \sim \\ \succ \end{matrix} \quad R \quad \right\rangle \quad \Longleftrightarrow \quad \left\langle \quad \phi(b_S, b_R) \quad \begin{matrix} < \\ = \\ > \end{matrix} \quad \xi(y_R, y_S) \quad \right\rangle. \tag{1.10}$$

PRAM is similar to the additive difference model in that they both compare gambles in terms of the gamble attributes. In other words, they all use between-gamble comparisons. The differences are specified as follows: PRAM specifies different functions for the gains and the probabilities of gains, whereas the additive difference model uses the same function for the gains and the probabilities of gains.

Loomes (2010b) also mentioned a more general predecessor to PRAM, which was reported in Loomes (2006). This model, referred to here as PRAM 2006, featured one more person-specific parameter γ in the function ω of Eq. 1.8, in addition to α. Letting $\gamma \geq 0$, PRAM 2006 assumes that

$$\omega = \left[\left(1 - \frac{p_1}{q_1}\right)\left(1 - \frac{q_2}{p_2}\right)\left(1 - \frac{p_3}{q_3}\right)\right]^{\gamma} \left(\frac{q_1 - p_1}{q_3 - p_3}\right)^{(q_1 - p_1 + q_3 - p_3)^{\alpha}}.$$

Loomes (2010b) used descriptive, across-participants, modal choice ("what did most people choose on this pair?") to provide qualitative evidence in support of PRAM's ability to explain data. There are major shortcomings to both, descriptive methods, and modal choice analyses across participants (see, e.g., Regenwetter et al., 2014, for a recent discussion). Loomes (2010b) used empirical illustrations to suggest that PRAM was able to predict the choice tendency of the majority of the participants. Thus, Loomes did not provide quantitative statistical tests of PRAM.

Buschena and Atwood (2011) specified a probabilistic specification of PRAM 2006 by adding an error term for the choice probability of each gamble pair and assuming that the error is a random draw from a logistic distribution (Hey and Orme, 1994), and tested that probabilistic specification of PRAM 2006 using a likelihood ratio test and Bayesian information criteria, and reported that PRAM 2006 was not supported by their data.

Baillon et al. (2015) analyzed their data against PRAM using a true and error model. They reported that the data did not support the existence of the regret cycle, which both PRAM and Regret Theory predicted

for their experimental stimuli. However, Baillon et al. used a different gamble format from what Loomes (2010) specified. The gambles in Baillon et al. (2015) had the same probabilities of gains and different gains for a gamble pair; yet Loomes (2010) required different probabilities of gains and the same gains for a gamble pair.

Guo and Regenwetter (2014) reported a new experiment and an individual-level analysis using both frequentist and Bayesian order-constrained statistical inference. In their experiment, 67 people participated in the first session, and 54 returned for a second, identical session the next day. The second session served as a replication of the first. There were 20 gamble pairs, which were designed to test PRAM in the experiment, and each pair was repeated 20 times per session. Guo et al. considered two types of "stochastic specification" of PRAM. In one, a decision maker has a fixed preference and choice variability is caused by occasional errors/trembles. In the other, the parameters of the perception functions for outcomes and for probabilities are random, with no constraints on their joint distribution. The results suggested that PRAM accounted poorly for individual subject laboratory data of nearly all 67 participants.

PRAM is an algebraic model. Buschena and Atwood (2011), Baillon et al. (2015), and Guo and Regenwetter (2014) all provided probabilistic specifications of PRAM and tested the probabilistic specifications of PRAM using statistical methods. All three papers concluded that PRAM did a poor job in explaining the participants' choices.

1.7 Conclusions

There are many transitive theories for decision making under risk and uncertainty, but only a few intransitive theories in the literature. I have reviewed six major intransitive decision theories in this paper: the additive difference model, Regret Theory, the lexicographic semiorder models, the similarity models, the priority heuristic and the perceived relative argument model. Intransitive theories attempt to explain reported behavioral violations of expected utility theory or prospect theory. However, for many intransitive theories, there are either no data collected or no data analyzed. For some intransitive theories, there are some data analyses but those data analyses usually have some methodological problems. The empirical evidence for and against all of these intransitive theories is weak. Moreover, there are pervasive methodological problems in the behavioral studies. Therefore, violation of transitivity is not very well documented. The entire literature on intransitivity of preferences should be reconsidered.

Therefore, the question is how to test these transitive and intransitive decision theories using a systematic and rigorous approach. The answer is to collect appropriate data (repeated choices from each

individual participant) and perform rigorous statistical analyses. Regenwetter et al. (2014) provided a novel and rigorous quantitative diagnostic framework for testing theories of binary choice. Their theoretical framework provided a solution to Luce's two-fold challenge and linked deterministic algebraic decision theory with observed variability in behavioral binary choice data. Regenwetter and Davis-Stober (2012) tested two transitive heuristic models, the weak order model and the linear order model, on individual data using order-constrained statistical methods. They have concluded that transitive theories account for decision makers' choices very well. I plan to test several major intransitive decision theories reviewed in this paper: the lexicographic semiorder models, the similairy models, the priority heuristic, and the perceived relative argument model. My results will help us understand to what extent these intransitive theories are adding value to behavioral decision research. For now, based on the lack of data for some intransitive decision theories and the poor methodology employed to test intransitive theories, we still do not know how well intransitive decision theories account for human behavior.

Chapter 2

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

2.1 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), and Cumulative Prospect Theory (Tversky and Kahneman, 1992b). 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., Brandstätter et al., 2006, González-Vallejo, 2002, Loomes and Sugden, 1987, Tversky, 1969). 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 (2018b) for details on these methodological problems). Thus, there is still little evidence of intransitivity (Davis-Stober et al., 2015, Regenwetter et al., 2011a, Regenwetter and Davis-Stober, 2012). 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 2.2 describes two intransitive decision heuristics: lexicographic semiorder models and similarity models; Section 2.3 describes the transitive linear order model and introduces two simple transitive heuristics; Section 2.4 introduces two kinds of probabilistic specifications for the algebraic models: distance-based models and mixture models. It also describes the statistical tools; Section 2.5 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 2.6 reports the data analysis results and Section 2.7 concludes the paper.

2.2 Intransitive Heuristic Models

In this section, I describe two intransitive heuristics, including the lexicographic semiorder model (Tversky, 1969) and the similarity model (Leland, 1994, Rubinstein, 1988). These two intransitive heuristics are illustrated using Tversky's (1969) stimulus set (see Panel A of Table 2.1). 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.

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

Table 2.1: 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	-	$\sim$	$\succ$	$\succ$	$\succ$
b		-	$\succsim$	$\succ$	$\succ$
c			-	$\succsim$	$\succ$
d				-	$\succsim$

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	-	$\sim$	$\succsim$	$\succsim$	$\succ$
b		-	$\succsim$	$\succsim$	$\succsim$
c			-	$\succsim$	$\succsim$
d				-	$\succsim$

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

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

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.

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 2.1 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 2.1 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 2.1, 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).

2.2.2 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 2.1. 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 2.1. 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 (2018b).

2.3 Transitive Models

2.3.1 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. (2017, 2011a,b) tested linear order models on risky and intertemporal data, and reported that the linear order model could explain the participants' behavior very well.

2.3.2 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.

2.4 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 2.4.1 and 2.4.2 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 (Davis-Stober, 2009, Iverson and Falmagne, 1985, Silvapulle and Sen, 2005). 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.

2.4.1 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 2.1. 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.

2.4.2 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 2.1, 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 2.2), as I vary the sequence of attributes and the threshold values. Row 16 in Table 2.2 shows the preference pattern that is depicted on the left side of Panel D in Table 2.1.

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

Table 2.2: 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	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$

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:

$$\theta_{ab} = \theta_{bc} = \theta_{cd} = \theta_{de}, \tag{2.1}$$

$$\theta_{ac} = \theta_{bd} = \theta_{ce}, \tag{2.2}$$

$$\theta_{ad} = \theta_{be}. \tag{2.3}$$

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

Inequalities:

$$0 \leq \theta_{ae}, \theta_{be}, \theta_{ce}, \theta_{de} \leq 1, \tag{2.4}$$

$$0 \leq \theta_{be} + \theta_{ce} - 2\theta_{de} \leq 2, \tag{2.5}$$

$$0 \leq \theta_{ae} + \theta_{ce} - 2\theta_{de} \leq 2, \tag{2.6}$$

$$0 \leq \theta_{ae} + \theta_{be} - 2\theta_{de} \leq 2, \tag{2.7}$$

$$0 \leq \theta_{ae} + \theta_{be} - 2\theta_{ce} \leq 2, \tag{2.8}$$

$$0 \leq -\theta_{ae} + 2\theta_{be} - 2\theta_{ce} + 2\theta_{de} \leq 2. \tag{2.9}$$

Equalities 2.1 to 2.3 show equal probabilities for certain gamble pairs. For example, Equality 2.1 shows equal probabilities for adjacent pairs in Tversky's stimuli. Equalities 2.1 to 2.3 show that this mixture polytope has four free parameters, θ_{ae}, θ_{be}, θ_{ce}, and θ_{de}, which are restricted by Inequalities 2.4 to 2.9. 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 parameters: 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.

2.4.3 Summary of Models

Table 2.3 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

Table 2.3: 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	-	-	-

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.

2.4.4 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

[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.

Table 2.4: 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

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.

2.5 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 2.1. 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.

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 2.4). Participants were 18 undergraduates at the University of Illinois at Urbana-Champaign. Gambles were presented as wheels of chance on computers, similar to Figure 2.1. 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 2.1), 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.

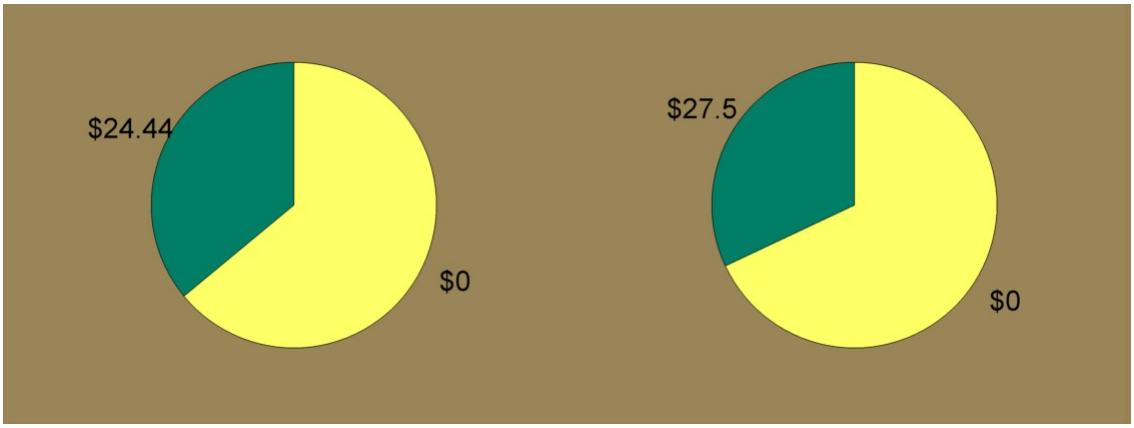

Figure 2.1: A gamble pair displayed in the experiment that I conducted in 2012.

2.6 Results

2.6.1 Distance-Based Model Results

Tables 2.5, 2.6, and 2.7 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 2.5, 2.6, and 2.7 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 2.5 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

Table 2.5: 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

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 2.5 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 2.6 and 2.7 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 2.6 and 2.7 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.

2.6.2 Mixture Model Results

Table 2.8 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

Table 2.6: 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 2.7: 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 2.8: 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	

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.

2.6.3 Model Comparison: Individual Level

I use Bayes factors to compare models. As I discuss in Section 2.4.4, 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 2.9 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

Table 2.9: 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

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 2.10: 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 2.10 – 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 2.10 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.

2.6.4 Model Comparison: Group Level

Table 2.11 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

Table 2.11: 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

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.

2.7 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 (2018b). 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.

2.8 One Published Article

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

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*, 119(2):408-416[3].

Please see Appendix A for the full published article.

2.9 Supplement Materials

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

Table 2.12: 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 2.13: 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 2.14: 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 2.15: 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 2.16: 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	$\surd$	$\surd$	0.09	0.15	$\star$	$\star$	0.97	$\surd$	$\star$	0.44	$\star$	$\star$
2	$\surd$	$\surd$	0.30	0.30	$\star$	$\star$	$\surd$	$\surd$	0.30	0.30	$\star$	$\star$
4	$\surd$	0.35	0.11	$\star$	$\star$	$\star$	$\surd$	0.46	0.42	$\star$	$\star$	$\star$
5	0.79	$\surd$	$\star$	0.49	$\star$	$\star$	0.81	$\surd$	$\star$	0.49	$\star$	$\star$
7	$\surd$	$\surd$	0.11	0.06	$\star$	$\star$	$\surd$	0.88	0.14	0.07	$\star$	$\star$
9	0.31	$\surd$	$\star$	$\surd$	$\star$	0.65	$\surd$	$\surd$	0.17	$\surd$	$\star$	0.65
11	$\surd$	$\surd$	0.08	0.59	$\star$	0.12	$\surd$	$\surd$	0.08	0.59	$\star$	0.12
12	$\surd$	$\star$	0.59	$\star$	$\star$	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	$\star$
13	$\surd$	0.50	0.28	$\star$	$\star$	$\star$	$\surd$	0.50	0.28	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.94	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.94
15	0.85	0.50	$\star$	0.11	$\star$	$\star$	0.87	$\surd$	$\star$	0.16	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	$\surd$	$\star$	0.15	$\surd$	$\surd$	$\star$	$\surd$	$\star$	0.15
17	0.57	0.83	$\star$	0.11	$\star$	$\star$	0.70	$\surd$	0.13	0.96	$\star$	0.09
18	0.52	$\surd$	0.07	0.56	$\star$	0.34	$\surd$	$\surd$	0.78	0.93	$\star$	0.34
19	$\surd$	$\surd$	0.79	0.52	0.67	0.29	$\surd$	$\surd$	0.79	0.52	0.67	0.29
20	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
21	0.71	$\star$	$\star$	$\star$	$\star$	$\star$	0.59	$\star$	$\star$	$\star$	$\star$	$\star$
22	0.91	0.90	0.37	0.06	0.08	$\star$	0.91	0.91	0.37	0.06	0.08	$\star$
23	0.83	0.68	$\star$	$\star$	$\star$	$\star$	0.83	0.68	$\star$	$\star$	$\star$	$\star$
24	$\surd$	$\surd$	$\surd$	$\surd$	0.51	0.97	$\surd$	$\surd$	$\surd$	$\surd$	0.51	0.97
25	0.70	0.96	0.06	0.29	$\star$	$\star$	$\surd$	$\surd$	0.73	0.55	0.21	0.13
26	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
27	$\surd$	$\surd$	0.38	0.38	$\star$	$\star$	$\surd$	$\surd$	0.38	0.45	$\star$	$\star$
28	$\surd$	$\surd$	0.66	0.76	0.67	0.19	$\surd$	$\surd$	0.66	0.76	0.67	0.19
29	0.71	$\surd$	0.16	0.25	$\star$	$\star$	0.71	$\surd$	0.16	0.25	$\star$	$\star$
30	$\surd$	$\surd$	0.32	0.60	0.18	0.23	$\surd$	$\surd$	0.32	0.60	0.18	0.23
31	$\star$	0.78	$\star$	0.18	$\star$	$\star$	$\star$	0.78	$\star$	0.28	$\star$	$\star$
32	$\surd$	$\surd$	$\surd$	$\surd$	0.94	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.94	$\surd$
33	$\star$	0.47	$\star$	$\star$	$\star$	$\star$	$\star$	0.23	$\star$	$\star$	$\star$	$\star$
34	0.98	$\surd$	0.10	0.77	$\star$	0.62	0.98	$\surd$	0.10	0.77	$\star$	0.62
35	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.09	$\surd$	$\surd$	$\surd$	$\surd$	0.36	0.09
36	$\surd$	$\surd$	$\surd$	0.31	0.62	0.12	$\surd$	$\surd$	$\surd$	0.31	0.62	0.12
37	$\surd$	0.90	$\surd$	0.07	0.98	$\star$	$\surd$	0.90	$\surd$	0.07	0.98	$\star$
38	$\surd$	$\surd$	0.36	0.17	$\star$	$\star$	$\surd$	$\surd$	0.36	0.73	$\star$	$\star$
39	$\surd$	0.81	0.28	0.38	$\star$	0.26	$\surd$	0.81	0.28	0.38	$\star$	0.26
41	0.55	$\surd$	0.09	$\star$	$\star$	$\star$	$\surd$	$\surd$	0.57	$\star$	$\star$	$\star$
42	$\surd$	$\surd$	$\star$	0.76	$\star$	$\star$	$\surd$	$\surd$	0.44	0.93	$\star$	0.29
43	0.71	0.71	$\star$	0.13	$\star$	$\star$	0.71	$\surd$	$\star$	0.16	$\star$	$\star$

Continued on next page

Table 2.16 – 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 2.17: 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	$\surd$	$\surd$	0.09	0.15	$\star$	$\star$	0.97	$\surd$	$\star$	0.44	$\star$	$\star$
2	$\surd$	$\surd$	0.43	0.30	$\star$	$\star$	$\surd$	$\surd$	0.50	0.30	$\star$	$\star$
4	$\surd$	0.38	0.11	$\star$	$\star$	$\star$	$\surd$	0.46	0.42	$\star$	$\star$	$\star$
5	0.19	$\surd$	$\star$	0.49	$\star$	$\star$	0.19	$\surd$	$\star$	0.49	$\star$	$\star$
7	$\surd$	$\surd$	0.48	0.22	$\star$	$\star$	$\surd$	0.88	0.48	0.27	$\star$	$\star$
9	$\surd$	$\surd$	0.17	$\surd$	$\star$	0.65	$\surd$	$\surd$	0.28	$\surd$	$\star$	0.65
11	$\surd$	$\surd$	0.08	0.59	$\star$	0.12	$\surd$	$\surd$	0.08	0.69	$\star$	0.12
12	$\surd$	$\star$	0.59	$\star$	$\star$	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	$\star$
13	$\surd$	0.50	0.28	$\star$	$\star$	$\star$	$\surd$	0.50	0.28	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.94	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.94
15	$\surd$	0.50	0.53	0.11	$\star$	$\star$	$\surd$	0.50	0.53	0.16	$\star$	$\star$
16	$\surd$	$\surd$	0.17	$\surd$	$\star$	0.15	$\surd$	$\surd$	0.32	$\surd$	$\star$	0.15
17	0.57	0.83	$\star$	0.11	$\star$	$\star$	0.65	$\surd$	0.13	0.8	$\star$	0.09
18	0.52	$\surd$	0.07	0.56	$\star$	0.34	$\surd$	$\surd$	0.78	0.93	$\star$	0.34
19	$\surd$	$\surd$	0.79	0.52	0.67	0.29	$\surd$	$\surd$	0.79	0.52	0.67	0.29
20	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
21	0.21	$\star$	$\star$	$\star$	$\star$	$\star$	0.18	$\star$	$\star$	$\star$	$\star$	$\star$
22	0.85	0.90	0.37	0.08	0.08	$\star$	$\surd$	0.90	0.37	0.14	0.08	$\star$
23	$\surd$	0.68	$\star$	$\star$	$\star$	$\star$	$\surd$	0.68	$\star$	$\star$	$\star$	$\star$
24	$\surd$	$\surd$	$\surd$	$\surd$	0.51	0.97	$\surd$	$\surd$	$\surd$	$\surd$	0.51	0.97
25	0.70	0.96	0.06	0.29	$\star$	$\star$	$\surd$	$\surd$	0.73	0.55	0.21	0.13
26	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
27	$\surd$	$\surd$	0.38	0.45	$\star$	$\star$	$\surd$	$\surd$	0.38	0.6	$\star$	$\star$
28	$\surd$	$\surd$	0.66	0.76	0.67	0.19	$\surd$	$\surd$	0.66	0.76	0.67	0.19
29	0.80	$\surd$	0.16	0.25	$\star$	$\star$	$\surd$	$\surd$	0.34	0.27	0.1	$\star$
30	$\surd$	$\surd$	0.32	0.60	0.18	0.23	$\surd$	$\surd$	0.32	0.60	0.18	0.23
31	0.08	0.85	$\star$	0.13	$\star$	$\star$	0.06	0.85	$\star$	0.13	$\star$	$\star$
32	$\surd$	$\surd$	$\surd$	$\surd$	0.94	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	0.94	$\surd$
33	0.12	0.47	$\star$	$\star$	$\star$	$\star$	0.13	0.23	$\star$	$\star$	$\star$	$\star$
34	$\surd$	$\surd$	0.20	0.77	$\star$	0.62	$\surd$	$\surd$	0.20	0.77	$\star$	0.62
35	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.09	$\surd$	$\surd$	$\surd$	$\surd$	0.36	0.09
36	$\surd$	$\surd$	$\surd$	0.31	0.62	0.12	$\surd$	$\surd$	$\surd$	0.31	0.62	0.12
37	$\surd$	$\surd$	$\surd$	0.19	0.98	$\star$	$\surd$	$\surd$	$\surd$	0.37	0.98	$\star$
38	$\surd$	$\surd$	0.36	0.16	$\star$	$\star$	$\surd$	$\surd$	0.36	0.68	$\star$	$\star$
39	$\surd$	0.83	0.28	0.38	$\star$	0.26	$\surd$	$\surd$	0.29	0.38	$\star$	0.26
41	0.55	$\surd$	0.09	$\star$	$\star$	$\star$	$\surd$	$\surd$	0.57	$\star$	$\star$	$\star$
42	$\surd$	$\surd$	$\star$	0.76	$\star$	$\star$	$\surd$	$\surd$	0.44	0.93	$\star$	0.29
43	0.71	0.71	$\star$	0.22	$\star$	$\star$	0.71	0.71	$\star$	0.22	$\star$	$\star$

Continued on next page

Table 2.17 – 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	√	0.34	0.25	⋆	⋆	0.53	√	0.34	0.77	⋆	0.29
46	√	√	0.07	√	⋆	0.09	√	√	⋆	√	⋆	0.09
47	√	√	√	0.39	0.98	⋆	√	√	√	0.39	0.98	⋆
48	√	0.87	0.36	0.19	⋆	0.12	√	0.87	0.36	0.19	⋆	0.12
49	0.79	√	⋆	0.6	⋆	0.24	0.80	√	⋆	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	√	√	0.40	0.50	⋆	0.15	√	√	0.68	0.50	⋆	0.15
55	⋆	√	⋆	√	⋆	√	⋆	√	⋆	√	⋆	√
56	√	√	0.92	0.79	0.34	0.59	√	√	0.93	0.98	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	√	√	0.83	√	⋆	√	√	√	0.83	√	⋆	√
66	√	√	√	0.43	√	0.1	√	√	√	0.89	√	0.28
67	√	√	√	√	0.98	0.99	√	√	√	√	0.98	0.99
3	√		0.17		⋆		√		0.2		⋆	
6	√		√		0.85		√		√		0.85	
8	√		0.6		⋆		√		0.6		⋆	
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	√		0.6		⋆		√		0.73		⋆	
57	√		0.65		0.69		√		0.73		0.69	
60	√		0.23		0.75		√		0.23		0.75	
62	√		0.13		⋆		√		0.13		⋆	
63	√		√		0.97		√		√		0.97	
64	0.88		0.28		⋆		0.88		0.19		⋆	
Fits	66	52	59	48	24	30	66	52	59	48	27	34

Table 2.18: 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 2.18 – continued from previous page

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

Table 2.19: The Bayes factors for the distance-based models for LSO-Diff 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-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	18891855275831140
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	21742296525682 43	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 2.20: 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 2.21: The Bayes factors for the distance-based models for SIM-Diff 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-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 2.22: 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 2.23: 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 2.24: 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 2.25: 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 2.26: 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 2.26 – 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 2.26 – 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 2.27: 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 2.27 – 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 3

Quantitative Tests of the Perceived Relative Argument Model Commentary on Loomes (2010)

3.1 Published Paper

A published 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 [1].

Please see Appendix B for the full published article.

3.2 Online Supplement Materials

This section reports the online supplement materials for the following 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.

It includes the analytical proofs for deterministic choice under PRAM. It reports participants' data and various tables of results.

Please see Appendix C for the full published supplement materials.

[1]

Chapter 4

Heterogeneity and Parsimony in Intertemporal Choice

A published paper:

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision*, 5(2), 63-94.

http://dx.doi.org/10.1037/dec0000069[1]

Please see Appendix D for the full published article.

Chapter 5

Testing 49 Different Forms of Cumulative Prospect Theory

5.1 Introduction

Tversky and Kahneman (1992b) proposed *Cumulative Prospect Theory*, henceforth CPT, to describe how people make decisions under risk. It is one of the most influential decision theories in the past few decades. Tversky and Kahneman (1992b) has been cited more than 4,700 times and CPT has been applied to many different contexts, for example, management (Becker and Gerhart, 1996, Steel and König, 2006), psychology (Lopes and Oden, 1999, Trepel et al., 2005), and transportation (Gao et al., 2010, Xu et al., 2011). The following describes how CPT works. For a gamble $G = (x_1, p_1; \ldots; x_n, p_n)$, where $x_1 \leq \ldots \leq x_k \leq 0 \leq x_{k+1} \leq \ldots \leq x_n$, let $w^+(p)$ and $w^-(p)$ be the probability weighting function to capture the subjective perception of the probabilities of gains and losses respectively. Let $u^+(x)$ and $u^-(x)$ be the utility function to capture the subjective perception of gains and losses respectively. CPT states that:

$$CPT(G) = \sum_{i=1}^{k} w_i^- u^-(x_i) + \sum_{i=k+1}^{n} w_i^+ u^+(x_i)$$

where

$$w_1^- = w^-(p_1), w_i^- = w^-(p_1 + ... + p_i) - w^-(p_1 + ... + p_{i-1}), \text{ for } 2 \leq i \leq k;$$

$$w_n^+ = w^+(p_n), w_i^+ = w^+(p_i + ... + p_n) - w^+(p_{i+1} + ... + p_n), \text{ for } k+1 \leq i \leq n-1.$$

In the current study, I only consider gambles with gains. I use $w(p)$ for the probability weighting function and $u(x)$ for the utility function. For example, for a two-outcome gamble with positive rewards, $G = (x_1, p_1; x_2, p_2)$, where $0 \leq x_1 \leq x_2$, CPT states that

$$CPT(G) = (1 - w(p_2))u(x_1) + w(p_2)u(x_2).$$

There has been a lot of work on fine-tuning the mathematical specifics of the theory (see, e.g., Stott, 2006, for a summary). The theory is algebraic, and hence inherently deterministic. Over time, there have been many different approaches to connect CPT to probabilistic data generating processes and, hence, to connect it to the statistical analysis of observations in the laboratory. Various papers, such as Stott (2006),

Blavatskyy and Pogrebna (2010), and Regenwetter et al. (2014) have highlighted that empirically evaluating CPT involves many moving parts: the outcome of such an analysis can strongly hinge on:

- the mathematical specification of the utility function for money
- the mathematical specification of the probability weighting function
- the probabilistic specification
- the stimuli used
- how observations are aggregated into data
- the statistical method

Stott (2006) investigated "Cumulative Prospect Theory's Functional Menagerie" by considering seven different functional forms for the utility function for gains, seven functional forms for the probability weighting function, and four probabilistic response mechanisms based on the assumption that the decision maker has a deterministic preference and that uncertainty in choice is due to noise/error. In this paper, I consider the same 49 combinations for functional forms on new stimuli, and with more general and more diverse non-parametric probabilistic specifications.

The rest of the paper is organized as follows: Section 5.2 describes two different stimulus sets used in this paper, from Experiment 2009 and Experiment 2012; Section 5.3 describes different functional forms for the probability weighting function and the utility function; Section 5.4 introduces two different probabilistic specifications and the relevant statistical methods; Section 5.5 reports the data analysis results; and Section 5.6 concludes the paper.

5.2 Experiments

I used two different stimulus sets in this paper, from Experiment 2009 and Experiment 2012. All gambles in Experiment 2009 have only two rewards, whereas gambles in Experiment 2012 have up to four rewards. Because the mathematical form of CPT is simple for two-outcome gambles, it is natural to start with two-outcome gambles when testing CPT. However, CPT makes richer and more restrictive predictions for more complicated gambles. Thus, it is important to test CPT with different kinds of stimuli.

Table 5.1: The 20 gamble pairs in Experiment 2009.

	Monetary gamble: Gamble 1				Monetary gamble: Gamble 0			
Pair	x_1	p_1	x_2	p_2	y_1	q_1	y_2	q_2
1	\$1.19	0.35	\$29.38	0.65	\$3.21	0.32	\$18.00	0.68
2	\$18.89	0.58	\$27.98	0.42	\$3.90	0.53	\$25.44	0.47
3	\$1.92	0.48	\$26.44	0.52	\$5.77	0.66	\$26.03	0.34
4	\$24.01	0.76	\$25.05	0.24	\$10.56	0.34	\$25.32	0.66
5	\$10.78	0.29	\$23.64	0.71	\$6.86	0.02	\$25.03	0.98
6	\$11.61	0.20	\$20.76	0.80	\$8.14	0.07	\$12.42	0.93
7	\$2.46	0.77	\$19.38	0.23	\$0.73	0.04	\$12.57	0.96
8	\$4.97	0.61	\$18.02	0.39	\$14.26	0.51	\$15.01	0.49
9	\$9.03	0.40	\$16.66	0.60	\$10.87	0.81	\$16.32	0.19
10	\$15.17	0.52	\$19.58	0.48	\$10.07	0.55	\$26.39	0.45
11	\$5.05	0.59	\$13.88	0.41	\$8.67	0.30	\$8.91	0.70
12	\$12.47	0.62	\$29.83	0.38	\$22.74	0.15	\$25.10	0.85
13	\$11.16	0.28	\$21.78	0.72	\$20.91	0.34	\$21.30	0.66
14	\$6.49	0.83	\$9.61	0.17	\$4.17	0.69	\$9.87	0.31
15	\$8.10	0.80	\$16.11	0.20	\$6.18	0.87	\$22.75	0.13
16	\$6.69	0.47	\$6.88	0.53	\$0.96	0.10	\$13.86	0.90
17	\$5.02	0.10	\$24.08	0.90	\$14.41	0.93	\$23.74	0.07
18	\$1.70	0.01	\$18.56	0.99	\$2.16	0.03	\$27.68	0.97
19	\$0.00	0.12	\$22.51	0.88	\$0.73	0.29	\$19.30	0.71
20	\$0.12	0.30	\$22.57	0.70	\$2.81	0.21	\$11.53	0.79

5.2.1 Experiment 2009

The Experiment 2009 was conducted on a laptop in our lab at the University of Illinois at Urbana-Champaign in the summer of 2009[1]. There were 40 participants in the experiment[2]. The experiment was conducted in two one-hour sessions on two different days. The participants first read instructions on the laptop, and then they were presented with a practice session. They were instructed to ask any questions whenever necessary. Following this, participants were presented with a sequence of gamble pairs, one pair at a time. The gamble pairs that were presented via computers using a two-alternative forced-choice (2AFC) paradigm, in which they were not allowed to state any preference or indifference. The gamble was shown as a wheel of chance. Probabilities were displayed in a colored area. At the end of the session, one of the choices made by the participants would randomly be selected and played for real. The average payment was \$20.97 per session. Gamble pairs were ordered by the computer in a quasi-random fashion with the condition that the same gamble pair was never presented twice in succession. For each session, each gamble pair was repeated 30 times. Participants made a total of 1200 choices across both sessions. Table 5.1 shows the 20 gamble pairs in Experiment 2009.

[1]The study was approved by the Institutional Review Board (IRB) of the University of Illinois under No. 08387.
[2]Due to a data-writing error, the data from Participants 11 and 35 were never analyzed. There remain 40 participants.

5.2.2 Experiment 2012

The Experiment 2012 were conducted at the University of Illinois at Urbana-Champaign in the summer of 2012[3]. 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 has 20 gamble pairs, shown in Table 5.2. The gamble pairs in this stimulus set were adapted from Birnbaum (2008b), with rewards adjusted to fit in the experimental paradigm. The gambles have two, three, or four positive rewards. 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, with colored areas to represent probabilities and numbers next to the wheels to represent payoffs. Before starting the experiment, participants were informed that one of their choices was randomly selected and played for real at the end of each session. The average payment was $21.76 per session. 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). The analysis results of one other stimulus set in this experiment were reported in my master thesis (Guo, 2018a).

5.3 Functional Forms

Table 5.3 reports seven different probability weighting functions $w(p)$ and seven different utility functions $u(x)$ (see also Tables 2 and 3 in Stott, 2006). Thus, there are 49 different versions of CPT. In particular, Tversky and Kahneman (1992b) used Tversky-Kahneman probability weighting function and power utility function, labeled *CPT-KT* in this paper.

5.4 Probabilistic Specifications

While CPT is a deterministic theory, experimental research collects variable choice data. How can one test an algebraic theory like CPT 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. This challenge 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,

[3]The study was approved by the Institutional Review Board (IRB) of the University of Illinois under No. 12632.

Table 5.2: The 20 gamble pairs in Experiment 2012.

	Monetary gamble: Gamble 1								Monetary gamble: Gamble 0							
Pair	x_1	p_1	x_2	p_2	x_3	p_3	x_4	p_4	y_1	q_1	y_2	q_2	y_3	q_3	y_4	q_4
1	\$13	0.05	\$13	0.10	\$23	0.85	-	-	\$3	0.05	\$23	0.10	\$23	0.85	-	-
2	\$13	0.15	\$23	0.85	-	-	-	-	\$3	0.05	\$23	0.95	-	-	-	-
3	\$5	0.50	\$6	0.50	\$21	0.90	-	-	\$2	0.50	\$5	0.10	\$21	0.85	-	-
4	\$5	0.50	\$6	0.50	\$21	0.50	\$21	0.85	\$2	0.50	\$5	0.50	\$5	0.50	\$21	0.85
5	\$1	0.10	\$11	0.10	\$24	0.80	-	-	\$4	0.10	\$21	0.10	\$24	0.80	-	-
6	\$1	0.10	\$11	0.10	\$21	0.80	-	-	\$4	0.10	\$21	0.90	-	-	-	-
7	\$1	0.10	\$11	0.10	\$24	0.80	-	-	\$4	0.10	\$22	0.10	\$24	0.80	-	-
8	\$1	0.20	\$22	0.80	-	-	-	-	\$4	0.10	\$22	0.90	-	-	-	-
9	\$1	0.90	\$5	0.50	\$21	0.50	-	-	\$1	0.90	\$12	0.50	\$13.5	0.50	-	-
1	\$5	0.95	\$21	0.50	-	-	-	-	\$5	0.90	\$13.5	0.10	-	-	-	-
11	\$1	0.25	\$2	0.50	\$11	0.25	-	-	\$2	0.50	\$4	0.25	\$22	0.25	-	-
12	\$1	0.25	\$11	0.25	\$24.5	0.50	-	-	\$4	0.25	\$22	0.25	\$24.5	0.50	-	-
13	\$1.5	0.59	\$11.5	0.20	\$12.5	0.20	\$24	0.10	\$1.5	0.59	\$4.5	0.20	\$21.5	0.20	\$24	0.10
14	\$1.5	0.10	\$11.5	0.20	\$12.5	0.20	\$24	0.59	\$1.5	0.10	\$4.5	0.20	\$21.5	0.20	\$24	0.59
15	\$0.5	0.60	\$14	0.20	\$14.5	0.20	-	-	\$0.5	0.60	\$1.5	0.20	\$21	0.20	-	-
16	\$0.5	0.10	\$14	0.45	\$14.5	0.45	-	-	\$0.5	0.10	\$1.5	0.45	\$21	0.45	-	-
17	\$1	0.50	\$11	0.50	-	-	-	-	\$1.5	0.50	\$21	0.50	-	-	-	-
18	\$0.5	0.40	\$1	0.48	\$11	0.48	-	-	\$0.5	0.40	\$1.5	0.48	\$21	0.48	-	-
19	\$1	0.10	\$11	0.10	\$23	0.80	-	-	\$1.5	0.10	\$21	0.10	\$23	0.80	-	-
20	\$1	0.45	\$11	0.45	\$23	0.10	-	-	\$1.5	0.45	\$21	0.45	\$23	0.10	-	-

Table 5.3: Summary of seven functional forms of the probability weighting function $w(p)$ and the utility function $u(x)$.

(a) Seven functional forms of the probability weighting function $w(p)$.

Name	Abbreviation	Equation
Linear	Lin	$w(p) = p$
Power	Pwr	$w(p) = p^{\gamma}$
Goldstein-Einhorn	GE	$w(p) = \frac{sp^{\gamma}}{sp^{\gamma}+(1-p)^{\gamma}}$
Tversky-Kahneman	TK	$w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{(\frac{1}{\gamma})}}$
Wu-Gonzalez	WG	$w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{s}}$
Prelec I	Prl-I	$w(p) = e^{-(-\ln p)^{\gamma}}$
Prelec II	Prl-II	$w(p) = e^{-s(-\ln p)^{\gamma}}$

(b) Seven functional forms of the utility function $u(x)$.

Name	Abbreviation	Equation
Linear	Lin	$u(x) = x$
Logarithmic	Log	$u(x) = ln(\alpha + x)$
Power	Pwr	$u(x) = x^{\alpha}$
Quadratic	Quad	$u(x) = \alpha x - x^2$
Exponential	Expo	$u(x) = 1 - e^{-\alpha x}$
Bell	Bell	$u(x) = \beta x - e^{-\alpha x}$
HARA	Hara	$u(x) = -(\beta + x)^{\alpha}$

2005). To perform an appropriate and rigorous test of CPT, researchers have to solve Luce's challenge. However, only a 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 CPT. 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 5.4.1 and 5.4.2 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 (Davis-Stober, 2009, Iverson and Falmagne, 1985, Silvapulle and Sen, 2005). 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 form of CPT and test those probabilistic models with both frequentist and Bayesian order-constrained statistical methods.

5.4.1 Distance-Based Models

A distance-based model, which is also called an 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

$$\begin{array}{rcl} x \succ y & \Leftrightarrow & \theta_{xy} \geq 1 - \tau \\ x \prec y & \Leftrightarrow & \theta_{xy} \leq \tau \end{array}$$

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. When $\tau = 0.50$, this model is also named *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).

Consider the two gambles in Pair 1 in Experiment 2009, Gamble 1 = $(\$1.19, 0.35; \$29.38, 0.65)$ and Gamble 0 = $(\$3.21, 0.32; \$18.00, 0.68)$. In Gamble 1, the decision maker has a 35% chance of winning \$1.19 and a 65% chance of winning \$29.38; In Gamble 0, the decision maker has a 32% chance of winning \$3.21 and a 68% chance of winning \$18.00. I consider a specific theoretical prediction of *CPT-KT*. For $G = (x_1, p_1; x_2, p_2)$, where $0 \leq x_1 \leq x_2$, *CPT-KT* states that

$$CPT(G) = \left(1 - \frac{p_2^{\gamma}}{(p_2^{\gamma} + p_1^{\gamma})^{(\frac{1}{\gamma})}}\right) x_1^{\alpha} + \frac{p_2^{\gamma}}{(p_2^{\gamma} + p_1^{\gamma})^{(\frac{1}{\gamma})}} x_2^{\alpha}$$

There are two parameters in *CPT-KT*, the weighting parameter γ and risk attitude α. I consider one specific prediction of *CPT-KT* with $\alpha = 0.88$ and $\gamma = 0.61$. In this case, the subjective value attached to Gamble 1 is

$$\left(1 - \frac{0.65^{0.61}}{(0.65^{0.61} + 0.35^{0.61})^{(\frac{1}{0.61})}}\right) 1.19^{0.88} + \frac{0.65^{0.61}}{(0.65^{0.61} + 0.35^{0.61})^{(\frac{1}{0.61})}} 29.38^{0.88} = 10.42.$$

The subjective value attached to Gamble 0 is

$$\left(1 - \frac{0.68^{0.61}}{(0.68^{0.61} + 0.32^{0.61})^{(\frac{1}{0.61})}}\right) 3.21^{0.88} + \frac{0.68^{0.61}}{(0.68^{0.61} + 0.32^{0.61})^{(\frac{1}{0.61})}} 18.00^{0.88} = 7.97.$$

Therefore, Gamble 1 is preferred to Gamble 0. Applying *CPT-KT* with $\alpha = 0.88$ and $\gamma = 0.61$ to the other pairs in Experiment 2009. For Pairs 1, 2, 4, 6, 9, 10, 14, 15, 17, 19, and 20, Gamble 1 is preferred to Gamble 0; for the other pairs, Gamble 0 is preferred to Gamble 1. When Gamble 1 is preferred to Gamble 0, I write the predicted gamble as 1; when Gamble 0 is preferred, I write the predicted gamble as 0. Thus a decision maker who satisfies *CPT-KT* with $\alpha = 0.88$ and $\gamma = 0.61$ has the following *preference pattern*:

1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, and 1. For Pair 1, the distance-based model with upper bound $\tau = 0.50$ means that a decision maker chooses Gamble 1 over Gamble 0 with probability at least 0.50. However, a distance-based model with upper bound $\tau = 0.50$ assumes that 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 Pair 1 as an example. The distance-based model with upper bound $\tau = 0.10$ means that the decision maker chooses Gamble 1 over Gamble 0 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.

5.4.2 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 the decision maker's 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 *CPT-KT* as an example. A mixture model treats the two parameters, α and γ, as random variables with any joint distribution whatsoever, hence permitting all possible probability distributions over the various permissible preference patterns.

I write $\succ$ for strict preference. I define $\mathcal{CPT}$ as a set of preference patterns predicted by CPT and $P(\succ_{CPT})$ as the probability of preference pattern $\succ_{CPT}$ in $\mathcal{CPT}$. According to the mixture model, for any pair (x, y), the binary choice probability θ_{xy} is

$$\theta_{xy} = \sum_{\substack{\succ_{CPT} \in \mathcal{CPT} \\ \text{in which } x \succ y}} P(\succ_{CPT}).$$

This equation shows that the probability of choosing x over y equals the total probability of those preference patterns predicted by CPT in which x is strictly preferred to y.

Take *CPT-KT* and the Experiment 2009 stimuli as an example. When allowing α and γ to be random variables with any joint distribution, I obtain 54 different preference patterns[4]. The mixture model of *CPT-KT* for gambles in Experiment 2009 can be cast geometrically as the convex hull (polytope) of 54 vertices

[4]I used the grid search to get predicted preference patterns. The grid search for α considered all values in the range [0.01, 10] with a step-size of 0.01 and the range [10.05, 50] with a step-size of 0.05. The grid search for γ considered all values in the range [0.279, 1] with a step-size of 0.01.

in a suitably chosen 20-dimensional unit hypercube of binary choice probabilities. Each vertex encodes the binary choice probabilities when the probability mass is concentrated on a single preference pattern predicted by *CPT-KT*. I provide the minimal description of the mixture polytope of *CPT-KT* for gambles in Experiment 2009 regarding its facet-defining equalities and inequalities presented in Section 5.7, via the public-domain software PORTA [5]. For equalities, I get $P_{18} = 0$ and $P_2 = P_6 = 1$. These equalities mean that this mixture polytope has 17 free parameters that are restricted by Inequalities (3) to (54)) in the Section. In this case, the mixture model is not full dimensional. It is a 17-dimensional polytope in a 20-dimensional space. I cannot test this mixture model with frequentist order-constrained statistical methods because the frequentist methods only apply for full dimensional models. The Bayesian methods, on the other hand, can handle non-full-dimensional polytopes, such as the mixture *CPT-KT* model described above. I provide the minimal descriptions of the mixture polytope for all 49 versions of CPT for Experiments 2009 and 2012 in the Supplemental Materials. I only use the Bayesian order-constrained methods to test the mixture model in this paper.

For the Experiment 2012 stimuli, the equalities $P_1 = P_2$ and $P_3 = P_4$ are obtained. The reason is that gambles in Pair 1 were generated by the branch splitting of gambles in Pair 2, and Pairs 1 and 2 are essentially the same gamble pair (for more information about the branch splitting, see Birnbaum, 2008b). So are Pairs 3 and 4.

I computed the preference patterns[6] and obtained the facet-defining inequalities for different CPT forms. For CPT with the linear utility function and the linear probability weighting function, there is only one preference pattern, and there is no mixture model in this case. There are also three other CPT forms[7] of which mathematical models are so complex that I could not have their mathematical descriptions computed within a reasonable of amount time. Therefore, there are only 45 mixture models of CPT for Experiments 2009 and 2012 respectively.

In sum, I test four different probabilistic specifications of 49 forms of CPT, the distance-based models with $\tau = 0.50$, 0.25, and 0.10, and the mixture model. Therefore, I test a total of $49 \times 3 + 45 = 192$ models on each of the two stimulus sets from Experiment 2009 and Experiment 2012.

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

[6]I used the grid search to get predicted preference patterns. The grid search for α and s considered all values in the range [0.01, 10] with a step-size of 0.01 and the range [10.05, 50] with a step-size of 0.05.; the grid search for β and γ considered all values in the range [.01, 1] with a step-size of 0.01; for "TK" probability weighting function, the grid search for γ considered all values in the range [0.279, 1] with a step-size of 0.01.

[7]For Experiment 2009, the three CPT forms are "Quad" utility function with "GE", "WG", and "Prl-II" probability weighting function; and for Experiment 2012, the three CPT forms are "Quad" utility function with "WG" and "Prl-II" probability weighting function, and "Bell" utility function with "GE" probability weighting function.

5.4.3 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, which permits all conceivable binary choice probabilities and additionally places a uniform prior on them.

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 (Gelfand and Smith, 1990, Myung et al., 2005, Sedransk et al., 1985). I completed all the analyses in this paper 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)[8].

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 according to Jeffreys (1998). I use: BF_A to represent the Bayes factor of model A; BF_B to represent the Bayes factor for model B; and $BF_{AB} = \frac{BF_A}{BF_B}$ 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. I will say that a decision model "fits" if its Bayes factor against the saturated model is larger than 3.2. 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 all 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 Bayes factors. The model with the highest GBF is the one that best accounts for all participants' data jointly.

[8]The analyses were supported by XSEDE Grant NSF SES No. 130016 (PI: Michel Regenwetter).

5.5 Results

5.5.1 The Distance-Based Models

Tables 5.4 and 5.5 summarize the results for the distance-based models using both frequentist and Bayesian methods for Experiment 2009. Tables 5.6 and 5.7 summarize the frequentist and Bayesian analyses for the distance-based models for Experiment 2012. The first two columns of Tables 5.4 - 5.7 display the utility function and the probability weighting function. Columns 3, 4, 6, 7, 9, and 10 report the total number of people who are fit by the distance-based models with the error rate upper bounds 0.50, 0.25, and 0.10 for each session. Columns 5, 8, and 11 in Tables 5.4 and 5.6 report the total number of people who are consistently fit by the distance-based models across both sessions with three different values of τ. By a consistent fit of the distance-based model, I mean that there exists a set of parameter values of CPT for which the distance-based model fits in both sessions using frequentist methods. Columns 5, 8, and 11 in Tables 5.5 and 5.7 report the total number of people who are simultaneously fit by the distance-based models for both sessions with three different values of τ using Bayes factor analyses.

Tables 5.4 - 5.7 show that, as expected, for each form of CPT, 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.

The frequentist analysis in Table 5.4 shows that most of the distance-based models with $\tau = 0.50$ of CPT explain more than half of the participants' data in the separate analysis for Experiment 2009. The distance-based CPT models with $\tau = 0.50$ explain no more than half of the participants' data consistently for both sessions. This result shows that there might be some degree of "over-fitting" for the distance-based models with $\tau = 0.50$. The number of participants who replicate across Session I and Session II is much smaller than the number of participants who are fit in the separate analysis. In other words, when a model fits the data of some participants in Session I, the estimated best-fitting parameters of that model need not predict the data of the same participants in Session II. Another reason why the distance-based CPT models with $\tau = 0.50$ involve extensive "over-fitting" is that the rejection rates leap up when I place stronger restrictions on error rates, and there are barely any successful replications (at most two for $\tau = 0.25$ and at most one for $\tau = 0.10$). Table 5.5 shows the Bayes factor analysis results for Experiment 2009 by comparing each model to the saturated model. The Bayes factor results are very similar to the frequentist analyses in Table 5.4.

Results in Table 5.6 and 5.7 indicate that the Experiment 2012 results are similar to the Experiment 2009 results. The number of outcomes in gambles does not seem to affect the analysis results.

All of these findings tell us that the distance-based models of CPT do not perform very well, and it is

Table 5.4: Results of the frequentist analysis of the distance-based models for Experiment 2009. The column labeled "S1" reports the number of people who are fit in Session I, the column labeled "S2" reports the number of people who are fit in Session II, and the column labeled "Consis" reports the number of people who are consistently fit in both sessions. There are 40 participants in each session.

		$\tau = 0.50$			$\tau = 0.25$			$\tau = 0.10$		
$u(x)$	$w(x)$	S1	S2	Consis	S1	S2	Consis	S1	S2	Consis
Lin	Lin	9	2	1	2	1	0	0	1	0
Lin	Pwr	10	7	3	3	1	0	0	1	0
Lin	GE	24	23	13	9	10	2	2	8	1
Lin	TK	17	16	9	5	9	2	2	7	1
Lin	WG	22	23	13	9	10	2	2	8	1
Lin	Prl-I	18	17	10	6	8	2	2	7	1
Lin	Prl-II	24	22	12	9	9	2	2	7	1
Log	Lin	13	14	9	4	7	2	2	6	1
Log	Pwr	18	21	12	8	10	2	2	9	1
Log	GE	27	25	14	10	13	2	2	10	1
Log	TK	15	18	10	4	10	2	2	7	1
Log	WG	25	25	14	10	13	2	2	10	1
Log	Prl-I	17	18	11	4	9	2	2	6	1
Log	Prl-II	27	25	14	10	13	2	2	10	1
Pwr	Lin	18	19	10	6	8	2	2	7	1
Pwr	Pwr	18	20	11	6	9	2	2	8	1
Pwr	GE	30	27	18	11	11	3	3	9	2
Pwr	TK	22	23	14	8	9	2	2	7	1
Pwr	WG	27	24	15	10	10	2	2	8	1
Pwr	Prl-I	22	22	14	10	9	2	2	7	1
Pwr	Prl-II	30	27	18	11	11	3	3	9	2
Quad	Lin	4	1	1	0	0	0	0	0	0
Quad	Pwr	10	7	3	3	1	0	0	1	0
Quad	GE	19	16	9	7	6	2	0	1	0
Quad	TK	4	2	1	0	0	0	0	0	0
Quad	WG	17	12	6	5	1	0	0	1	0
Quad	Prl-I	6	3	1	0	0	0	0	0	0
Quad	Prl-II	18	12	6	5	1	0	0	1	0
Expo	Lin	14	14	9	4	7	2	2	6	1
Expo	Pwr	19	22	12	8	10	2	2	9	1
Expo	GE	30	30	18	12	16	4	3	10	2
Expo	TK	18	17	10	4	8	2	2	6	1
Expo	WG	27	24	14	10	11	2	2	9	1
Expo	Prl-I	19	17	11	4	8	2	2	6	1
Expo	Prl-II	27	24	14	10	11	2	2	9	1
Bell	Lin	17	15	9	6	8	2	2	7	1
Bell	Pwr	18	21	12	8	10	2	2	9	1
Bell	GE	31	31	19	14	18	4	3	11	2
Bell	TK	22	20	10	6	11	2	2	8	1
Bell	WG	28	26	15	12	13	2	2	10	1
Bell	Prl-I	24	22	12	7	11	2	2	8	1
Bell	Prl-II	28	26	15	12	13	2	2	10	1
HARA	Lin	1	0	0	0	0	0	0	0	0
HARA	Pwr	1	0	0	0	0	0	0	0	0
HARA	GE	1	0	0	0	0	0	0	0	0
HARA	TK	1	0	0	0	0	0	0	0	0
HARA	WG	1	0	0	0	0	0	0	0	0
HARA	Prl-I	1	0	0	0	0	0	0	0	0
HARA	Prl-II	1	0	0	0	0	0	0	0	0

important to perform analyses at the individual level and use replications in this kind of decision research.

5.5.2 Mixture Model

Tables 5.8 and 5.9 show the number of people who are fit by the mixture model of CPT by Bayes factor on the data of Experiment 2009 and Experiment 2012, separately and simultaneously. Most of the mixture models of CPT fails to win over the saturated model for all of the participants for both stimulus sets. For Experiment 2009, the model that fits the highest number of participants' data uses Tversky-Kahneman probability weighting function and the power utility function (seven out of 40 for Session I, five out of 40 for Session II, and zero out of 54 for both sessions); for Experiment 2012, the model that fits the highest number of participants' data uses the Goldstein-Einhorn probability weighting function and the quadratic utility

Table 5.5: Results of the Bayesian analysis of the distance-based models for Experiment 2009. The column labeled "S1" reports the number of people who are fit in Session I, the column labeled "S2" reports the number of people who are fit in Session II, and the column labeled "Both" reports the number of people who are simultaneously fit in both sessions. There are 40 participants in each session.

		$\tau = 0.50$			$\tau = 0.25$			$\tau = 0.10$		
$u(x)$	$w(x)$	S1	S2	Both	S1	S2	Both	S1	S2	Both
Lin	Lin	10	2	1	3	1	0	0	1	0
Lin	Pwr	11	5	2	6	3	1	1	1	0
Lin	GE	24	22	17	16	15	11	6	8	4
Lin	TK	18	16	9	9	11	5	4	7	3
Lin	WG	22	22	16	15	15	11	6	8	4
Lin	Prl-I	18	15	10	10	10	6	5	7	3
Lin	Prl-II	24	21	16	16	13	9	6	7	3
Log	Lin	16	14	9	9	10	6	5	6	3
Log	Pwr	19	18	11	14	15	9	6	9	4
Log	GE	27	23	18	17	16	10	7	11	6
Log	TK	17	17	11	9	12	7	5	8	4
Log	WG	25	23	17	16	16	10	7	11	6
Log	Prl-I	18	17	12	9	12	7	5	7	3
Log	Prl-II	26	23	18	17	16	10	7	11	6
Pwr	Lin	19	16	9	12	12	6	5	7	3
Pwr	Pwr	18	17	10	11	14	8	5	8	4
Pwr	GE	29	25	21	19	20	14	8	9	5
Pwr	TK	22	22	15	15	15	10	7	7	3
Pwr	WG	26	22	17	17	17	12	7	8	4
Pwr	Prl-I	23	21	16	15	15	10	7	7	3
Pwr	Prl-II	29	25	21	19	20	14	8	9	5
Quad	Lin	2	0	0	0	0	0	0	0	0
Quad	Pwr	9	6	3	4	1	0	3	1	0
Quad	GE	18	13	9	8	7	3	5	4	1
Quad	TK	3	0	0	0	0	0	0	0	0
Quad	WG	16	8	6	7	1	1	4	1	0
Quad	Prl-I	4	1	0	1	0	0	0	0	0
Quad	Prl-II	16	7	5	6	1	0	4	1	0
Expo	Lin	15	12	8	9	10	6	5	6	3
Expo	Pwr	19	19	11	13	14	8	6	9	4
Expo	GE	29	27	22	20	20	14	9	13	7
Expo	TK	18	16	11	10	11	7	5	6	3
Expo	WG	26	22	17	17	14	9	7	9	5
Expo	Prl-I	19	16	12	10	11	7	5	6	3
Expo	Prl-II	26	22	18	17	14	9	7	9	5
Bell	Lin	17	13	8	10	12	7	5	7	3
Bell	Pwr	18	18	11	13	15	9	6	9	4
Bell	GE	30	28	24	20	21	13	9	15	8
Bell	TK	20	18	11	12	14	8	5	9	4
Bell	WG	27	23	19	18	16	10	7	11	6
Bell	Prl-I	23	20	15	13	14	9	6	9	4
Bell	Prl-II	27	23	19	18	16	10	7	11	6
HARA	Lin	0	0	0	0	0	0	0	0	0
HARA	Pwr	0	0	0	0	0	0	0	0	0
HARA	GE	0	0	0	0	0	0	0	0	0
HARA	TK	0	0	0	0	0	0	0	0	0
HARA	WG	0	0	0	0	0	0	0	0	0
HARA	Prl-I	0	0	0	0	0	0	0	0	0
HARA	Prl-II	0	0	0	0	0	0	0	0	0

Table 5.6: Results of the frequentist analysis of the distance-based models for Experiment 2012. The column labeled "S1" reports the number of people who are fit in Session I, the column labeled "S2" reports the number of people who are fit in Session II, and the column labeled "Consis" reports the number of people who are consistently fit in both sessions. There are 67 participants in Session I and 54 in Session II.

		$\tau = 0.50$			$\tau = 0.25$			$\tau = 0.10$		
$u(x)$	$w(x)$	S1	S2	Consis	S1	S2	Consis	S1	S2	Consis
Lin	Lin	8	12	3	0	0	0	0	0	0
Lin	Pwr	9	13	3	1	0	0	1	0	0
Lin	GE	29	29	16	5	5	2	3	2	1
Lin	TK	13	19	7	0	0	0	0	0	0
Lin	WG	19	20	8	1	0	0	1	0	0
Lin	Prl-I	14	17	7	0	0	0	0	0	0
Lin	Prl-II	30	34	16	6	9	2	3	4	1
Log	Lin	24	27	14	5	7	2	2	3	1
Log	Pwr	28	32	15	6	7	2	3	3	1
Log	GE	30	33	16	6	8	2	3	4	1
Log	TK	24	29	15	5	7	2	2	3	1
Log	WG	30	32	16	6	7	2	3	3	1
Log	Prl-I	25	29	15	5	7	2	2	3	1
Log	Prl-II	35	37	18	6	8	2	3	4	1
Pwr	Lin	26	29	14	6	7	2	3	3	1
Pwr	Pwr	29	32	15	6	7	2	3	3	1
Pwr	GE	30	33	17	6	8	2	3	4	1
Pwr	TK	29	32	16	6	7	2	3	3	1
Pwr	WG	31	32	17	6	7	2	3	3	1
Pwr	Prl-I	29	31	16	6	7	2	3	3	1
Pwr	Prl-II	31	35	17	6	9	2	3	4	1
Quad	Lin	24	26	14	4	5	2	2	2	1
Quad	Pwr	28	28	16	4	7	2	2	3	1
Quad	GE	29	30	16	4	7	2	2	3	1
Quad	TK	27	29	16	4	5	2	2	2	1
Quad	WG	31	32	17	4	7	2	2	3	1
Quad	Prl-I	26	29	16	4	5	2	2	2	1
Quad	Prl-II	35	36	20	4	8	2	2	4	1
Expo	Lin	24	27	14	5	7	2	2	3	1
Expo	Pwr	27	31	15	6	7	2	3	3	1
Expo	GE	36	35	18	6	8	2	3	4	1
Expo	TK	24	30	15	5	7	2	2	3	1
Expo	WG	34	35	19	6	7	2	3	3	1
Expo	Prl-I	31	32	17	5	7	2	2	3	1
Expo	Prl-II	35	37	18	6	8	2	3	4	1
Bell	Lin	25	29	14	5	7	2	2	3	1
Bell	Pwr	29	32	15	6	7	2	3	3	1
Bell	GE	30	33	17	6	8	2	3	4	1
Bell	TK	26	32	15	5	7	2	2	3	1
Bell	WG	26	29	17	4	5	2	2	2	1
Bell	Prl-I	26	31	15	5	7	2	2	3	1
Bell	Prl-II	30	35	16	6	9	2	3	4	1
HARA	Lin	20	20	11	0	0	0	0	0	0
HARA	Pwr	22	20	11	0	0	0	0	0	0
HARA	GE	26	21	13	0	0	0	0	0	0
HARA	TK	22	20	11	0	0	0	0	0	0
HARA	WG	23	21	12	0	0	0	0	0	0
HARA	Prl-I	22	20	11	0	0	0	0	0	0
HARA	Prl-II	23	22	12	0	0	0	0	0	0

Table 5.7: Results of the Bayes factor analysis of the distance-based models for Experiment 2012. The column labeled "S1" reports the number of people who are fit in Session I, the column labeled "S2" reports the number of people who are fit in Session II, and the column labeled "Both" reports the number of people who are simultaneously fit in both sessions. There are 67 participants in Session I and 54 in Session II.

		$\tau = 0.50$			$\tau = 0.25$			$\tau = 0.10$		
$u(x)$	$w(x)$	S1	S2	Both	S1	S2	Both	S1	S2	Both
Lin	Lin	2	0	0	0	0	0	0	0	0
Lin	Pwr	5	3	2	1	0	0	1	0	0
Lin	GE	14	14	6	8	7	4	4	5	2
Lin	TK	2	2	0	0	0	0	0	0	0
Lin	WG	4	2	0	1	0	0	1	0	0
Lin	Prl-I	3	2	0	0	0	0	0	0	0
Lin	Prl-II	17	17	8	10	11	6	4	9	3
Log	Lin	17	20	10	9	11	5	3	7	2
Log	Pwr	16	17	8	10	9	5	4	7	2
Log	GE	14	16	8	10	9	5	4	8	2
Log	TK	13	17	8	9	9	5	3	7	2
Log	WG	15	15	8	10	9	5	4	7	2
Log	Prl-I	13	19	9	9	9	5	3	7	2
Log	Prl-II	21	21	11	10	11	5	4	8	2
Pwr	Lin	18	20	10	10	10	5	4	7	2
Pwr	Pwr	17	17	8	10	9	5	4	7	2
Pwr	GE	14	15	8	10	9	5	4	8	2
Pwr	TK	15	17	8	10	9	5	4	7	2
Pwr	WG	14	13	8	10	9	5	4	7	2
Pwr	Prl-I	15	16	7	10	9	5	4	7	2
Pwr	Prl-II	17	18	9	10	11	6	4	9	3
Quad	Lin	19	22	11	8	9	5	3	6	2
Quad	Pwr	18	20	12	7	10	5	3	7	2
Quad	GE	18	16	8	7	8	4	3	7	2
Quad	TK	17	20	11	7	8	5	3	5	2
Quad	WG	21	16	9	7	9	4	3	7	2
Quad	Prl-I	18	21	12	7	8	5	3	5	2
Quad	Prl-II	21	24	10	8	12	4	3	8	2
Expo	Lin	15	19	9	8	10	4	3	7	2
Expo	Pwr	16	16	7	9	9	4	4	7	2
Expo	GE	21	19	9	10	11	5	4	8	2
Expo	TK	14	17	7	8	9	4	3	7	2
Expo	WG	21	19	10	9	10	4	4	7	2
Expo	Prl-I	20	20	10	9	10	4	3	7	2
Expo	Prl-II	21	21	12	10	12	6	4	9	3
Bell	Lin	16	18	10	9	10	5	3	7	2
Bell	Pwr	16	17	8	10	9	5	4	7	2
Bell	GE	13	15	8	10	9	5	4	8	2
Bell	TK	13	16	7	9	9	5	3	7	2
Bell	WG	13	11	8	7	7	4	3	5	2
Bell	Prl-I	13	15	7	9	9	5	3	7	2
Bell	Prl-II	18	18	9	10	11	6	4	9	3
HARA	Lin	16	14	4	2	2	1	0	1	0
HARA	Pwr	9	6	3	1	1	0	0	0	0
HARA	GE	8	7	2	1	2	0	0	0	0
HARA	TK	8	6	3	1	1	0	0	0	0
HARA	WG	9	6	2	1	1	0	0	0	0
HARA	Prl-I	9	6	3	1	1	0	0	0	0
HARA	Prl-II	7	7	1	1	1	0	0	0	0

Table 5.8: Results of the Bayes factor analysis for the mixture models for Experiment 2009. It shows the number of people who are fit by the mixture model, separately and simultaneously. There are 40 participants in each session.

u(x)	w(x)	Session I	Session II	Both Sessions
Lin	Pwr	0	0	0
Lin	GE	3	2	0
Lin	TK	2	4	0
Lin	WG	2	1	0
Lin	Prl-I	1	4	0
Lin	Prl-II	2	2	0
Log	Lin	0	0	0
Log	Pwr	0	4	0
Log	GE	3	1	0
Log	TK	1	2	0
Log	WG	1	1	0
Log	Prl-I	2	2	0
Log	Prl-II	1	1	0
Pwr	Lin	0	3	0
Pwr	Pwr	0	4	0
Pwr	GE	4	2	1
Pwr	TK	7	5	0
Pwr	WG	3	1	0
Pwr	Prl-I	4	2	0
Pwr	Prl-II	5	2	0
Quad	Lin	0	0	0
Quad	Pwr	0	0	0
Quad	TK	0	0	0
Quad	Prl-I	0	0	0
Expo	Lin	0	3	0
Expo	Pwr	0	1	0
Expo	GE	3	3	0
Expo	TK	1	3	0
Expo	WG	2	1	0
Expo	Prl-I	2	2	0
Expo	Prl-II	1	1	0
Bell	Lin	0	2	0
Bell	Pwr	0	3	0
Bell	GE	2	2	0
Bell	TK	2	5	0
Bell	WG	1	1	0
Bell	Prl-I	3	4	0
Bell	Prl-II	2	1	0
Hara	Lin	0	0	0
Hara	Pwr	0	0	0
Hara	GE	0	0	0
Hara	TK	0	0	0
Hara	WG	0	0	0
Hara	Prl-I	0	0	0
Hara	Prl-II	0	0	0

function (12 out of 67 for Session I, 21 out of 54 for Session II, and nine out of 54 for both sessions). Overall, the Bayes factor analysis shows that none of the mixture models of CPT could explain the participants' data very well.

5.5.3 Model Comparison: Individual Level

I use Bayes factors to compare models. As I discuss in Section 5.4.3, 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 Experiments 2009 and 2012.

Table 5.10 and 5.11 report the results of the model comparison by Bayes factor for Experiments 2009 and 2012. The first column shows the participant ID. The second and third columns show the utility function and the probability weighting function of the best form of CPT. The fourth column shows the stochastic

Table 5.9: Results of the Bayes factor analysis for the mixture models for Experiment 2012. The table shows the number of people who are fit by the mixture model, separately and simultaneously. There are 67 participants in Session I and 54 in Session II.

u(x)	w(x)	Session I	Session II	Both Sessions
Lin	Pwr	0	0	0
Lin	GE	0	0	0
Lin	TK	0	0	0
Lin	WG	0	0	0
Lin	Prl-I	0	0	0
Lin	Prl-II	1	2	0
Log	Lin	0	0	0
Log	Pwr	0	0	0
Log	GE	0	0	0
Log	TK	0	1	0
Log	WG	0	0	0
Log	Prl-I	0	1	0
Log	Prl-II	3	1	0
Pwr	Lin	0	2	0
Pwr	Pwr	0	0	0
Pwr	GE	0	0	0
Pwr	TK	0	0	0
Pwr	WG	0	0	0
Pwr	Prl-I	0	0	0
Pwr	Prl-II	2	3	0
Quad	Lin	0	0	0
Quad	Pwr	0	1	0
Quad	GE	12	21	8
Quad	TK	3	1	0
Quad	Prl-I	2	1	1
Expo	Lin	0	1	0
Expo	Pwr	0	0	0
Expo	GE	4	2	1
Expo	TK	0	1	0
Expo	WG	4	3	0
Expo	Prl-I	1	3	0
Expo	Prl-II	4	1	1
Bell	Lin	0	1	0
Bell	Pwr	0	0	0
Bell	TK	0	0	0
Bell	WG	0	0	0
Bell	Prl-I	0	0	0
Bell	Prl-II	1	0	0
HARA	Lin	0	0	0
HARA	Pwr	0	0	0
HARA	GE	0	0	0
HARA	TK	0	0	0
HARA	WG	0	1	0
HARA	Prl-I	0	0	0
HARA	Prl-II	0	6	0

form and the upper bound τ on the error rate (when applicable). I use the value of the upper bound τ for the distance-based model and "Mixture" to represent the mixture model. The fifth column shows the Bayes factor for the best model compared to the saturated model. The sixth column shows the Bayes factor between the best and second-best models. There are two sessions in each experiment.

In Experiment 2009, for eight out of 40 participants in Session I and 11 out of 40 participants in Session II, none of the 192 probabilistic models of CPT win over the saturated model. For Session I, out of the 32 participants who are best fit by CPT, five are best fit by the mixture model, and 27 are best fit by the distance-based model. For Session II, out of the 29 participants who are best fit by CPT, six are best fit by the mixture model, and 23 are best fit by the distance-based model. There are many variations regarding the functional forms of the utility function and the probability weighting function of the best CPT. Overall, there is no particular form of CPT that is the best across the board.

In Experiment 2012, for 28 out of 67 participants in Session I and 12 out of 54 participants in Session II, none of the 192 probabilistic models of CPT win over the saturated model. For Session I, out of the 39 participants who are best fit by CPT, ten are best fit by the mixture model and 29 are best fit by the distance-based model. For Session II, out of the 32 participants who are best fit by CPT, 16 are best fit by the mixture model, and 16 are best fit by the distance-based model. One thing to note is that CPT with the quadratic utility function and the Goldstein-Einhorn probability weighting function is the core theory of the winners for all ten participants who are best fit by a mixture model in Session I, and 13 out of 16 participants who are best fit by a mixture model in Session II. The distance-based model of CPT with the quadratic utility function and the linear probability weighting function wins out for the highest number of participants separately in each session and simultaneously for both sessions (15 in Session I, 13 in Session II, and eight for both sessions simultaneously). For Experiment 2012, the distance-based model of CPT with the quadratic utility function and the linear probability weighting function, and the mixture model of CPT with the quadratic utility function and the Goldstein-Einhorn probability weighting function, are the two best performing theories [9].

Overall, there is much evidence for heterogeneity across individuals and stimulus sets in terms of the best model. No single form of CPT, or type of probabilistic specifications, is robust across all participants and stimulus sets.

[9]The mixture model of CPT with the quadratic utility function and the Goldstein-Einhorn probability weighting function is not available for Experiment 2009.

Table 5.10: The individual model comparison results for Experiment 2009 by Bayes factor. The Bayes factors larger than one billion are shown in scientific form. There are 40 participants in each session.

	Session I					Session II				
ID	$u(x)$	$w(p)$	Form	Best BF	Best/Second	$u(x)$	$w(p)$	Form	Best BF	Best/Second
1	Log	Lin	0.1	6.84E+16	2	Log	Lin	0.1	2.85E+18	2
2	-	-	-	-	-	-	-	-	-	-
3	Log	Pwr	0.25	236564122	2	-	-	-	-	-
4	Pwr	Prl-I	Mixture	4677359	1	Pwr	TK	0.5	1737	1
5	Log	Prl-I	Mixture	25716519529	27	Log	Lin	0.1	1.50E+15	2
6	Log	TK	Mixture	4704363370	16	Lin	Prl-I	Mixture	9.03E+19	11
7	Pwr	Prl-II	0.1	2249404552328	1	Pwr	Prl-II	0.1	1.54E+17	1
8	Log	Prl-I	0.5	22969	1	Lin	TK	0.5	70164	2
9	-	-	-	-	-	Expo	GE	0.1	576801	2
10	Log	Prl-I	0.5	36	2	Log	Prl-I	0.5	148	1
11	Lin	Prl-I	0.1	21961380736	4	Log	Pwr	0.1	137649511683	2
12	Log	Lin	0.5	106300	1	Lin	GE	Mixture	96634843	1
13	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	Lin	Lin	0.1	2.41E+17	8
15	Expo	GE	0.25	12237507	2	Expo	GE	0.25	3342184	2
16	Log	Lin	0.5	5010	2	Expo	GE	0.25	19300	2
17	Log	Lin	0.25	621875	2	Log	Lin	0.5	2470	2
18	Lin	TK	Mixture	1.85E+18	30	Bell	Lin	Mixture	1.93E+23	906
19	-	-	-	-	-	Bell	Lin	Mixture	1.97E+23	739
20	Lin	Lin	0.5	26834	7	Expo	GE	0.5	92	2
21	Lin	Lin	0.5	67023	1	Expo	GE	0.25	43176	2
22	Expo	WG	Mixture	9343129	3	Quad	GE	0.25	45629	6
23	Pwr	Prl-II	0.5	8	1	Lin	Pwr	0.5	7651	2
24	Lin	Lin	0.5	36332	1	-	-	-	-	-
25	Bell	Prl-II	0.25	2148589	1	Bell	TK	Mixture	68043073655810	1581179
26	Log	Lin	0.5	5757	1	Log	Lin	0.5	113	2
27	Pwr	Prl-II	0.5	2479	1	Log	Lin	0.25	3138476	2
28	Expo	TK	0.5	21	1	-	-	-	-	-
29	Quad	Pwr	0.25	1169080	4	-	-	-	-	-
30	Lin	WG	0.5	10307	1	Expo	WG	0.5	4627	3
31	Lin	Lin	0.25	134117784	8	Log	Lin	0.1	4.87E+16	2
32	Expo	WG	0.5	744	1	-	-	-	-	-
33	Lin	Pwr	0.25	85519818	2	Log	Pwr	0.1	2.81E+17	1
34	Lin	Lin	0.25	11594220	8	-	-	-	-	-
35	Expo	TK	0.25	7489	1	Pwr	Prl-II	Mixture	44069	37
36	-	-	-	-	-	-	-	-	-	-
37	-	-	-	-	-	-	-	-	-	-
38	-	-	-	-	-	-	-	-	-	-
39	Quad	GE	0.25	15469	2	Quad	GE	0.1	2288266172	47
40	Log	Lin	0.25	13518355372	1	Log	TK	0.1	26135885021292	3

Table 5.11: The individual model comparison results for Experiment 2012 by Bayes factor. The Bayes factors larger than one billion are shown in scientific form. There are 67 participants in Session I and 54 in Session II.

	Session I					Session II				
ID	$u(x)$	$w(p)$	Form	Best BF	Best/Second	$u(x)$	$w(p)$	Form	Best BF	Best/Second
1	Quad	Lin	0.25	3.21E+09	3	Quad	Lin	0.1	1.84E+10	3
2	Pwr	Prl-II	0.5	9	1	Lin	Pwr	0.5	11	2
4	Quad	Lin	0.25	6706	3	Quad	Lin	0.25	74	1
5	-	-	-	-	-	-	-	-	-	-
7	Quad	Lin	0.5	405	5	Quad	Lin	0.5	18621	3
9	Quad	Lin	0.25	1.09E+10	3	Quad	Lin	0.5	74	3
11	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	Lin	Pwr	0.5	491	2
14	Log	Lin	0.5	4813	2	Expo	Lin	0.1	6.04E+12	1
15	-	-	-	-	-	Quad	Lin	0.5	22	6
16	-	-	-	-	-	Log	Lin	0.5	42629	1
17	Quad	Lin	0.5	22046	2	Log	Lin	0.25	43441100	1
18	-	-	-	-	-	Quad	Lin	0.5	53	3
19	Quad	GE	Mixture	37	55	Pwr	Prl-II	Mixture	5	2
20	-	-	-	-	-	-	-	-	-	-
21	Hara	Lin	0.25	13	1	Hara	Lin	0.25	5301	6
22	Quad	Lin	0.5	244	2	Quad	GE	Mixture	72	3
23	Quad	GE	Mixture	252	122	Quad	GE	Mixture	903	60
24	-	-	-	-	-	Expo	Prl-I	0.5	124	2
25	-	-	-	-	-	Quad	Lin	0.5	59	3
26	Quad	Lin	0.1	2.25E+16	3	Quad	Lin	0.1	9.10E+17	3
27	-	-	-	-	-	Quad	GE	Mixture	1614	1
28	Quad	GE	Mixture	3479	34	Quad	GE	Mixture	1039	6
29	Quad	GE	Mixture	2020	40	Quad	WG	0.5	9	1
30	-	-	-	-	-	Hara	Prl-II	Mixture	31	12
31	-	-	-	-	-	-	-	-	-	-
32	-	-	-	-	-	-	-	-	-	-
33	-	-	-	-	-	-	-	-	-	-
34	Quad	Lin	0.5	419	1	Quad	GE	Mixture	6391	763
35	-	-	-	-	-	Expo	TK	0.5	340	1
36	Quad	Lin	0.5	609	5	-	-	-	-	-
37	Quad	Pwr	0.5	260	1	Quad	TK	0.5	45	1
38	Quad	Lin	0.5	569	5	Quad	Lin	0.5	166	3
39	Quad	GE	Mixture	11	3	Quad	GE	Mixture	488	158
41	-	-	-	-	-	Quad	Lin	0.25	1.30E+09	3
42	Quad	Lin	0.5	32274	1	Quad	Lin	0.25	103029067	3
43	-	-	-	-	-	-	-	-	-	-
44	Quad	Lin	0.5	478	6	Log	Lin	0.5	19	1
46	-	-	-	-	-	-	-	-	-	-
47	Quad	Lin	0.25	5664441	3	Quad	Lin	0.5	10551	3
48	Expo	Prl-I	0.5	5344	3	Quad	GE	Mixture	403	81
49	Hara	Lin	0.5	18	6	Hara	Prl-II	Mixture	140	3
50	Quad	Lin	0.5	32928	3	Quad	GE	Mixture	64024	21
52	Quad	GE	Mixture	8	3	Quad	GE	Mixture	16	11
53	-	-	-	-	-	Quad	GE	Mixture	20	33
55	-	-	-	-	-	Quad	Lin	0.1	3.51E+18	3
56	Quad	GE	Mixture	4333	370	Quad	GE	Mixture	1972	174
58	Lin	Pwr	0.5	254	3	-	-	-	-	-
59	Quad	GE	Mixture	2409	701	Quad	GE	Mixture	285	78
61	-	-	-	-	-	Quad	GE	Mixture	154	29
65	-	-	-	-	-	-	-	-	-	-
66	Lin	Pwr	0.1	1.52E+15	3	Lin	Prl-II	0.25	8645523	2
67						Lin	Prl-II	0.1	2.58E+10	1
3	Quad	GE	Mixture	157	13					
6	-	-	-	-	-					
8	-	-	-	-	-					
10	Quad	Lin	0.5	21	3					
40	-	-	-	-	-					
45	Quad	TK	0.5	298	1					
51	Quad	GE	Mixture	464	358					
54	-	-	-	-	-					
57	Hara	Lin	0.5	4	3					
60	Lin	Prl-I	0.5	11	1					
62	Hara	Lin	0.25	2730	6					
63	Log	Lin	0.5	56408	1					
64	Expo	TK	0.5	40	1					

Table 5.12: The top five models by GBF in Session I and Session II of Experiment 2009.

	Session I				Session II			
Ranking	$u(x)$	$w(p)$	Form	Log(GBF)	$u(x)$	$w(p)$	Form	Log(GBF)
1	Pwr	Prl-II	0.5	43.07	Expo	GE	0.5	12.81
2	Pwr	GE	0.5	41.90	Bell	GE	0.5	9.38
3	Bell	GE	0.5	40.24	-	-	-	-
4	Expo	GE	0.5	39.44	-	-	-	-
5	Lin	Prl-II	0.5	35.54	-	-	-	-

5.5.4 Model Comparison: Group Level

I use group Bayes factor (GBF; Stephan et al., 2007) to select among models at the group level. The GBF aggregates likelihoods across participants and is the product of the individual-level Bayes factors. Table 5.12 shows the top five models by GBF for Experiment 2009. The first column shows the ranking of each model; the second and third columns show the functional forms of the utility function and the probability weighting function; the fourth column shows the stochastic form of the model; and the fifth column shows the log10 value of GBF for each model. For Session II, only two models beat the saturated model by GBF. The distance-based models with $\tau = 0.50$ of CPT with the exponential and Bell utility function, and the Goldstein-Einhorn probability weighting function rank top five in terms of accounting for all participants' choices jointly for each session of Experiment 2009. For Experiment 2012, none of the 192 models of CPT win over the saturated model for Session I or Session II. In other words, all of the 192 models of CPT perform poorly in terms of accounting for all participants' choices jointly for Experiment 2012.

5.6 Conclusions

Cumulative Prospect Theory is the most famous contemporary theory of risky choice. Many papers studying CPT used only one specific form of the utility function and the probability weighting function, and some even only used one specific set of parameter values for the utility function and the probability weighting function (e.g., Birnbaum, 2008b, Harrison et al., 2010, Rieger et al., 2017). In this paper, I consider 49 combinations for functional forms with four different probabilistic specifications, the distance-based models with the error rate upper bound τ =0.50, 0.25, 0.10, and the mixture model, on two different stimulus sets.

The analysis shows the distance-based model of CPT have been systematically violated by the participants, no matter whether using frequentist or Bayesian methods and no matter for Experiment 2009 or Experiment 2012. Out of the 49 forms of CPT, the most lenient distance-based model could consistently account across two sessions for only at most $\frac{1}{2}$ of the participants from Experiment 2009 and $\frac{1}{3}$ of the participants from Experiment 2012 in the frequentist tests, and for at most $\frac{1}{2}$ of participants from Experiment 2009 and $\frac{1}{5}$ of the participants from Experiment 2012 in the Bayesian analysis. When the error rate bound

τ goes down to 0.25 or less, the distance-based model could only account (with replication) for around $\frac{1}{7}$ of the participants from Experiment 2009 and around $\frac{1}{25}$ of the participants from Experiment 2012 in the frequentist tests, and at most $\frac{1}{3}$ of the participants from Experiment 2009 and $\frac{1}{10}$ of the participants from Experiment 2012 in the Bayesian tests. In sum, the distance-based model analysis of CPT consistently shows poor model performance, regardless of which functional form of CPT, whether I permit high or low error rate upper bound, whether I use frequentist or Bayesian methods, and whether I use gambles with only two rewards or gambles with up to four rewards. Most of the mixture models of CPT fails to win over the saturated model for all of the participants using the Bayes factor analysis for both Experiment 2009 and Experiment 2012, suggesting poor performance. One thing to mention is that the distance-based model and mixture model analysis depends on the assumption that the grid search of the parameter space identified all preference patterns of interest.

The model comparison at the individual level 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), Guo (2018a), Hey (2005), Loomes et al. (2002), and Regenwetter et al. (2014).

The model comparison at the group level shows that the distance-based models with $\tau = 0.50$ of CPT with the exponential and Bell utility function, and the Goldstein-Einhorn probability weighting function rank top 5 for each session of Experiment 2009; and all of the 192 probabilistic models of CPT perform poorly in terms of accounting for all participants' choices jointly for Experiment 2012.

The paper also shows the dangers of overfitting and the need for replication in decision-making research. Last but not least, I would like to point out that the paper is the largest-scale project for a systematic test of CPT. All the quantitative analysis in this paper consumed about 307,000 CPU hours on the supercomputer at the Pittsburgh Supercomputing Center.

5.7 Supplement Materials

The following provides the minimal description of the mixture polytope of *CPT-KT* for gambles in Experiment 2009 in terms of its facet-defining equalities and inequalities.

Equalities:

$$P_{18} = 0 \tag{5.1}$$

$$P_2 = P_6 = 1 \tag{5.2}$$

Inequalities:

$$P_5 - P_7 - P_{10} - P_{14} - P_{17} \leq -2 \tag{5.3}$$

$$P_1 + P_5 - P_7 - P_{10} - P_{14} - P_{17} - P_{19} \leq -2 \tag{5.4}$$

$$-P_3 - P_{10} \leq -1 \tag{5.5}$$

$$-P_4 - P_8 \leq -1 \tag{5.6}$$

$$-P_4 - P_{12} \leq -1 \tag{5.7}$$

$$-P_9 - P_{14} \leq -1 \tag{5.8}$$

$$-P_{11} - P_{14} \leq -1 \tag{5.9}$$

$$-P_{15} - P_{20} \leq -1 \tag{5.10}$$

$$P_5 - P_7 - P_{14} \leq -1 \tag{5.11}$$

$$P_1 - P_9 - P_{14} - P_{19} \leq -1 \tag{5.12}$$

$$-P_5 \leq 0 \tag{5.13}$$

$$-P_{13} \leq 0 \tag{5.14}$$

$$-P_1 + P_{20} \leq 0 \tag{5.15}$$

$$-P_{14} + P_{15} \leq 0 \tag{5.16}$$

$$-P_{19} + P_{20} \leq 0 \tag{5.17}$$

$$P_{11} - P_{17} \leq 0 \tag{5.18}$$

$$P_5 - P_{16} \leq 0 \tag{5.19}$$

$$P_3 - P_{20} \leq 0 \tag{5.20}$$

$$-P_9 - P_{14} + P_{16} + P_{17} \leq 0 \tag{5.21}$$

$$P_3 - P_9 - P_{14} + P_{15} \leq 0 \tag{5.22}$$

$$P_3 - P_{14} + P_{15} - P_{17} \leq 0 \tag{5.23}$$

$$P_1 + P_{11} - P_{17} - P_{19} \leq 0 \tag{5.24}$$

$$P_1 - P_9 - P_{14} + P_{16} + P_{17} - P_{19} \leq 0 \tag{5.25}$$

$$P_1 + P_3 - P_9 - P_{14} + P_{15} - P_{19} \leq 0 \tag{5.26}$$

$$P_1 + P_3 - P_{14} + P_{15} - P_{17} - P_{19} \leq 0 \tag{5.27}$$

$$P_{19} \leq 1 \tag{5.28}$$

$$P_7 \leq 1 \tag{5.29}$$

$$P_{12} + P_{14} \leq 1 \tag{5.30}$$

$$P_{10} + P_{11} \leq 1 \tag{5.31}$$

$$P_8 + P_{14} \leq 1 \tag{5.32}$$

$$P_5 + P_{17} \leq 1 \tag{5.33}$$

$$P_5 + P_9 \leq 1 \tag{5.34}$$

$$P_4 + P_{13} \leq 1 \tag{5.35}$$

$$P_3 + P_{16} \leq 1 \tag{5.36}$$

$$P_1 + P_5 \leq 1 \tag{5.37}$$

$$-P_3 + P_7 + P_{17} \leq 1 \tag{5.38}$$

$$P_9 - P_{17} + P_{19} \leq 1 \tag{5.39}$$

$$P_7 + P_{15} - P_{16} \leq 1 \tag{5.40}$$

$$P_7 + P_9 - P_{11} \leq 1 \tag{5.41}$$

$$P_1 + P_7 - P_{19} \leq 1 \tag{5.42}$$

$$P_1 + P_5 + P_9 - P_{17} \leq 1 \tag{5.43}$$

$$P_1 + P_3 + P_{16} - P_{19} \leq 1 \tag{5.44}$$

$$P_3 - P_9 - P_{14} + P_{15} + P_{16} + P_{17} \leq 1 \tag{5.45}$$

$$P_1 + P_3 - P_9 - P_{14} + P_{15} + P_{16} + P_{17} - P_{19} \leq 1 \tag{5.46}$$

$$P_{16} + P_{17} + P_{19} \leq 2 \tag{5.47}$$

$$P_9 + P_{16} + P_{19} \leq 2 \tag{5.48}$$

$$P_7 + P_{15} + P_{17} \leq 2 \tag{5.49}$$

$$P_7 + P_9 + P_{15} \leq 2 \tag{5.50}$$

$$P_1 + P_5 + P_{16} + P_{17} \leq 2 \tag{5.51}$$

$$P_1 + P_5 + P_9 + P_{16} \leq 2 \tag{5.52}$$

$$P_3 + P_7 + P_9 + P_{15} - P_{17} \leq 2 \tag{5.53}$$

$$P_1 + P_3 + P_7 + P_9 + P_{15} - P_{17} - P_{19} \leq 2 \tag{5.54}$$

Chapter 6

Testing Cumulative Prospect Theory and Intransitive Heuristics for Gambles With Gains and Losses

6.1 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), and Cumulative Prospect Theory (Tversky and Kahneman, 1992b).

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., Brandstätter et al., 2006, González-Vallejo, 2002, Loomes and Sugden, 1987, Tversky, 1969). However, these studies contain pervasive methodological problems in collecting, modeling, and analyzing empirical data (see Section 2 of Guo (2018b) for details on these methodological problems). 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.

In this paper, I test the leading theory of risky choice, Cumulative Prospect Theory (Tversky and Kahneman, 1992b), which only permits transitive preferences. I also test two kinds of intransitive decision heuristics, the lexicographic semiorder model and the similarity model. I test these theories on three different types of gambles, i.e., gambles with gains only, gambles with losses only, and gambles with a mixture of gains and losses.

In Guo (2018a), I tested the lexicographic semiorder model and the similarity model, and compared them with the linear order model and two simple transitive heuristics. The results showed that the intransitive heuristics perform well and win over the linear order model for most participants in model comparison. One interpretation is that Bayes factor rewards parsimonious models and penalizes complex models. The linear order model gets penalized for being complex because it permits all possible transitive linear orders.

Cumulative Prospect Theory, on the other hand, allows only transitive preference patterns, and it is more parsimonious compared to the linear order model. In this paper, I test 49 different versions of Cumulative Prospect Theory and compare them with the two intransitive heuristics to find out which theory performs the best, the transitive Cumulative Prospect Theory or the intransitive heuristics.

The rest of the paper is organized as follows: Section 6.2 introduces the 49 different forms of Cumulative Prospect Theory and two kinds of intransitive heuristics, the lexicographic semiorder model and the similarity model; Section 6.3 describes five different stimulus sets used in this paper, including two sets of gambles with gains only, two sets of gambles with losses only, and one set of gambles with a mixture of gains and losses; Section 6.4 introduces two different probabilistic specifications and the statistical methods employed in this paper; Section 6.5 reports the data analysis results; and Section 6.6 concludes the paper.

6.2 Decision Theories

In this section, I introduce the transitive theory, Cumulative Prospect Theory, and describe two kinds of intransitive heuristics, the lexicographic semiorder model and the similarity model.

6.2.1 Cumulative Prospect Theory

Tversky and Kahneman (1992b) proposed *Cumulative Prospect Theory*, henceforth CPT, to describe how people make decisions under risk. It is one of the most influential decision theories in the past few decades. Tversky and Kahneman (1992b) has been cited more than 4,700 times and CPT has been applied to many different contexts, for example, management (Becker and Gerhart, 1996, Steel and König, 2006), psychology (Lopes and Oden, 1999, Trepel et al., 2005), and transportation (Gao et al., 2010, Xu et al., 2011). The following describes how CPT works. For a gamble $G = (x_1, p_1; \ldots; x_n, p_n)$, where $x_1 \leq \ldots \leq x_k \leq 0 \leq x_{k+1} \leq \ldots \leq x_n$, let $w^+(p)$ and $w^-(p)$ be the probability weighting function to capture the subjective perception of the probabilities of gains and losses respectively. Let $u^+(x)$ and $u^-(x)$ be the utility function to capture the subjective perception of gains and losses respectively. CPT states that:

$$CPT(G) = \sum_{i=1}^{k} w_i^- u^-(x_i) + \sum_{i=k+1}^{n} w_i^+ u^+(x_i)$$

where

$$w_1^- = w^-(p_1), w_i^- = w^-(p_1 + ... + p_i) - w^-(p_1 + ... + p_{i-1}), \text{ for } 2 \leq i \leq k;$$

$$w_n^+ = w^+(p_n), w_i^+ = w^+(p_i + ... + p_n) - w^+(p_{i+1} + ... + p_n), \text{ for } k+1 \leq i \leq n-1.$$

Stott (2006) investigated "Cumulative Prospect Theory's Functional Menagerie" by considering seven

different functional forms for the utility function for gains, seven functional forms for the probability weighting function, and four probabilistic response mechanisms based on the assumption that the decision maker has a deterministic preference and that uncertainty in choice is due to noise/error. In this paper, I consider the same 49 combinations for functional forms on new stimuli, and with more general and more diverse non-parametric probabilistic specifications.

Table 6.1: Summary of seven functional forms of the probability weighting function and the utility function, for gains and losses.

(a) Seven functional forms of the probability weighting function for gains and losses.

Name	Abbreviation	Equation for Gains	Equation for Losses
Linear	Lin	$w^+(p) = p$	$w^-(p) = p$
Power	Pwr	$w^+(p) = p^{\gamma^+}$	$w^-(p) = p^{\gamma^-}$
Goldstein-Einhorn	GE	$w^+(p) = \frac{s^+ p^{\gamma^+}}{s^+ p^{\gamma^+} + (1-p)^{\gamma^+}}$	$w^-(p) = \frac{s^- p^{\gamma^-}}{s^- p^{\gamma^-} + (1-p)^{\gamma^-}}$
Tversky-Kahneman	TK	$w^+(p) = \frac{p^{\gamma^+}}{(p^{\gamma^+} + (1-p)^{\gamma^+})^{(\frac{1}{\gamma^+})}}$	$w^-(p) = \frac{p^{\gamma^-}}{(p^{\gamma^-} + (1-p)^{\gamma^-})^{(\frac{1}{\gamma^-})}}$
Wu-Gonzalez	WG	$w^+(p) = \frac{p^{\gamma^+}}{(p^{\gamma^+} + (1-p)^{\gamma^+})^{s^+}}$	$w^-(p) = \frac{p^{\gamma^-}}{(p^{\gamma^-} + (1-p)^{\gamma^-})^{s^-}}$
Prelec I	Prl-I	$w^+(p) = e^{-(-\ln p)^{\gamma^+}}$	$w^-(p) = e^{-(-\ln p)^{\gamma^-}}$
Prelec II	Prl-II	$w^+(p) = e^{-s^+(-\ln p)^{\gamma^+}}$	$w^-(p) = e^{-s^-(-\ln p)^{\gamma^-}}$

(b) Seven functional forms of the utility function for gains and losses.

Name	Abbreviation	Equation for Gains	Equation for Losses
Linear	Lin	$u^+(x) = x$	$u^-(x) = x$
Logarithmic	Log	$u^+(x) = ln(\alpha^+ + x)$	$u^-(x) = -ln(\alpha^- - x)$
Power	Pwr	$u^+(x) = x^{\alpha^+}$	$u^-(x) = -(-x)^{\alpha^-}$
Quadratic	Quad	$u^+(x) = \alpha^+ x - x^2$	$u^-(x) = \alpha^- x + x^2$
Exponential	Expo	$u^+(x) = 1 - e^{-\alpha^+ x}$	$u^-(x) = -1 + e^{\alpha^- x}$
Bell	Bell	$u^+(x) = \beta^+ x - e^{-\alpha^+ x}$	$u^-(x) = \beta^- x + e^{\alpha^- x}$
HARA	Hara	$u^+(x) = -(\beta^+ + x)^{\alpha^+}$	$u^-(x) = (\beta^- - x)^{\alpha^-}$

Table 6.1 reports seven different functional forms of the probability weighting functions and seven different functional forms of the utility functions for both gains and losses (see also Tables 2 and 3 for the functional forms of gains in Stott, 2006). Thus, there are 49 different versions of CPT. One thing to point out is that parameters in the probability weighting function and the utility function are different for gains and losses. In particular, Tversky and Kahneman (1992b) used Tversky-Kahneman probability weighting function and power utility function, labeled *CPT-KT* in this paper.

6.2.2 Lexicographic Semiorder Model

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 are intransitive decision models.

In this paragraph, I explain how a lexicographic semiorder works. 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, $x \prec_i y$ to denote that the decision maker prefers y to x on attribute i, and $x \sim_i y$ to denote that the decision maker is indifferent between x and y on attribute i. A lexicographic semiorder model works as follows:

1. The decision maker considers gamble attributes sequentially, for example, first the maximum gain and then the probability of maximum gain, or first the probability of maximum gain and then the maximum gain. For each attribute i, the decision maker uses a threshold $\epsilon_i > 0$.

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 reach a decision 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, $x \sim y$.

One may use the linear utility function in Table 6.1. 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, the log utility function in Table 6.1 is used. In this paper, I consider two kinds of lexicographic semiorder models, one uses the linear function for utility (labeled as LSO-Diff), and the other one uses the log utility function (labeled as LSO-Ratio). For details of the lexicographic semiorder model, please see Section 2 of Chapter 3.

6.2.3 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.

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 $x \succ_2 y$	$x \prec_1 y$ and $x \prec_2 y$	$x \succ_1 y$ and $x \prec_2 y$
$x \succ_1 y$ and $x \sim_2 y$	$x \prec_1 y$ and $x \sim_2 y$	$x \prec_1 y$ and $x \succ_2 y$
$x \sim_1 y$ and $x \succ_2 y$	$x \sim_1 y$ and $x \prec_2 y$	$x \sim_1 y$ and $x \sim_2 y$
$x \succ y$	$x \prec y$	$x \sim y$

The first column shows all three conditions to derive $x \succ y$, the second column shows all three conditions to derive $x \prec y$, and the third column shows all three conditions to derive $x \sim y$.

In this paper, I consider two types of similarity models, one uses the linear utility function (labeled as SIM-Diff), and the other one uses the log utility function (labeled as SIM-Ratio). For details of the similarity model, please see Section 2 of Chapter 3.

6.3 The 2010 Experiment

The 2010 experiment was conducted at the University of Illinois at Urbana-Champaign using laptop computers [1]. There were 50 participants in the experiment (29 were males, and 21 were females). There are five stimulus sets in the experiment, each with five different gambles. Participants made repeated choices (20 times for each pair) over gambles pairs that were presented via computers using a two-alternative forced-choice paradigm. Each gamble was displayed as a wheel of chance, with colored areas to represent probabilities and numbers next to the wheels to represent payoffs. At the end of the experiment, two of the participants' chosen prospects were drawn at random and played out for real money. These two prospects were obtained with constraints; the first prospect only involved a positive monetary outcome and the second involved at least one negative monetary outcome. The average payment for this experiment was $21.10.

Table 6.2 shows the five different stimulus sets in the experiment. In each stimulus set, people choose from all ten possible pairwise comparisons of five gambles. All of the gambles in the experiment have only two outcomes. There are five different types of gambles in the experiment. Set 1 has gambles with only non-negative monetary outcomes, and one of the two outcomes for each gamble is $0. Set 2 has gambles with only positive monetary outcomes. Set 3 has gambles with only non-positive monetary outcomes, and one of the two outcomes for each gamble is $0. Set 4 has gambles with only negative monetary outcomes. Set 5 has gambles with a mixture of positive and negative outcomes.

6.4 Probabilistic Specifications

CPT, the lexicographic semiorder model, and the similarity model are all deterministic theories. At the same time, experimental research collects variable choice data. How can one test an algebraic theory using

[1] The study was approved by the Institutional Review Board (IRB) of the University of Illinois under No. 10718.

Table 6.2: The five gamble sets in the 2010 experiment.

	x_1	p_1	x_2	p_2
Set 1	\$0.00	0.23	\$8.75	0.77
	\$0.00	0.33	\$9.00	0.67
	\$0.00	0.37	\$9.50	0.63
	\$0.00	0.39	\$10.50	0.61
	\$0.00	0.40	\$12.50	0.60
Set 2	\$4.75	0.30	\$7.00	0.70
	\$3.75	0.40	\$8.00	0.60
	\$2.75	0.44	\$9.00	0.56
	\$1.75	0.46	\$10.00	0.54
	\$0.75	0.47	\$11.00	0.53
Set 3	-\$2.50	0.27	\$0.00	0.73
	-\$2.75	0.17	\$0.00	0.83
	-\$3.25	0.13	\$0.00	0.87
	-\$4.25	0.11	\$0.00	0.89
	-\$6.25	0.10	\$0.00	0.90
Set 4	-\$5.00	0.29	-\$2.25	0.71
	-\$5.25	0.19	-\$1.75	0.81
	-\$5.75	0.15	-\$1.25	0.85
	-\$6.75	0.13	-\$0.75	0.87
	-\$8.75	0.12	-\$0.25	0.88
Set 5	-\$0.50	0.13	\$5.00	0.87
	-\$1.50	0.23	\$6.00	0.77
	-\$2.50	0.27	\$7.00	0.73
	-\$3.50	0.29	\$8.00	0.71
	-\$4.50	0.30	\$9.00	0.70

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). To perform an appropriate and rigorous test of the deterministic decision theory, researchers have to solve Luce's challenge. However, only a 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. 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 6.4.1 and 6.4.2 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 (Davis-Stober, 2009, Iverson and Falmagne, 1985, Silvapulle and Sen, 2005). 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 decision theory and test those probabilistic models with both frequentist and Bayesian order-constrained statistical methods.

6.4.1 Distance-Based Models

A distance-based 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

$$\begin{array}{lcc} x \succ y & \Leftrightarrow & \theta_{xy} \geq 1 - \tau \\ x \prec y & \Leftrightarrow & \theta_{xy} \leq \tau \\ x \sim y^2 & \Leftrightarrow & \frac{1-\tau}{2} \leq \theta_{xy} \leq \frac{1+\tau}{2} \end{array}$$

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). However, a distance-based model with upper bound $\tau = 0.50$ might be too lenient. To compensate for this, one could place a more restrictive constraint on τ for each binary pair, for example, $\tau = 0.10$, which means that the decision maker chooses the preferred prospect with probability at least 0.10. In this paper, I use three different upper bounds, $\tau = 0.50$, 0.25, and 0.10, on the error rate.

6.4.2 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

[2]The lexicographic semiorder models and the similarity models predict indifferences. The 49 forms of CPT do not predict indifferences.

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 *CPT-KT* as example, a mixture model treats the two parameters, risk attitude α and weighting parameter γ, as random variables with any joint distribution whatsoever, hence permitting all possible probability distributions over the various permissible preference patterns. I write $\succ$ for strict preference. I define $\mathcal{CPT}$ as a set of preference patterns predicted by CPT and $P(\succ_{CPT})$ as the probability of preference pattern $\succ_{CPT}$ in $\mathcal{CPT}$. According to the mixture model, for any pair (x, y), the binary choice probability θ_{xy} is

$$\theta_{xy} = \sum_{\substack{\succ_{CPT} \in \mathcal{CPT} \\ \text{in which } x \succ y}} P(\succ_{CPT}).$$

This equation shows that the probability of choosing x over y equals the total probability of those preference patterns predicted by CPT in which x is strictly preferred to y.

For theories that predict indifferences, like the lexicographic semiorder models, I define the binary choice probability θ_{xy} as the following. 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.

To get the mixture model of a theory, a researcher needs to get the permissible preference patterns predicted by that theory. Here I use *CPT-KT* and Set 1 as an example. When allowing α and γ to be random variables with any joint distribution, I get 11 different preference patterns[3]. The mixture model of *CPT-KT* for Set 1 can be cast geometrically as the convex hull (polytope) of 11 vertices in a suitably chosen ten dimensional unit hypercube of binary choice probabilities. Each vertex encodes the binary choice probabilities when the probability mass is concentrated on a single preference pattern predicted by *CPT-KT*. I provide the minimal description of the mixture polytope of *CPT-KT* for Set 1 in terms of its facet-defining

[3]I used the grid search to get predicted preference patterns. The grid search for α considered all values in the range [0.01, 10] with a step-size of 0.01 and the range [10.05, 50] with a step-size of 0.05. The grid search for γ considered all values in the range [0.279, 1] with a step-size of 0.01.

inequalities, via the public-domain software PORTA[4].

$$P_{bc} \leq P_{ad} \leq P_{ac} \leq P_{ab} \leq 1 \tag{6.1}$$

$$0 \leq P_{de} \leq P_{ce} \leq P_{cd} \leq P_{be} \leq P_{bd} \leq P_{ae} \tag{6.2}$$

This mixture model has ten free parameters that are restricted by Inequalities (1) and (2). In this case, the mixture model is full dimensional. I can test this mixture model with both frequentist and Bayesian order-constrained statistical methods. However, when the mixture model is not full dimensional[5], the frequentist methods do not work. The Bayesian methods, on the other hand, can handle non-full-dimensional models. I provide the minimal descriptions of the mixture models for all of the theories tested in this paper in the Supplemental Materials. I only use the Bayesian order-constrained methods to test the mixture models in this paper.

I computed the predicted pattens[6] and obtained the facet-defining inequalities of the lexicographic semiorder model and the similarity model for Sets 1 - 5, and of the 49 forms of CPT for Sets 1 - 4. Because the gambles in Set 5 have both gains and losses, the parameter number of CPT for Set 5 doubles compared to the parameter number of CPT for gambles with only gains or losses. Due to limitations on computation ability and resources, I selected and computed predicted patterns for six forms of CPT for Set 5[7], based on their performance for Sets 1 - 4 and their parameter numbers for Set 5 (no more than four parameters). Table 6.3 shows different forms of CPT (in each block) that make identical predictions for each stimulus set.

In sum, I test four different probabilistic specifications, the distance-based models with τ =0.50, 0.25, and 0.10 and the mixture model of the lexicographic semiorder model, the similarity model, and the 49 forms of CPT on Sets 1 - 5 (as applicable). Altogether, I test a total of 864 probabilistic models in this paper.

6.4.3 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

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

[5]Some of the mixture models of the 49 forms of CPT are not full dimensional. See Supplemental Materials for more details.

[6]I used the grid search to get predicted preference patterns. The grid search for α and s considered all values in the range [0.01, 10] with a step-size of 0.01 and the range [10.05, 50] with a step-size of 0.05.; the grid search for β and γ considered all values in the range [.01, 1] with a step-size of 0.01; for Tversky-Kahneman probability weighting function, the grid search for γ considered all values in the range [0.279, 1] with a step-size of 0.01.

[7]The six CPT forms are: CPT with the linear utility function with the linear probability weighting function, CPT with the linear utility function and the power probability weighting function, CPT with the linear utility function and Tversky-Kahneman probability weighting function, CPT with the linear utility function and Wu-Gonzalez probability weighting function, CPT with the linear utility function and Prelec II probability weighting function, and CPT with the power utility function and the linear probability weighting function.

Table 6.3: Different CPT forms that make identical predications. The different CPT forms in each block have the same predictions.

Set 1		Set 2		Set 3		Set 4	
$u(x)$	$w(p)$	$u(x)$	$w(p)$	$u(x)$	$w(p)$	$u(x)$	$w(p)$
Lin	Pwr	Lin	Lin	Lin	Pwr	Lin	Lin
Lin	Prl-I	Lin	TK	Lin	TK	Lin	TK
Lin	Prl-II	Lin	Prl-I	Lin	WG	Log	Lin
Log	GE	Log	Lin	Lin	Prl-I	Log	TK
Pwr	Lin	Log	TK	Lin	Prl-II	Expo	Lin
Pwr	Pwr	Log	Prl-I	Log	Pwr	Expo	TK
Pwr	GE	Quad	Lin	Log	TK	Bell	Lin
Pwr	TK	Quad	Prl-I	Log	Prl-I	Bell	TK
Pwr	WG	Expo	Lin	Pwr	Lin	Lin	Pwr
Pwr	Prl-I	Expo	TK	Pwr	Pwr	Lin	GE
Pwr	Prl-II	Expo	Prl-I	Pwr	TK	Lin	WG
Log	Lin	Bell	Lin	Pwr	Prl-I	Lin	Prl-II
Bell	Lin	Bell	TK	Pwr	Prl-II	Pwr	Lin
Log	Pwr	Bell	Prl-I	Log	Lin	Pwr	TK
Log	Prl-I	Log	Pwr	Expo	Lin	Pwr	Prl-I
Bell	Prl-I	Log	GE	Log	GE	Lin	Prl-I
Quad	Pwr	Log	WG	Pwr	GE	Expo	Prl-I
Quad	Prl-I	Log	Prl-II	Bell	Pwr	Log	GE
Expo	Pwr	Pwr	WG	Bell	GE	Log	Prl-II
Expo	WG	Pwr	Prl-II	Quad	Pwr	Expo	Lin
Expo	Prl-I	Bell	GE	Quad	TK	Expo	TK
Expo	GE	Bell	WG	Quad	WG	Bell	Lin
Expo	Prl-II	Hara	TK	Quad	Prl-I	Bell	TK
Hara	Lin	Hara	Prl-I	Hara	Lin	Expo	GE
Hara	Pwr	Hara	WG	Hara	Pwr	Expo	Prl-II
Hara	TK	Hara	Prl-II	Hara	TK	Bell	Pwr
Hara	Prl-I			Hara	Prl-I	Bell	WG
						Hara	Lin
						Hara	Prl-I

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, which permits all conceivable binary choice probabilities and additionally places a uniform prior on them.

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 (Gelfand and Smith, 1990, Myung et al., 2005, Sedransk et al., 1985). I completed all the analyses in this paper 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)[8].

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 cutoffs for "substantial" 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. I will say that a decision model "fits" if its Bayes factor against the saturated model is larger than 3.2. 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 all 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 the individual Bayes factors. The model with the highest GBF is the one that best accounts for all participants' data jointly.

Table 6.4: The total number of people who are fit by the distance-based models using the frequentist analysis for Sets 1 and 2 (gambles with non-negative outcomes). There are 50 participants in the experiment.

			Set 1			Set 2		
Decision Theory			$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$
LSO-Diff			49	41	20	49	46	25
LSO-Ratio			49	40	20	50	47	28
SIM-Diff			45	40	25	49	35	20
SIM-Ratio			45	39	25	50	36	20
CPT	Lin	Lin	4	0	0	24	18	10
	Lin	Pwr	35	24	12	43	31	20
	Lin	GE	42	24	12	47	32	20
	Lin	TK	20	6	0	24	18	10
	Lin	WG	45	25	13	47	32	20
	Lin	Prl-I	35	24	12	24	18	10
	Lin	Prl-II	46	28	14	48	32	20
	Log	Lin	15	2	1	24	18	10
	Log	Pwr	45	26	13	42	31	20
	Log	GE	46	28	14	42	31	20
	Log	TK	26	6	1	24	18	10
	Log	WG	49	33	14	42	31	20
	Log	Prl-I	45	26	13	24	18	10
	Log	Prl-II	49	33	14	42	31	20
	Pwr	Lin	46	28	14	46	32	20
	Pwr	Pwr	46	28	14	47	32	20
	Pwr	GE	46	28	14	49	32	20
	Pwr	TK	46	28	14	48	32	20
	Pwr	WG	46	28	14	49	32	20
	Pwr	Prl-I	46	28	14	47	32	20
	Pwr	Prl-II	46	28	14	49	32	20
	Quad	Lin	16	4	2	24	18	10
	Quad	Pwr	49	33	14	42	31	20
	Quad	GE	49	33	14	42	31	20
	Quad	TK	27	8	2	28	18	10
	Quad	WG	49	33	14	42	31	20
	Quad	Prl-I	49	33	14	24	18	10
	Quad	Prl-II	49	33	14	42	31	20
	Expo	Lin	16	4	2	24	18	10
	Expo	Pwr	49	33	14	42	31	20
	Expo	GE	49	33	14	42	31	20
	Expo	TK	31	10	2	24	18	10
	Expo	WG	49	33	14	42	31	20
	Expo	Prl-I	49	33	14	24	18	10
	Expo	Prl-II	49	33	14	42	31	20
	Bell	Lin	15	2	1	24	18	10
	Bell	Pwr	45	26	13	44	32	20
	Bell	GE	46	28	14	49	33	20
	Bell	TK	30	8	1	24	18	10
	Bell	WG	49	33	14	49	33	20
	Bell	Prl-I	45	26	13	24	18	10
	Bell	Prl-II	49	33	14	50	34	20
	HARA	Lin	7	3	1	45	32	20
	HARA	Pwr	7	3	1	50	34	20
	HARA	GE	32	20	13	50	34	20
	HARA	TK	7	3	1	45	32	20
	HARA	WG	32	20	13	50	34	20
	HARA	Prl-I	7	3	1	45	32	20
	HARA	Prl-II	32	20	13	50	34	20

Table 6.5: The total number of people who are fit by the distance-based models using the frequentist analysis for Sets 3 and 4 (gambles with non-positive outcomes). There are 50 participants in the experiment.

			Set 3			Set 4		
Decision Theory			$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$
LSO-Diff			50	47	20	50	47	25
LSO-Ratio			49	44	19	50	45	25
SIM-Diff			49	46	21	50	22	15
SIM-Ratio			49	43	19	45	23	13
CPT	Lin	Lin	10	1	0	38	23	13
	Lin	Pwr	41	29	6	49	33	15
	Lin	GE	42	30	6	49	33	15
	Lin	TK	41	29	6	38	23	13
	Lin	WG	41	29	6	49	33	15
	Lin	Prl-I	41	29	6	47	33	15
	Lin	Prl-II	46	33	7	49	33	15
	Log	Lin	14	4	1	38	23	13
	Log	Pwr	46	33	7	50	33	15
	Log	GE	46	33	7	50	33	15
	Log	TK	46	33	7	38	23	13
	Log	WG	46	33	7	50	33	15
	Log	Prl-I	46	33	7	45	33	15
	Log	Prl-II	48	35	7	50	33	15
	Pwr	Lin	46	33	7	49	33	15
	Pwr	Pwr	46	33	7	50	33	15
	Pwr	GE	46	33	7	50	33	15
	Pwr	TK	46	33	7	49	33	15
	Pwr	WG	48	35	7	50	33	15
	Pwr	Prl-I	46	33	7	49	33	15
	Pwr	Prl-II	46	33	7	50	33	15
	Quad	Lin	14	4	1	38	23	13
	Quad	Pwr	48	35	7	50	33	15
	Quad	GE	48	35	7	50	33	15
	Quad	TK	48	35	7	38	23	13
	Quad	WG	48	35	7	50	33	15
	Quad	Prl-I	48	35	7	45	33	15
	Quad	Prl-II	48	35	7	50	33	15
	Expo	Lin	14	4	1	38	23	13
	Expo	Pwr	48	35	7	50	33	15
	Expo	GE	48	35	7	50	33	15
	Expo	TK	48	35	7	38	23	13
	Expo	WG	48	35	7	50	33	15
	Expo	Prl-I	48	35	7	47	33	15
	Expo	Prl-II	48	35	7	50	33	15
	Bell	Lin	14	4	1	38	23	13
	Bell	Pwr	46	33	7	50	33	15
	Bell	GE	46	33	7	50	33	15
	Bell	TK	48	35	7	38	23	13
	Bell	WG	48	35	7	50	33	15
	Bell	Prl-I	48	35	7	48	33	15
	Bell	Prl-II	48	35	7	50	33	15
	HARA	Lin	36	24	7	41	23	13
	HARA	Pwr	36	24	7	41	23	13
	HARA	GE	36	24	7	41	23	13
	HARA	TK	36	24	7	41	23	13
	HARA	WG	36	24	7	41	23	13
	HARA	Prl-I	36	24	7	41	23	13
	HARA	Prl-II	36	24	7	41	23	13

Table 6.6: The total number of people who are fit by the distance-based models using the frequentist analysis for Set 5 (gambles with a mixture of gains and losses). There are 50 participants in the experiment.

			Set 5		
Decision Theory			$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$
LSO-Diff			50	47	30
LSO-Ratio			50	47	30
SIM-Diff			50	38	23
SIM-Ratio			50	38	23
CPT	Lin	Lin	6	0	0
	Lin	Pwr	48	37	23
	Lin	TK	49	38	23
	Lin	WG	50	38	23
	Lin	Prl-II	50	38	23
	Pwr	Lin	50	38	23

6.5 Results

6.5.1 The Distance-Based Models

Tables 6.4 - 6.6 summarize the results for the distance-based models using frequentist methods for all five stimulus sets. The first three columns display the decision theories, including two kinds of lexicographic semiorder model and the similarity model, and the utility function and probability weighting function of the 49 forms of CPT. The following columns report the total number of people who are fit by the distance-based models with τ =0.50, 0.25, and 0.10 for each stimulus set.

Tables 6.4 - 6.6 show 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. The frequentist analyses show that most of the distance-based models of CPT with $\tau = 0.50$ explain almost all of the participants' data for every stimulus set. When I put higher restrictions on the error rate bound, for example, $\tau = 0.10$, the number of fits decreases a lot compared to when $\tau = 0.50$. The frequentist analysis of distance-based model shows that the numbers of fits are similar for different stimulus sets, except for the 49 forms of CPT for Set 3 when $\tau = 0.10$. For the distance-based models of CPT, the number of fits for Set 3 is much smaller than the other stimulus sets when $\tau = 0.10$. Overall, for all three error rate bounds, the intransitive heuristics explain equal or more participants' data than the 49 forms of CPT.

Tables 6.7 - 6.9 summarize the results for the distance-based models using the Bayes factor analysis for all five stimulus sets. I find a close alignment between frequentist and Bayesian results. In sum, for all three error rate bounds, the intransitive heuristics explain more participants' data than the 49 forms of CPT for Sets 1 - 4, and the intransitive heuristics and CPT have a similar number of fits for Set 5.

Table 6.10 summarizes the number of people who are fit consistently and simultaneously across different stimulus sets by the distance-based models with three different values of τ, using the frequentist and Bayes

[8]The analyses were supported by XSEDE Grant NSF SES No. 130016 (PI: Michel Regenwetter).

Table 6.7: The total number of people who are fit by the distance-based models using the Bayesian analysis for Sets 1 and 2 (gambles with non-negative outcomes). There are 50 participants in the experiment.

			Set 1			Set 2		
Decision Theory			$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$
LSO-Diff			46	36	21	48	43	25
LSO-Ratio			46	35	21	25	38	22
SIM-Diff			38	41	31	48	46	24
SIM-Ratio			37	40	31	29	42	22
CPT	Lin	Lin	4	1	0	25	17	11
	Lin	Pwr	34	23	15	41	29	21
	Lin	GE	39	24	15	41	29	21
	Lin	TK	18	5	1	25	17	11
	Lin	WG	42	25	16	41	29	21
	Lin	Prl-I	34	23	15	25	17	11
	Lin	Prl-II	44	28	17	41	29	21
	Log	Lin	14	4	1	25	17	11
	Log	Pwr	44	26	16	38	29	21
	Log	GE	44	28	17	38	29	21
	Log	TK	24	7	1	25	17	11
	Log	WG	49	32	17	38	29	21
	Log	Prl-I	44	26	16	25	17	11
	Log	Prl-II	49	32	17	38	29	21
	Pwr	Lin	44	28	17	41	29	21
	Pwr	Pwr	44	28	17	41	29	21
	Pwr	GE	44	28	17	40	29	20
	Pwr	TK	44	28	17	44	29	21
	Pwr	WG	44	28	17	40	29	20
	Pwr	Prl-I	44	28	17	41	29	21
	Pwr	Prl-II	44	28	17	40	29	20
	Quad	Lin	14	6	2	25	17	11
	Quad	Pwr	48	31	16	38	29	21
	Quad	GE	48	31	16	38	29	21
	Quad	TK	19	9	2	22	16	11
	Quad	WG	48	31	16	38	29	21
	Quad	Prl-I	48	31	16	25	17	11
	Quad	Prl-II	48	31	16	38	29	21
	Expo	Lin	14	6	2	25	17	11
	Expo	Pwr	49	33	17	38	29	21
	Expo	GE	49	33	17	37	29	21
	Expo	TK	27	10	3	25	17	11
	Expo	WG	49	33	17	37	29	21
	Expo	Prl-I	49	33	17	25	17	11
	Expo	Prl-II	49	33	17	37	29	21
	Bell	Lin	14	4	1	25	17	11
	Bell	Pwr	44	26	16	38	30	20
	Bell	GE	44	28	17	40	30	20
	Bell	TK	29	8	2	25	17	11
	Bell	WG	49	33	17	40	30	20
	Bell	Prl-I	44	26	16	25	17	11
	Bell	Prl-II	49	33	17	41	30	20
	HARA	Lin	4	3	1	39	30	21
	HARA	Pwr	4	3	1	47	30	20
	HARA	GE	28	20	14	47	30	20
	HARA	TK	4	3	1	39	30	21
	HARA	WG	28	19	14	47	30	20
	HARA	Prl-I	4	3	1	39	30	21
	HARA	Prl-II	28	19	14	47	30	20

Table 6.8: The total number of people who are fit by the distance-based models using the Bayesian analysis for Sets 3 and 4 (gambles with non-positive outcomes). There are 50 participants in the experiment.

			Set 3			Set 4		
Decision Theory			$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$
LSO-Diff			49	37	19	49	37	22
LSO-Ratio			46	34	18	40	35	22
SIM-Diff			44	37	21	34	41	25
SIM-Ratio			40	35	19	27	39	23
CPT	Lin	Lin	10	1	0	36	25	16
	Lin	Pwr	40	23	14	46	31	19
	Lin	GE	41	24	12	46	31	19
	Lin	TK	40	23	14	36	25	16
	Lin	WG	40	23	14	46	31	19
	Lin	Prl-I	40	23	14	42	32	19
	Lin	Prl-II	45	27	14	46	31	19
	Log	Lin	12	4	2	36	25	16
	Log	Pwr	45	27	14	48	31	19
	Log	GE	44	27	14	47	31	19
	Log	TK	45	27	14	36	25	16
	Log	WG	44	27	14	49	31	19
	Log	Prl-I	45	27	14	42	32	19
	Log	Prl-II	47	29	14	47	31	19
	Pwr	Lin	45	27	14	46	31	19
	Pwr	Pwr	45	27	14	49	31	19
	Pwr	GE	44	27	14	47	31	19
	Pwr	TK	45	27	14	46	31	19
	Pwr	WG	44	27	11	47	31	19
	Pwr	Prl-I	45	27	14	46	31	19
	Pwr	Prl-II	45	27	14	47	31	19
	Quad	Lin	12	4	2	33	25	15
	Quad	Pwr	45	29	11	47	29	18
	Quad	GE	44	29	11	46	29	18
	Quad	TK	45	29	11	33	25	15
	Quad	WG	45	29	11	46	29	18
	Quad	Prl-I	45	29	11	39	31	19
	Quad	Prl-II	44	29	11	45	29	18
	Expo	Lin	12	4	2	36	25	16
	Expo	Pwr	46	29	12	47	30	19
	Expo	GE	45	29	11	47	29	18
	Expo	TK	47	29	14	36	25	16
	Expo	WG	44	29	11	47	29	18
	Expo	Prl-I	46	29	13	42	32	19
	Expo	Prl-II	44	29	11	47	29	18
	Bell	Lin	12	4	2	36	25	16
	Bell	Pwr	44	27	14	46	29	18
	Bell	GE	44	27	14	45	29	18
	Bell	TK	47	29	15	36	25	16
	Bell	WG	47	29	12	46	29	18
	Bell	Prl-I	47	29	15	44	32	19
	Bell	Prl-II	47	29	14	45	29	18
	HARA	Lin	31	19	13	34	23	15
	HARA	Pwr	31	19	13	33	23	15
	HARA	GE	31	19	13	30	23	15
	HARA	TK	31	19	13	33	23	15
	HARA	WG	30	19	13	32	23	15
	HARA	Prl-I	31	19	13	34	23	15
	HARA	Prl-II	30	19	13	30	23	15

Table 6.9: The total number of people who are fit by the distance-based models using the Bayesian analysis for Set 5 (gambles with a mixture of gains and losses). There are 50 participants in the experiment.

			Set 5		
Decision Theory			$\tau = 0.50$	$\tau = 0.25$	$\tau = 0.10$
LSO-Diff			49	41	29
LSO-Ratio			28	37	24
SIM-Diff			49	42	28
SIM-Ratio			32	45	27
CPT	Lin	Lin	7	0	0
	Lin	Pwr	48	31	26
	Lin	TK	49	32	27
	Lin	WG	45	32	24
	Lin	Prl-II	45	32	24
	Pwr	Lin	47	32	25

factor analysis. For the frequentist analysis, I report the total number of people who are consistently fit across different stimulus sets by the distance-based models. By a consistent fit of the distance-based model, I mean that there exists a set of parameter values of decision theories for which the distance-based model fits in different stimulus sets. For the Bayes factor analysis, I report the total number of people who are simultaneously fit across different stimulus sets by the distance-based models. The column labeled 'Gains' represents fits across Sets 1 and 2, which have only non-negative monetary outcomes; the column labeled 'Losses' represents fits across Sets 3 and 4, which have only non-positive monetary outcomes; the column labeled 'All' represents fits across Sets 1 - 4[9].

Results in Tables 6.4 - 6.10 show that there might be some degree of 'over-fitting' for the distance-based models. For both the frequentist and Bayes factor analysis, the number of participants who replicate across different stimulus sets is much smaller than the number of participants who are fit in separate stimulus set. For the frequentist analysis, it means that when a model fits the data of some participants in one stimulus set, the estimated best-fitting parameters of that model need not predict the data of the same participants in another stimulus set. For the Bayes factor analysis, it means that the model cannot explain the data of a participant simultaneously across different stimulus sets, even though the model might explain the data well for some stimulus sets separately. Again, the results of the frequentist and Bayes factor analysis are in close alignment with each other.

In sum, the results show that the distance-based models of intransitive heuristics fit more participants' data than the 49 forms of CPT for each stimulus set, no matter whether using the frequentist or the Bayes factor analysis.

[9]I exclude Set 5 in this analysis, since the predicted patterns for most forms of CPT are not computed. Also, because CPT has different parameters for gains and losses, I only consider consistent fits separately for gains and losses for the frequentist analysis.

Table 6.10: The number of participants who are consistently (using frequentist analysis) and simultaneously (using Bayes factor analysis) fit across different stimulus sets by the distance-based models. The 'Gains' column represents fits across Sets 1 and 2; the 'Losses' column represents fits across Sets 3 and 4; and the 'All' column represents fits across Sets 1 - 4. There are 50 participants in the experiment.

		Frequentist		Bayes Factor					Frequentist		Bayes Factor		
Theory	τ	Gains	Losses	Gains	Losses	All	Theory	τ	Gains	Losses	Gains	Losses	All
LS-Diff	0.50	46	46	45	48	43	Quad GE	0.50	41	0	36	43	32
	0.25	29	28	32	27	17		0.25	21	0	20	19	10
	0.10	15	10	16	11	9		0.10	10	0	12	5	3
LS-Ratio	0.50	43	40	23	37	19	Quad TK	0.50	20	37	14	32	9
	0.25	17	10	27	22	14		0.25	5	16	6	13	2
	0.10	9	4	15	10	8		0.10	2	3	2	4	0
SIM-Diff	0.50	37	22	30	38	25	Quad WG	0.50	41	0	36	44	32
	0.25	17	8	23	20	11		0.25	21	0	20	19	10
	0.10	9	4	15	8	6		0.10	10	0	12	5	3
SIM-Ratio	0.50	35	25	26	26	18	Quad Prl-I	0.50	23	44	23	38	16
	0.25	15	8	20	13	9		0.25	8	25	8	20	5
	0.10	9	4	13	6	5		0.10	2	3	3	6	0
Lin Lin	0.50	2	16	3	4	0	Quad Prl-II	0.50	41	0	36	42	31
	0.25	0	7	1	1	0		0.25	21	0	20	19	10
	0.10	0	3	0	0	0		0.10	10	0	12	5	3
Lin Pwr	0.50	29	41	26	38	22	Expo Lin	0.50	13	14	14	11	6
	0.25	15	19	15	14	6		0.25	4	4	6	4	1
	0.10	8	2	10	5	4		0.10	2	1	2	2	0
Lin GE	0.50	40	0	31	39	27	Expo Pwr	0.50	41	48	37	46	35
	0.25	16	0	16	15	7		0.25	21	25	21	20	10
	0.10	8	0	10	4	4		0.10	10	3	12	6	4
Lin TK	0.50	10	31	10	28	5	Expo GE	0.50	41	0	36	45	33
	0.25	1	12	2	8	0		0.25	21	0	21	19	10
	0.10	0	2	0	3	0		0.10	10	0	12	5	3
Lin WG	0.50	43	0	34	38	26	Expo TK	0.50	21	37	19	35	12
	0.25	17	0	17	14	8		0.25	5	16	7	13	3
	0.10	9	0	11	5	4		0.10	2	3	2	4	0
Lin Prl-I	0.50	10	39	10	33	5	Expo WG	0.50	41	0	36	44	33
	0.25	2	19	3	14	0		0.25	21	0	21	19	10
	0.10	0	2	1	5	0		0.10	10	0	12	5	3
Lin Prl-II	0.50	45	0	36	43	32	Expo Prl-I	0.50	23	46	24	41	19
	0.25	20	0	20	18	9		0.25	8	25	9	20	5
	0.10	10	0	12	6	4		0.10	2	3	3	6	0
Log Lin	0.50	12	14	13	11	6	Expo Prl-II	0.50	41	0	36	44	33
	0.25	2	4	4	4	1		0.25	21	0	21	19	10
	0.10	1	1	1	2	0		0.10	10	0	12	5	3
Log Pwr	0.50	38	46	34	44	32	Bell Lin	0.50	12	14	13	11	6
	0.25	17	23	18	18	9		0.25	2	4	4	4	1
	0.10	9	3	11	6	4		0.10	1	1	1	2	0
Log GE	0.50	39	0	34	43	32	Bell Pwr	0.50	40	0	34	42	31
	0.25	19	0	20	18	9		0.25	18	0	18	17	9
	0.10	10	0	12	6	4		0.10	9	0	11	5	3

Continued on next page

Table 6.10 – Continued from previous page

Theory		τ	Frequentist Gains	Frequentist Losses	Bayes Factor Gains	Bayes Factor Losses	Bayes Factor All	Theory		τ	Frequentist Gains	Frequentist Losses	Bayes Factor Gains	Bayes Factor Losses	Bayes Factor All
Log	TK	0.50	18	36	18	34	11	Bell	GE	0.50	45	0	36	41	30
		0.25	3	16	5	12	3			0.25	21	0	20	17	9
		0.10	1	3	1	4	0			0.10	10	0	12	5	3
Log	WG	0.50	41	0	37	43	35	Bell	TK	0.50	20	0	21	35	14
		0.25	21	0	20	18	10			0.25	3	0	5	13	3
		0.10	10	0	12	6	4			0.10	1	0	1	4	0
Log	Prl-I	0.50	20	42	20	39	15	Bell	WG	0.50	48	0	39	45	35
		0.25	4	23	6	18	4			0.25	23	0	21	19	10
		0.10	1	3	2	6	0			0.10	10	0	12	5	3
Log	Prl-II	0.50	41	0	37	46	35	Bell	Prl-I	0.50	20	0	20	42	15
		0.25	21	0	20	20	10			0.25	4	0	6	20	4
		0.10	10	0	12	6	4			0.10	1	0	2	7	0
Pwr	Lin	0.50	43	46	36	43	32	Bell	Prl-II	0.50	49	0	40	44	34
		0.25	20	23	20	18	9			0.25	24	0	21	19	10
		0.10	10	3	12	6	4			0.10	10	0	12	5	3
Pwr	Pwr	0.50	44	46	36	44	33	Hara	Lin	0.50	7	32	4	22	2
		0.25	20	23	20	18	9			0.25	3	12	3	9	1
		0.10	10	3	12	6	4			0.10	1	3	1	4	0
Pwr	GE	0.50	45	0	35	43	33	Hara	Pwr	0.50	7	0	4	22	2
		0.25	20	0	20	18	9			0.25	3	0	3	9	1
		0.10	10	0	12	6	4			0.10	1	0	1	4	0
Pwr	TK	0.50	44	46	38	43	33	Hara	GE	0.50	32	0	26	19	15
		0.25	20	23	20	18	9			0.25	17	0	17	9	6
		0.10	10	3	12	6	4			0.10	9	0	10	4	2
Pwr	WG	0.50	45	0	35	44	32	Hara	TK	0.50	7	0	4	22	2
		0.25	20	0	20	18	9			0.25	3	0	3	9	1
		0.10	10	0	12	6	4			0.10	1	0	1	4	0
Pwr	Prl-I	0.50	44	46	36	43	32	Hara	WG	0.50	32	0	26	21	16
		0.25	20	23	20	18	9			0.25	17	0	16	9	6
		0.10	10	3	12	6	4			0.10	9	0	10	4	2
Pwr	Prl-II	0.50	45	0	35	44	33	Hara	Prl-I	0.50	7	0	4	22	2
		0.25	20	0	20	18	9			0.25	3	0	3	9	1
		0.10	10	0	12	6	4			0.10	1	0	1	4	0
Quad	Lin	0.50	13	14	14	11	6	Hara	Prl-II	0.50	32	0	26	19	15
		0.25	4	4	6	4	1			0.25	17	0	16	9	6
		0.10	2	1	2	2	0			0.10	9	0	10	4	2
Quad	Pwr	0.50	41	48	36	45	33								
		0.25	21	25	20	19	10								
		0.10	10	3	12	5	3								

6.5.2 Mixture Model

Table 6.11 shows the number of people who are fit by the mixture model of the lexicographic semiorder, the similarity model, and the 49 forms of CPT by the Bayes factor analysis for Sets 1 - 5, separately for each stimulus set and simultaneously across all stimulus sets. The first column shows the name of each theory; Columns 2 - 6 shows the number of people who are fit by the mixture model for each stimulus set; Column 7 represents the number of simultaneous fit across Sets 1 and 2; Column 8 represents the number of simultaneous fit across Sets 3 and 4; and Column 9 represents the number of simultaneous fit across Sets 1 - 4. The cell with "-" means that the mixture model is not available for that stimulus set.

The mixture models of the lexicographic semiorder and the similarity models explain at most 52% of the participants' data for different stimulus sets. The mixture models of intransitive heuristics explain at most 14 out of 50 participants simultaneously for gains, at most five for losses, and at most two across five stimulus sets, suggesting that these models could not explain the participants' data very well simultaneously across different stimulus sets.

The 49 forms of CPT differ a lot in terms of their numbers of fit for different stimulus sets. The numbers of fit for different CPT are similar for every stimulus set, except for Set 2, of which the number of fits is much smaller. The mixture model of CPT that perform the best for Set 1 is CPT with the exponential utility function and Prelec I probability weighting function, which explains 34 out of 50 participants' data; for Set 2, CPT with Hara utility function and Goldstein-Einhorn probability weighting function, which explains 35 out of 50 participants' data; for Set 3, CPT with the power utility function and Wu-Gonzalez probability weighting function, which explains 36 out of 50 participants' data; for Set 4, CPT with Bell utility function and the power probability weighting function, which explains 44 out of 50 participants' data. No mixture models of one single form of CPT could explain the data of the most participants well across all stimulus sets.

Most of the CPT forms fail to win over the saturated model for all of the participants simultaneously across Sets 1 and 2. The mixture model of CPT with Hara utility function and Prelec II probability weighting function explains the most participants' data (23 out of 50) simultaneously across Sets 1 and 2. There are more mixture models of CPT that could explain more than half of the participants' data simultaneously across Sets 3 and 4. The mixture model of CPT with the exponential utility function and the power probability weighting function explains the most participants' data (29 out of 50) simultaneously across Sets 3 and 4. The mixture model of CPT with Hara utility function and Prelec II probability weighting function explains the most participants' data (15 out of 50) simultaneously across all five stimulus sets. Most mixture models of CPT could not explain any participants' data simultaneously across Sets 1 - 4.

Table 6.11: The number of people who are fit by the mixture model using the Bayes factor analysis.

Theory			Set 1	Set 2	Set 3	Set 4	Set 5	Gains	Losses	All
LSO-Diff			23	25	2	10	26	10	0	0
LSO-Ratio			26	30	11	20	22	14	5	2
SIM-Diff			26	20	13	21	19	10	5	2
SIM-Ratio			25	9	14	14	16	2	5	2
	Lin	Lin	-	-	-	-	-	-	-	-
	Lin	Pwr	0	0	0	25	8	0	0	0
	Lin	GE	0	3	0	24	-	0	0	0
	Lin	TK	0	-	0	-	10	-	-	-
	Lin	WG	0	3	0	25	31	0	0	0
	Lin	Prl-I	0	-	0	0	-	-	0	-
	Lin	Prl-II	24	6	30	26	31	4	17	0
	Log	Lin	0	-	0	-	-	-	-	-
	Log	Pwr	0	2	31	26	-	0	18	0
	Log	GE	25	1	34	23	-	0	18	0
	Log	TK	0	-	32	-	-	-	-	-
	Log	WG	30	2	32	23	-	0	14	0
	Log	Prl-I	0	-	28	0	-	-	0	-
	Log	Prl-II	32	2	35	24	-	0	18	0
	Pwr	Lin	22	0	30	25	35	0	17	0
	Pwr	Pwr	23	15	29	23	-	6	14	1
	Pwr	GE	26	27	31	23	-	14	16	3
	Pwr	TK	24	8	32	23	-	4	18	1
	Pwr	WG	23	31	36	22	-	14	17	5
	Pwr	Prl-I	23	5	31	27	-	3	19	0
	Pwr	Prl-II	24	28	31	25	-	15	17	5
	Quad	Lin	0	-	0	0	-	-	0	-
	Quad	Pwr	27	1	28	42	-	0	25	0
	Quad	GE	30	3	31	41	-	1	26	0
CPT	Quad	TK	0	0	28	0	-	0	0	0
	Quad	WG	27	1	26	43	-	0	24	0
	Quad	Prl-I	28	-	27	0	-	-	0	-
	Quad	Prl-II	29	2	30	41	-	0	26	0
	Expo	Lin	0	-	0	-	-	-	-	-
	Expo	Pwr	31	3	33	43	-	0	29	0
	Expo	GE	32	1	32	41	-	0	26	0
	Expo	TK	0	-	32	-	-	-	-	-
	Expo	WG	31	2	30	42	-	1	27	0
	Expo	Prl-I	34	-	31	0	-	-	0	-
	Expo	Prl-II	32	2	32	40	-	0	25	0
	Bell	Lin	0	-	0	-	-	-	-	-
	Bell	Pwr	0	2	30	44	-	0	27	0
	Bell	GE	25	6	31	39	-	2	22	1
	Bell	TK	0	-	33	-	-	-	-	-
	Bell	WG	33	7	27	43	-	3	23	3
	Bell	Prl-I	0	-	31	0	-	-	0	-
	Bell	Prl-II	33	33	35	40	-	23	27	15
	HARA	Lin	0	3	0	0	-	0	0	0
	HARA	Pwr	0	29	0	2	-	0	0	0
	HARA	GE	0	35	0	0	-	0	0	0
	HARA	TK	0	3	0	1	-	0	0	0
	HARA	WG	0	34	0	3	-	0	0	0
	HARA	Prl-I	0	3	0	1	-	0	0	0
	HARA	Prl-II	0	33	0	1	-	0	0	0

Overall, there are a lot of variations in terms of model performance within each stimulus set and across different stimulus sets.

6.5.3 Model Comparison: Individual Level

I use Bayes factors to compare models. As I discuss in Section 6.4.3, 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 every stimulus set.

Tables 6.12 - 6.16 report the results of model comparison by Bayes factor for Sets 1 - 5. The first column shows the participant ID. The second column shows the best probabilistic model with the decision theory and the stochastic form. I use the value of the upper bound τ for the distance-based model and "Mix' to represent the mixture model. The third column shows the Bayes factor for the best model compared to the saturated model. The fourth and fifth columns show the second best probabilistic model and its Bayes factor. The last column shows the Bayes factor between the best and second-best models.

For Set 1, the best models are transitive for 34 out of 50 participants. Out of these 34 transitive models, 18 models are the distance-based models of CPT with the linear utility function and the power probability weighting function[10]. The best models are the mixture models for 17 out of 50 participants and are the distance-based models for 33 participants.

For Set 2, the best models are transitive for 39 out of 50 participants. Out of these 39 transitive models, 18 models are the distance-based models of CPT with the linear utility function and the linear probability weighting function[10], which is equivalent to the expected value theory. The best models are the mixture models for 12 out of 50 participants and are the distance-based models for 38 participants.

For Set 3, the best models are transitive for 39 out of 50 participants. Out of these 39 transitive models, 20 models are the distance-based models of CPT with the linear utility function and the power probability weighting function[10]. The best models are the mixture models for 19 out of 50 participants and are the distance-based models for 31 out of 50 participants.

For Set 4, the best models are transitive for 46 out of 50 participants. Out of these 46 transitive models, 22 models are the distance-based models of CPT with the linear utility function and the linear probability weighting function[10], which is equivalent to the expected value theory. The best models are the mixture models for 12 out of 50 participants and are the distance-based models for 38 participants.

For Set 5, the best models are transitive for 43 out of 50 participants. Out of these 43 transitive models, 31 models are the distance-based models of CPT with the linear utility function and the power probability

[10]See Table 6.3 for other CPT forms that make the same predictions.

weighting function. The best models are the mixture models for 10 out of 50 participants and are the distance-based models for 40 participants.

For Sets 2 and 4, which involve only positive or negative outcomes, the distance-based model of CPT that makes a single predicted pattern (for example, CPT with the linear utility function and the linear probability weighting function) wins out for the most participants. Moreover, CPT with the linear utility function and the linear probability weighting function is equivalent to computing the expect value of a gamble. For Sets 1, 3, and 5, the distance-based model of CPT with the linear utility function and the power probability weighting function wins out for the most participants.

Overall, regarding the core decision theory, there are more transitive theories, CPT, winning out than the intransitive heuristics for Sets 1 - 5. Regarding the probability specification, there are more distance-based models winning out than mixture models for Sets 1 - 5. There is a lot of evidence for heterogeneity across individuals and stimulus sets in terms of the best model. No single decision theory, or type of probabilistic specifications, is robust across all participants and stimulus sets.

6.5.4 Model Comparison: Group Level

I use group Bayes factor (GBF, Stephan et al., 2007) to select among models at the group level. The GBF aggregates likelihoods across participants and is the product of individual-level Bayes factors. 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. Table 6.17 shows the top ten models ranked by GBF for every stimulus set. The top panel shows the top ten models for each stimulus set and the bottom panel shows the log10 value of GBF for the corresponding model. For the top panel, the first column shows the ranking of a model; the second to sixth columns show the name of the probabilistic model and its stochastic form.

The distance-based models of the intransitive heuristics rank among top ten for Sets 1 - 4, especially for Set 2, the top four models are all distance-based models of the intransitive heuristics. For example, the distance-based model of LSO-Diff with $\tau = 0.25$ ranks the 10th for Set 1, the 9th for Set 2, the 4th for Set 3, the 1st for Set 4. For Set 5, the top ten models are all probabilistic models of CPT. There does not seem to be one form of CPT that ranks consistently among top ten across all stimulus sets.

For Sets 2 and 4, none of the distance-based models of CPT that makes a single predicted pattern rank among the top ten by GBF, even though these models explain the data of the most participants in those two stimulus sets. This result means that the distance-based models of those CPT could fit for a lot of individual participants in a given stimulus set, but they could not fit for some other participants at all in the same stimulus set. With these substantial individual differences, these CPT models could not account for all participants' data jointly.

Table 6.12: The best and second-best models and their Bayes factors for Set 1.

ID	Best Theory	Best BF	Sec. Theory	Sec. BF	Sec./Best
1	SIM-Diff-Mix	272	SIM-Ratio-Mix	270	1
2	Lin-Pwr-0.10	1019813	Lin-GE-0.10	728438	1
3	Lin-Pwr-0.25	6401	Lin-GE-0.25	4572	1
4	LSO-Diff-0.10	6134	LSO-Ratio-0.10	6134	1
5	Lin-TK-0.25	2334	Lin-Pwr-0.25	1400	2
6	Lin-Pwr-0.10	555095	Lin-GE-0.10	396496	1
7	Pwr-Pwr-Mix	5381	Pwr-Prl-I-Mix	5363	1
8	LSO-Diff-0.25	375	LSO-Ratio-0.25	375	1
9	Lin-Pwr-0.10	169050506	Lin-GE-0.10	120750361	1
10	SIM-Diff-Mix	2706	Expo-Pwr-Mix	2534	1
11	Lin-Pwr-0.50	87	SIM-Ratio-Mix	86	1
12	Lin-Pwr-0.25	60026	Lin-GE-0.25	42876	1
13	Expo-WG-Mix	3244	Bell-Prl-II-Mix	2732	1
14	Lin-Pwr-0.25	65449	Lin-GE-0.25	46750	1
15	SIM-Diff-0.10	70441	SIM-Diff-0.25	613	115
16	Lin-Pwr-0.25	760	Lin-GE-0.25	543	1
17	Quad-Prl-I-Mix	67	SIM-Ratio-Mix	66	1
18	Lin-Pwr-0.25	5372	Lin-GE-0.25	3837	1
19	Lin-Pwr-0.25	397	Lin-GE-0.25	284	1
20	Expo-Prl-II-Mix	80	Expo-GE-Mix	73	1
21	Expo-Prl-II-Mix	4990	Expo-GE-Mix	4753	1
22	SIM-Diff-0.10	242640	LSO-Ratio-0.10	240462	1
23	Quad-GE-Mix	202	Quad-Pwr-Mix	178	1
24	Lin-Pwr-0.50	102	Lin-GE-0.50	73	1
25	Pwr-Prl-I-Mix	8260	Pwr-TK-Mix	8241	1
26	Lin-Pwr-0.10	627717597	Lin-GE-0.10	448369712	1
27	Lin-Pwr-0.10	597923	Lin-GE-0.10	427088	1
28	Expo-Prl-II-Mix	96	Expo-GE-Mix	96	1
29	Bell-Prl-II-Mix	3588	Bell-WG-Mix	2410	1
30	LSO-Diff-0.25	153	LSO-Ratio-0.25	153	1
31	Expo-Pwr-0.50	60	Bell-Prl-II-0.50	56	1
32	SIM-Diff-0.10	17216	LSO-Diff-0.10	17061	1
33	Expo-WG-Mix	22959	Pwr-Lin-Mix	22554	1
34	Expo-Lin-0.10	230945490	Expo-TK-0.10	179624270	1
35	Lin-Pwr-0.10	6829808	Lin-GE-0.10	4878434	1
36	LSO-Diff-Mix	842	LSO-Ratio-Mix	791	1
37	Expo-WG-Mix	189	Expo-Pwr-Mix	187	1
38	SIM-Ratio-Mix	690	SIM-Diff-Mix	638	1
39	Lin-Pwr-0.10	92928	Lin-Pwr-0.25	66783	1
40	Lin-TK-0.25	206	Lin-TK-0.50	168	1
41	SIM-Diff-0.25	77	Expo-Pwr-0.50	66	1
42	LSO-Ratio-0.25	297	LSO-Diff-0.25	297	1
43	Lin-Pwr-0.10	447764073	Lin-GE-0.10	319831481	1
44	SIM-Diff-0.10	348	LSO-Ratio-0.10	344	1
45	Lin-Pwr-0.25	24961	Lin-GE-0.25	17829	1
46	log-Lin-0.25	19662	log-TK-0.25	16854	1
47	Lin-Pwr-0.10	275305	Lin-GE-0.10	196647	1
48	SIM-Diff-0.10	2555	SIM-Diff-0.25	1092	2
49	SIM-Ratio-Mix	131	SIM-Diff-Mix	118	1
50	SIM-Diff-Mix	2136	Expo-Prl-I-Mix	1160	2

Table 6.13: The best and second-best models and their Bayes factors for Set 2.

ID	Best Theory	Best BF	Sec. Theory	Sec. BF	Sec./Best
1	SIM-Diff-0.50	97	HARA-GE-Mix	88	1
2	HARA-Pwr-Mix	55	HARA-GE-Mix	48	1
3	Lin-Pwr-0.25	19086	Pwr-Lin-0.25	10604	2
4	Lin-Lin-0.10	630476240	Lin-Pwr-0.10	126095248	5
5	SIM-Diff-0.25	83	LSO-Diff-0.25	71	1
6	Pwr-Prl-I-Mix	22846	Pwr-TK-Mix	11646	2
7	SIM-Diff-Mix	42	LSO-Diff-Mix	36	1
8	LSO-Ratio-0.10	24129	Lin-Lin-0.25	11841	2
9	Lin-Pwr-0.10	177579	Pwr-Lin-0.10	98655	2
10	LSO-Diff-Mix	917	Pwr-TK-Mix	372	2
11	Pwr-WG-Mix	64	Pwr-Prl-II-Mix	64	1
12	Lin-Pwr-0.25	53427	Pwr-Lin-0.25	29682	2
13	HARA-GE-Mix	105	HARA-WG-Mix	102	1
14	Lin-Pwr-0.25	14508	Pwr-Lin-0.25	8060	2
15	HARA-Lin-0.25	205	HARA-TK-0.25	133	2
16	Lin-Pwr-0.50	128	Lin-Pwr-0.25	84	2
17	Lin-Lin-0.25	312041	Lin-Lin-0.10	309644	1
18	Lin-Lin-0.25	420468	Lin-Lin-0.10	190253	2
19	Lin-Pwr-0.50	155	Lin-Pwr-0.25	95	2
20	Lin-Lin-0.10	106113391	Lin-Pwr-0.10	21222678	5
21	SIM-Diff-0.50	23	LSO-Diff-0.50	12	2
22	Lin-Pwr-0.25	45909	Lin-Pwr-0.10	32294	1
23	Lin-Lin-0.50	818	Lin-Lin-0.25	296	3
24	HARA-GE-Mix	77	HARA-WG-Mix	75	1
25	Lin-Lin-0.25	2937	Lin-Lin-0.50	925	3
26	Lin-Pwr-0.10	627717597	Pwr-Lin-0.10	348731998	2
27	Lin-Pwr-0.10	4294098	Pwr-Lin-0.10	2385610	2
28	Lin-Lin-0.10	387645	Lin-Lin-0.25	376708	1
29	LSO-Diff-Mix	80	LSO-Ratio-0.25	67	1
30	Lin-Lin-0.50	191	Quad-TK-0.50	58	3
31	SIM-Diff-0.50	50	LSO-Diff-Mix	41	1
32	Lin-Lin-0.10	871794696	Lin-Pwr-0.10	174358939	5
33	LSO-Diff-Mix	169	SIM-Diff-0.50	39	4
34	Lin-Lin-0.10	630476240	Lin-Pwr-0.10	126095248	5
35	Lin-Pwr-0.10	319399465	Pwr-Lin-0.10	177444147	2
36	Lin-Lin-0.25	1706	Lin-Lin-0.50	788	2
37	Lin-Lin-0.25	18858	Lin-Pwr-0.25	3772	5
38	Lin-Lin-0.25	1553	Lin-Lin-0.50	847	2
39	Lin-Pwr-0.25	53427	Pwr-Lin-0.25	29682	2
40	Lin-Lin-0.50	533	LSO-Diff-Mix	169	3
41	Lin-Lin-0.50	411	LSO-Diff-Mix	101	4
42	Lin-Lin-0.25	166436	Lin-Pwr-0.25	33287	5
43	Lin-Pwr-0.10	1706352	Pwr-Lin-0.10	947974	2
44	Lin-Lin-0.10	15593636	Lin-Pwr-0.10	3118727	5
45	Lin-Pwr-0.25	1522	Pwr-Lin-0.25	846	2
46	Lin-Lin-0.10	2238820364	Lin-Pwr-0.10	447764073	5
47	Pwr-WG-Mix	57	Pwr-Prl-II-Mix	54	1
48	Pwr-Prl-II-Mix	79	Pwr-WG-Mix	78	1
49	LSO-Diff-Mix	425	SIM-Diff-Mix	86	5
50	LSO-Diff-0.10	719	LSO-Diff-0.25	695	1

Table 6.14: The best and second-best models and their Bayes factors for Set 3.

ID	Best Theory	Best BF	Sec. Theory	Sec. BF	Sec./Best
1	SIM-Diff-0.50	65	SIM-Ratio-Mix	63	1
2	Lin-Pwr-0.25	9632	Lin-GE-0.25	7224	1
3	Lin-Pwr-0.25	713	Lin-GE-0.25	535	1
4	Bell-TK-Mix	14256	Bell-Prl-I-Mix	11250	1
5	Lin-Pwr-0.25	6023	Lin-GE-0.25	4517	1
6	LSO-Diff-0.50	18	LSO-Ratio-0.50	17	1
7	log-Prl-II-Mix	3381	Bell-Prl-I-Mix	1564	2
8	log-Prl-II-Mix	1047	Expo-TK-Mix	653	2
9	Lin-Pwr-0.10	1681126	Lin-GE-0.10	1260844	1
10	Bell-Prl-I-Mix	462	Bell-TK-Mix	326	1
11	Lin-Pwr-0.50	81	Lin-GE-0.50	61	1
12	Lin-Pwr-0.25	10325	Lin-GE-0.25	7744	1
13	Lin-Pwr-0.50	126	Expo-TK-Mix	125	1
14	Lin-Pwr-0.50	65	Lin-GE-0.50	49	1
15	Bell-Prl-I-0.25	2758	Bell-TK-0.25	2561	1
16	Lin-Pwr-0.50	85	Lin-GE-0.50	64	1
17	Lin-Pwr-0.25	8465	Lin-GE-0.25	6349	1
18	SIM-Diff-0.10	11157	LSO-Diff-0.10	11056	1
19	SIM-Ratio-Mix	1398	SIM-Diff-Mix	643	2
20	Lin-Pwr-0.10	22119333	Lin-GE-0.10	16589500	1
21	Bell-Prl-I-0.25	2255	Bell-TK-0.25	2094	1
22	Lin-Pwr-0.25	7399	Lin-GE-0.25	5549	1
23	Lin-Pwr-0.25	1189	Lin-GE-0.25	891	1
24	Pwr-TK-Mix	111	Lin-Prl-II-Mix	110	1
25	Pwr-GE-Mix	5554	log-GE-Mix	4785	1
26	Lin-Pwr-0.10	373136727	Lin-GE-0.10	279852546	1
27	SIM-Ratio-Mix	585	SIM-Diff-Mix	370	2
28	Bell-Prl-I-Mix	944	log-TK-Mix	921	1
29	Bell-Prl-I-Mix	13207	Bell-TK-Mix	9182	1
30	Lin-Pwr-0.25	2216	Lin-GE-0.25	1662	1
31	LSO-Ratio-0.50	15	LSO-Diff-0.50	13	1
32	log-Lin-0.25	20856	Bell-Lin-0.25	17380	1
33	Expo-Prl-I-Mix	217	Expo-Pwr-Mix	190	1
34	Lin-Pwr-0.10	49156125	Lin-GE-0.10	36867094	1
35	log-WG-Mix	90	log-Prl-II-Mix	82	1
36	LSO-Diff-0.25	1503	LSO-Ratio-0.25	1502	1
37	log-Lin-0.50	28	Quad-WG-Mix	27	1
38	SIM-Ratio-Mix	182	Quad-TK-Mix	181	1
39	SIM-Ratio-Mix	606	SIM-Diff-Mix	338	2
40	Pwr-TK-Mix	347	log-Prl-I-Mix	340	1
41	SIM-Ratio-Mix	152	SIM-Diff-0.25	115	1
42	Pwr-Prl-II-Mix	523	Pwr-Pwr-Mix	515	1
43	Lin-Pwr-0.10	262532683	Lin-GE-0.10	196899512	1
44	SIM-Diff-0.10	348442	LSO-Diff-0.10	345301	1
45	Lin-Pwr-0.25	3515	Lin-GE-0.25	2636	1
46	Lin-Pwr-0.25	8500	Lin-GE-0.25	6375	1
47	Lin-Pwr-0.25	5599	Lin-GE-0.25	4199	1
48	Bell-Prl-I-Mix	334	Pwr-Prl-I-Mix	311	1
49	Expo-Prl-I-Mix	210	Expo-Pwr-Mix	198	1
50	log-WG-Mix	200	Lin-Pwr-0.50	133	2

Table 6.15: The best and second-best models and their Bayes factors for Set 4.

ID	Best Theory	Best BF	Sec. Theory	Sec. BF	Sec./Best
1	Quad-Prl-II-Mix	33	SIM-Diff-0.50	33	1
2	Lin-Lin-0.25	6892	log-Prl-I-0.25	1379	5
3	Lin-Lin-0.25	739	Lin-Lin-0.50	659	1
4	Lin-Lin-0.10	4037249	log-Prl-I-0.10	807450	5
5	Lin-WG-Mix	1672	Lin-Prl-II-Mix	1619	1
6	LSO-Diff-0.50	16	LSO-Diff-0.25	9	2
7	Lin-Lin-0.25	30500	log-Prl-I-0.25	6100	5
8	LSO-Diff-0.10	2599503	SIM-Diff-0.10	1313653	2
9	log-Prl-I-0.25	3039	Lin-Prl-I-0.25	2532	1
10	Lin-Lin-0.10	390181	Lin-Lin-0.25	363493	1
11	Expo-Pwr-Mix	147	Lin-Lin-0.50	101	1
12	Lin-Lin-0.25	29213	log-Prl-I-0.25	5843	5
13	Lin-Lin-0.50	716	log-Prl-I-0.50	158	5
14	Lin-Lin-0.10	23202820	log-Prl-I-0.10	4640564	5
15	log-Prl-I-0.10	106901	Lin-Prl-I-0.10	89084	1
16	Expo-Pwr-Mix	26	Expo-GE-Mix	22	1
17	SIM-Diff-0.50	45	Pwr-TK-Mix	26	2
18	SIM-Diff-0.10	11530	SIM-Diff-0.25	469	25
19	Lin-Lin-0.50	707	log-Prl-I-0.50	175	4
20	Pwr-Lin-Mix	3950	Lin-Prl-II-Mix	3921	1
21	log-Prl-I-0.25	12296	Lin-Prl-I-0.25	10246	1
22	log-Prl-I-0.10	18704016	Lin-Prl-I-0.10	15586680	1
23	log-WG-Mix	186	Expo-Pwr-Mix	177	1
24	log-Prl-I-0.50	160	Lin-Lin-0.50	141	1
25	Pwr-Prl-I-Mix	4124	Lin-Prl-II-Mix	4106	1
26	Lin-Lin-0.10	2266327066	log-Prl-I-0.10	453265413	5
27	Lin-Lin-0.10	1917294	Lin-Lin-0.25	436877	4
28	Lin-Lin-0.25	99520	LSO-Diff-0.10	47471	2
29	Pwr-Prl-I-Mix	537	Lin-WG-Mix	509	1
30	Lin-Lin-0.10	1139171549	log-Prl-I-0.10	227834310	5
31	Expo-Pwr-Mix	167	Expo-WG-Mix	103	2
32	Lin-Lin-0.10	1596997327	log-Prl-I-0.10	319399465	5
33	Lin-Lin-0.25	2266	Lin-Lin-0.50	781	3
34	Lin-Prl-I-0.50	48	Bell-Pwr-Mix	44	1
35	Lin-Lin-0.10	15415553	log-Prl-I-0.10	3083111	5
36	Lin-WG-Mix	5152	Pwr-Prl-I-Mix	5142	1
37	Lin-Lin-0.50	481	log-Prl-I-0.50	193	2
38	Expo-Pwr-Mix	121	log-Prl-I-0.50	83	1
39	Lin-Lin-0.10	22371287	log-Prl-I-0.10	4474257	5
40	Lin-Lin-0.10	2266327066	log-Prl-I-0.10	453265413	5
41	Expo-Pwr-Mix	215	Quad-Pwr-Mix	137	2
42	log-Prl-I-0.25	573	Lin-Prl-I-0.25	478	1
43	Lin-Lin-0.10	1593866	Lin-Lin-0.25	472358	3
44	Lin-Lin-0.10	20097919	log-Prl-I-0.10	4019584	5
45	Lin-Pwr-0.50	25	SIM-Diff-0.50	25	1
46	log-Prl-I-0.25	396	Lin-Prl-I-0.25	330	1
47	Lin-Lin-0.50	113	Quad-Lin-0.50	34	3
48	Lin-Lin-0.10	8587583	log-Prl-I-0.10	1717517	5
49	log-Prl-I-0.25	265	Lin-Prl-I-0.25	221	1
50	Lin-WG-Mix	5935	Pwr-Prl-I-Mix	5925	1

Table 6.16: The best and second-best models and their Bayes factors for Set 5.

ID	Best Theory	Best BF	Sec. Theory	Sec. BF	Sec./Best
1	Pwr-Lin-Mix	242	SIM-Diff-0.50	43	6
2	Pwr-Lin-Mix	140	LSO-Diff-Mix	87	2
3	Lin-Pwr-0.25	141	Lin-TK-0.50	79	2
4	Lin-Pwr-0.10	43863283	Pwr-Lin-0.10	16083204	3
5	Lin-Pwr-0.25	2883	Pwr-Lin-0.25	1064	3
6	Lin-Pwr-0.10	145181575	Pwr-Lin-0.10	53233244	3
7	SIM-Diff-0.10	9666	SIM-Diff-0.25	2774	3
8	Lin-Pwr-0.10	104833427	Pwr-Lin-0.10	38438923	3
9	Lin-Pwr-0.50	49	Lin-TK-0.50	49	1
10	Lin-Pwr-0.25	5416	Pwr-Lin-0.25	1986	3
11	Pwr-Lin-Mix	105	SIM-Diff-0.50	49	2
12	Lin-Pwr-0.25	11516	Pwr-Lin-0.25	4223	3
13	Lin-Pwr-0.25	2307	LSO-Diff-0.25	915	3
14	LSO-Diff-Mix	89	LSO-Ratio-Mix	58	2
15	Pwr-Lin-Mix	35	Lin-Pre-II-Mix	18	2
16	Lin-Lin-0.50	94	Lin-Pwr-0.50	45	1
17	Lin-Pwr-0.10	53852	Lin-Pwr-0.25	24643	2
18	Lin-Pwr-0.10	145181575	Pwr-Lin-0.10	53233244	3
19	Lin-Pwr-0.25	97	Lin-Lin-0.50	90	1
20	Lin-Pwr-0.10	285326180	Pwr-Lin-0.10	104619600	3
21	LSO-Diff-0.25	759	SIM-Diff-0.25	575	1
22	Pwr-Lin-Mix	29	SIM-Diff-0.50	13	2
23	Lin-Pwr-0.10	7550673	Pwr-Lin-0.10	2768580	3
24	Pwr-Lin-Mix	38	Lin-TK-0.50	24	2
25	Lin-Pwr-0.25	14714	Pwr-Lin-0.25	5395	3
26	Lin-Pwr-0.10	200750669	Pwr-Lin-0.10	73608579	3
27	SIM-Diff-0.50	20	LSO-Ratio-Mix	19	1
28	Lin-Pwr-0.10	16140817	Pwr-Lin-0.10	5918300	3
29	LSO-Diff-0.25	195	SIM-Diff-0.25	147	1
30	Lin-Pwr-0.10	1880441	Pwr-Lin-0.10	689495	3
31	Lin-Pwr-0.25	2993	Pwr-Lin-0.25	1097	3
32	Lin-Pwr-0.10	145181575	Pwr-Lin-0.10	53233244	3
33	Lin-Lin-0.50	42	SIM-Diff-0.50	24	1
34	Lin-Pwr-0.10	139007231	Pwr-Lin-0.10	50969318	3
35	Lin-Pwr-0.10	9780185	Pwr-Lin-0.10	3586068	3
36	LSO-Diff-0.10	531554	SIM-Diff-0.10	410190	1
37	Lin-Pwr-0.25	7172	Pwr-Lin-0.25	2630	3
38	Pwr-Lin-Mix	298	LSO-Diff-Mix	196	2
39	Lin-Pwr-0.10	93793	Lin-Pwr-0.25	46480	2
40	Lin-Pwr-0.25	27691	Lin-Pwr-0.10	11895	2
41	Lin-Pwr-0.25	100	Lin-Pwr-0.50	79	1
42	Lin-Pwr-0.10	39145953	Pwr-Lin-0.10	14353516	3
43	Pwr-Lin-Mix	320	Lin-Pre-II-Mix	20	16
44	Lin-Pwr-0.10	40884716	Pwr-Lin-0.10	14991062	3
45	Lin-Pwr-0.10	601784	Pwr-Lin-0.10	220654	3
46	Lin-Pwr-0.10	288831769	Pwr-Lin-0.10	105904982	3
47	Pwr-Lin-Mix	373	LSO-Diff-0.25	84	4
48	Lin-TK-0.50	42	Lin-Pwr-0.50	42	1
49	LSO-Diff-0.25	187	SIM-Diff-0.25	130	1
50	Lin-Pwr-0.25	16007	Pwr-Lin-0.25	5869	3

Table 6.17: The top ten models ranked by GBF in each stimulus set, i.e., the best model (highest GBF) to worst model (lowest GBF). The bottom panel shows the log10 value of GBF for the corresponding model.

	Set 1	Set 2	Set 3	Set 4	Set 5
Ranking	Probabilistic Models				
1	Expo-Pwr-0.50	SIM-Diff-0.25	Bell-Prl-I-0.50	LSO-Diff-0.25	log-Lin-0.50
2	Expo-WG-0.50	SIM-Ratio-0.25	Bell-TK-0.50	Lin-Pwr-0.50	log-Pwr-0.50
3	Expo-Prl-I-0.50	LSO-Diff-0.25	Bell-Prl-II-0.50	Lin-GE-0.50	Lin-Prl-I-0.50
4	Bell-Prl-II-0.50	LSO-Ratio-0.25	LSO-Diff-0.25	Lin-WG-0.50	Pwr-GE-0.50
5	SIM-Diff-0.25	Pwr-Lin-0.50	Lin-Prl-II-0.50	Lin-Prl-II-0.50	Pwr-TK-0.50
6	SIM-Ratio-0.25	Lin-Pwr-0.50	log-Pwr-0.50	Pwr-Lin-0.50	Lin-Prl-II-0.50
7	Expo-GE-0.50	HARA-Pwr-0.50	log-TK-0.50	Pwr-TK-0.50	Lin-Pwr-0.50
8	Expo-Prl-II-0.50	HARA-Lin-0.50	log-Prl-I-0.50	Pwr-Prl-I-0.50	Lin-Lin-0.50
9	Bell-WG-0.50	SIM-Diff-0.50	Pwr-Lin-0.50	SIM-Diff-0.25	Pwr-Lin-0.50
10	LSO-Diff-0.25	SIM-Ratio-0.50	Pwr-Pwr-0.50	SIM-Ratio-0.25	Lin-WG-0.50
Ranking	The log10 value of GBF for each model				
1	84.34	97.11	81.75	86.00	120.65
2	84.34	97.11	80.27	81.83	120.65
3	84.34	87.77	78.99	81.83	108.06
4	82.63	70.84	78.02	81.83	94.40
5	82.59	56.99	77.18	81.83	94.40
6	82.59	56.36	77.18	81.83	85.94
7	81.33	56.16	77.18	81.83	81.05
8	81.33	55.99	77.18	81.83	78.53
9	81.04	54.79	77.18	80.64	70.12
10	79.48	54.79	77.18	80.64	66.22

6.6 Conclusions

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. 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. On the other hand, CPT is the most famous contemporary theory of risky choice. This paper presents a comprehensive analysis of the lexicographic semiorder model, the similarity model, and 49 forms of CPT, and compares the intransitive heuristics with the transitive CPT. This paper tries to find out which model can explain human choice behavior better, transitive theories or intransitive heuristics.

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 864 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. The Bayesian order-constrained method allows me to put all of the 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. Overall, there is a close alignment between frequentist and Bayes factor results for the distance-based models, even though these two methods are conceptually and computationally quite distinct.

The analysis of the distance-based models shows that the intransitive heuristics fit more participants' data than CPT. The mixture model analysis shows that the number of fits differs a lot among different theories within the same stimulus set and across different stimulus sets.

The model comparison at the individual level shows that for Sets 2 and 4, the distance-based model of CPT that makes a single predicted pattern, for example, CPT with the linear utility function and the linear probability weighting function, wins out for the most participants. This result means that there is a group of participants who might simplify the task and compute the expected value of gambles for Sets 2 and 4. However, these models do not win out at the model comparison at the group level by GBF, which means that these CPT models could not account for all participants' data jointly.

For Sets 1, 3, and 5, the distance-based model of CPT with the linear utility function and the power probability weighting function wins out for the most participants. For each stimulus set, there are more probabilistic models of CPT winning out than the intransitive heuristics. In this sense, it seems that CPT is doing a better job in terms of explaining the participants' data.

Moreover, the model comparison 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 result 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), Guo (2018a), Hey (2005), Loomes et al. (2002), and Regenwetter et al. (2014).

In this paper, I use five different stimulus sets, two of which have only non-negative outcomes, two of which have only non-positive outcomes, and one of which have both gains and losses. Overall, the analysis results for different stimulus sets do not differ a lot. I do not see much difference regarding model performance for various stimulus sets.

One thing to point out is 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 result speaks directly to Birnbaum (2011)'s concern about model mimicry. My analyses show that many participants are fit by both the intransitive heuristics and different forms of CPT. One explanation for this finding might be that many preference patterns predicted by the intransitive heuristics are transitive. 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. In this case, the model comparison by Bayes factor is very important since Bayes factor appropriately penalizes the complexity of the model. In this study, there are more probabilistic models of CPT winning out than the intransitive heuristics for each stimulus set.

Last but not the least, I would like to point out that the paper is a large scale project for a systematic test of both transitive and intransitive decision theories. All the quantitative analyses in this paper consumed about 304,000 CPU hours on the supercomputer at Pittsburgh Supercomputing Center.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4):503–546.

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

Arló-Costa, H. and Pedersen, A. P. (2013). Fast and frugal heuristics: Rationality and the limits of naturalism. *Synthese*, 190(5):831–850.

Baillon, A., Bleichrodt, H., and Cillo, A. (2015). A tailor-made test of intransitive choice. *Operations Research*, 63(1):198–211.

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

Becker, B. and Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. *Academy of Management Journal*, 39(4):779–801.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, 30(5):961–981.

Bell, D. E. (1983). Risk premiums for decision regret. *Management Science*, 29(10):1156–1166.

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

Birnbaum, M. and Schmidt, U. (2008). An experimental investigation of violations of transitivity in choice under uncertainty. *Journal of Risk and Uncertainty*, 37(1):77–91.

Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes*, 95(1):40–65.

Birnbaum, M. H. (2008a). Evaluation of the priority heuristic as a descriptive model of risky decision making: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, 115(1):253–260.

Birnbaum, M. H. (2008b). New paradoxes of risky decision making. *Psychological Review*, 115(2):463.

Birnbaum, M. H. (2010). Testing lexicographic semiorders as models of decision making: Priority dominance, integration, interaction, and transitivity. *Journal of Mathematical Psychology*, 54(4):363–386.

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 Bahra, J. P. (2007). Gain-loss separability and coalescing in risky decision making. *Management Science*, 53(6):1016–1028.

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.

Birnbaum, M. H. and LaCroix, A. R. (2008). Dimension integration: Testing models without trade-offs. *Organizational Behavior and Human Decision Processes*, 105(1):122–133.

Blavatskyy, P. R. and Pogrebna, G. (2010). Models of stochastic choice and decision theories: why both are important for analyzing decisions. *Journal of Applied Econometrics*, 25(6):963–986.

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

Bouyssou, D. and Pirlot, M. (2002). Nontransitive decomposable conjoint measurement. *Journal of Mathematical Psychology*, 46(6):677–703.

Bouyssou, D. and Pirlot, M. (2004). Additive difference models without additivity and subtractivity. *Journal of Mathematical Psychology*, 48(4):263–291.

Bouyssou, D. and Vansnick, J.-C. (1986). Noncompensatory and generalized noncompensatory preference structures. *Theory and Decision*, 21(3):251–266.

Bradbury, H. and Nelson, T. M. (1974). Transitivity and the patterns of children's preferences. *Developmental Psychology*, 10(1):55–64.

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

Brandstätter, E. and Gussmack, M. (2013). The cognitive processes underlying risky choice. *Journal of Behavioral Decision Making*, 26(2):185–197.

Buschena, D. E. and Atwood, J. A. (2011). Evaluation of similarity models for expected utility violations. *Journal of Econometrics*, 162(1):105–113.

Buschena, D. E. and Zilberman, D. (1999). Testing the effects of similarity on risky choice: Implications for violations of expected utility. *Theory and Decision*, 46(3):253–280.

Butler, D. J. (1998). A choice-rule formulation of intransitive utility theory. *Economics Letters*, 59(3):323–329.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304(5674):1167–1170.

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

Cha, Y., Choi, M., Guo, Y., Regenwetter, M., and Zwilling, C. (2013). Reply: Birnbaum's (2012) statistical tests of independence have unknown Type-I error rates and do not replicate within participant. *Judgment and Decision Making*, 8(1):55–73.

Chen, Y.-J. and Corter, J. E. (2006). When mixed options are preferred in multiple-trial decisions. *Journal of Behavioral Decision Making*, 19(1):17–42.

Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions: rendues à la pluralité des voix [Essay on the application of the probabilistic analysis of majority vote decisions]*. Pairs, France: Imprimerie Royale.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., and Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience*, 8(9):1255–1262.

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. (2012). A lexicographic semiorder polytope and probabilistic representations of choice. *Journal of Mathematical Psychology*, 56(2):86–94.

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.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Fishburn, P. C. (1980). Lexicographic additive differences. *Journal of Mathematical Psychology*, 21(3):191 – 218.

Fishburn, P. C. (1990). Continuous nontransitive additive conjoint measurement. *Mathematical Social Sciences*, 20(2):165–193.

Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4(2):113–134.

Fishburn, P. C. (1992). Additive differences and simple preference comparisons. *Journal of Mathematical Psychology*, 36(1):21–31.

Gao, S., Frejinger, E., and Ben-Akiva, M. (2010). Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation Research Part C: Emerging Technologies*, 18(5):727–740.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Glöckner, A. and Betsch, T. (2008). Do people make decisions under risk based on ignorance? An empirical test of the priority heuristic against cumulative prospect theory. *Organizational Behavior and Human Decision Processes*, 107(1):75–95.

Glöckner, A. and Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, 24(1):71–98.

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

Gonzalez-Vallejo, C., Bonazzi, A., and J. Shapiro, A. (1996). Effects of vague probabilities and of vague payoffs on preference: A model comparison analysis. *Journal of Mathematical Psychology*, 40(2):130–140.

Guo, Y. (2018a). Rationality or irrationality of preferences? A quantitative test of intransitive decision models. Master's thesis, University of Illinois at Urbana-Champaign.

Guo, Y. (2018b). 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.

Harrison, G. W., Humphrey, S. J., and Verschoor, A. (2010). Choice under uncertainty: evidence from ethiopia, india and uganda. *The Economic Journal*, 120(543):80–104.

Hays, W. L. (1988). *Statistics (4th Edition)*. Holt, Rinehart &Winston, New York: NY.

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.

Humphrey, S. J. (2001). Non-transitive choice: Event-splitting effects or framing effects. *Econometrica*, 68(269):77–96.

Humphrey, S. J. (2004). Feedback-conditional regret theory and testing regret-aversion in risky choice. *Journal of Economic Psychology*, 25(6):839–857.

Inman, J. J., Dyer, J. S., and Jia, J. (1997). A generalized utility model of disappointment and regret effects on post-choice valuation. *Marketing Science*, 16(2):97–111.

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.

Johnson, E. J., Schulte-Mecklenbeck, M., and Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, 115(1):263–273.

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.

Kohli, R. and Jedidi, K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science*, 26(3):380–399.

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

Leland, J. W. (2002). Similarity judgments and anomalies in intertemporal choice. *Economic Inquiry*, 40(4):574–581.

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

Loomes, G. (2006). The improbability of a general, rational and descriptively adequate theory of decision under risk. Manuscript.

Loomes, G. (2010a). Modeling choice and valuation in decision experiments. *Psychological Review*, 117(3):902–924.

Loomes, G. (2010b). Modeling choice and valuation in decision experiments. *Psychological Review*, 117(3):902–924.

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., Starmer, C., and Sugden, R. (1991). Observing violations of transitivity by experimental methods. *Econometrica*, 59(2):425–439.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824.

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.

Loomes, G. and Taylor, C. (1992). Non-transitive preferences over gains and losses. *The Economic Journal*, 102(411):357–365.

Lopes, L. and Oden, G. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and sp/a theory. *Journal of Mathematical Psychology*, 43(2):286–313.

Lorentziadis, P. L. (2013). Preference under risk in the presence of indistinguishable probabilities. *Operational Research*, 13(3):429–446.

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. (1978). Lexicographic tradeoff structures. *Theory and Decision*, 9(2):187–193.

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.

Luce, R. D. and Suppes, P. (1965). Preference, utility, and subjective probability. *Handbook of Mathematical Psychology*, 3:249–410.

Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *The American Economic Review*, 97(5):1824–1839.

Manzini, P. and Mariotti, M. (2012). Choice by lexicographic semiorders. *Theoretical Economics*, 7(1):1–23.

McNamara, T. and Diwadkar, V. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology*, 34(2):160–190.

Mellers, B., Chang, S., Birnbaum, M., and Ordóñez, L. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2):347–361.

Mellers, B., Schwartz, A., and Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, 128(3):332–345.

Mood, A. M. (1950). *Introduction to the Theory of Statistics.* McGraw-Hill.

Morrison, H. W. (1962). *Intransitivity of paired comparison choices.* PhD thesis, University of Michigan.

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

Pachur, T., Hertwig, R., Gigerenzer, G., and Brandstätter, E. (2013). Testing process predictions of models of risky choice: A quantitative model comparison approach. *Frontiers in Psychology*, 4:646.

Raeva, D., Mittone, L., and Schwarzbach, J. (2010). Regret now, take it now: On the role of experienced regret on intertemporal choice. *Journal of Economic Psychology*, 31(4):634–642.

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.

Riechard, D. E. (1991). Intransitivity of paired comparisons related to gender and community socioeconomic setting. *The Journal of Experimental Education*, 59(2):197–205.

Rieger, M. O. and Wang, M. (2008). What is behind the priority heuristic? A mathematical analysis and comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, 115(1):274–280.

Rieger, M. O., Wang, M., and Hens, T. (2017). Estimating cumulative prospect theory parameters from an international survey. *Theory and Decision*, 82(4):567–596.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6):1446–1465.

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.

Schuck-Paim, C. and Kacelnik, A. (2002). Rationality in risk-sensitive foraging choices by starlings. *Animal Behaviour*, 64(6):869–879.

Sedransk, J., Monahan, J., and Chiu, H. Y. (1985). Bayesian estimation of finite population parameters in categorical data models incorporating order restrictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):519–527.

Shafir, S. (1994). Intransitivity of preferences in honey bees: support for 'comparative' evaluation of foraging options. *Animal Behaviour*, 48(1):55–67.

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.

Starmer, C. and Sugden, R. (1998). Testing alternative explanations of cyclical choices. *Economica*, 65(259):347–361.

Steel, P. and König, C. (2006). Integrating theories of motivation. *Academy of Management Review*, 31(4):889–913.

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.

Stott, H. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, 32(2):101–130.

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.

Trepel, C., Fox, C., and Poldrack, R. (2005). Prospect theory on the brain? toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research*, 23(1):34–50.

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

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4):281–299.

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4):251–278.

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

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

Vilà, X. (1998). On the intransitivity of preferences consistent with similarity relations. *Journal of Economic Theory*, 79(2):281–287.

Vind, K. (1991). Independent preferences. *Journal of Mathematical Economics*, 20(1):119–135.

Waite, T. A. (2001). Intransitive preferences in hoarding gray jays (Perisoreus canadensis). *Behavioral Ecology and Sociobiology*, 50(2):116–121.

Xu, H., Zhou, J., and Xu, W. (2011). A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. *Transportation Research Part C: Emerging Technologies*, 19(2):218–228.

Yee, M., Dahan, E., Hauser, J. R., and Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, 26(4):532–549.

Appendix A: Parsimonious Testing of Transitive or Intransitive Preferences: Reply to Birnbaum (2011)

Psychological Review
2011, Vol. 118, No. 4, 684–688

0033-295X/11/$12.00 DOI: 10.1037/a0025291

REPLY

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

Michel Regenwetter
University of Illinois at Urbana–Champaign

Jason Dana
University of Pennsylvania

Clintin P. Davis-Stober
University of Missouri

Ying Guo
University of Illinois at Urbana–Champaign

Birnbaum (2011) raised 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 5 choice alternatives. While the approach of Regenwetter, Dana, and Davis-Stober (2011) achieves high power with modest sample sizes for 5 alternatives, pattern-counting approaches face the difficulty of combinatoric explosion in permissible response patterns. Even for fewer than 5 alternatives, if the choice of how to "block" individual responses into response patterns is slightly mistaken, intransitive preferences can mimic transitive ones. At the same time, statistical tests on proportions of response patterns rely on similar "independent and identically distributed" 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 data and, consistent with Birnbaum's (2011) concern, we find evidence for model mimicry in some cases.

Keywords: parsimonious testing, random utility, rationality, transitivity of preferences

Regenwetter, Dana, and Davis-Stober (2011) 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 independent and identically distributed (iid) sampling.

Michel Regenwetter and Ying Guo, Department of Psychology, University of Illinois at Urbana–Champaign; Jason Dana, Department of Psychology, University of Pennsylvania; Clintin P. Davis-Stober, Department of Psychological Sciences, University of Missouri.

Michel Regenwetter is supported by National Science Foundation DRMS Award SES 08-20009. Any opinions expressed in this publication are those of the authors and do not necessarily reflect the views of universities or the funding agency.

Correspondence concerning this article should be addressed to Michel Regenwetter, Quantitative Psychology, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: regenwet@illinois.edu

Birnbaum (2011) described an alternative quantitative approach to testing transitivity on within-subject data. He agreed with the substantive conclusion of Regenwetter et al. (2011), henceforth RDDS, that evidence for intransitivity is lacking and also with their criticism of weak stochastic transitivity. He argued, however, that RDDS did not go far enough in criticizing past approaches, particularly because they analyzed proportions of binary responses. He contrasted their approach with that of Birnbaum and Gutierrez (2007), who, instead, analyzed 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 showed how using the RDDS approach could suggest that choices are transitive when in fact the decision maker has intransitive preferences, a phenomenon we call *model mimicry*. His example showed how analyzing patterns, as Birnbaum and Gutierrez did, could diagnose this true intransitivity, because their approach identifies a preference distribution and that of RDDS does not. Birnbaum's comment also questioned 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 focuses on a small number of key points. We start and end with a 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 did, 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 three gambles, there are eight 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 the 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 used to test transitivity. Retaining transitivity with three gambles is not very informative because most conceivable data sets will support transitivity.

On the other hand, if one uses five 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 three gambles to five 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 five choice alternatives.

Because there are 1,024 possible response patterns for five gambles, combinatoric explosion will pose a formidable problem for any pattern-counting approach. Consider again Birnbaum's (2011) Table 2. The example data used 200 repetitions, so that there are 25 observations for each of the eight possible response patterns. To obtain an average of 25 observations per pattern with five 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 many more permissible patterns (see Table 1).

Because a strong test of transitivity requires five 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) questioned, 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. Because we could assign a 0 or 1 to each item (as Birnbaum did for patterns of three in his Table 2) and 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 Equation 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 three 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 eight kinds of patterns occur, reducing the data to 7 degrees of freedom (the number is 7 because once seven pattern frequencies are observed, the eighth is determined, as 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 eight binary relations. The three 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 differed in that they did not preserve any temporal information about the sequence of these decisions. They assumed 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 reduced 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 1 summarizes these and other insights. Pattern-counting approaches such as 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?

Although RDDS estimated binary choice probabilities and tested transitivity, they did not estimate the unique distribution of preferences that their model assumes exists. Birnbaum (2011) gave a hypothetical mixture of intransitive states that

Table 1
Commonalities (Centered) and Differences (Split) Between Birnbaum's (2011) and Regenwetter et al.'s (2010, 2011) Approaches to Testing Transitivity of Preferences, as Well as Key Strengths and Weaknesses

(Birnbaum, 2011)	**(Regenwetter et al., 2010, 2011)**
Uses blocking assumptions in order to group observed binary responses into patterns.	Does not make blocking assumptions.
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	
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.*	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.*
Can identify a unique preference distribution from *pattern* frequencies.	*Cannot* identify a unique preference distribution from *response* frequencies.
Tests transitivity under assumption that the decision makers are never indifferent between any two prospects. (Tests "strict linear orders.")	
Can be extended to a more direct test of transitivity ("strict weak orders") by permitting additional "no preference" response category ("ternary choice")	
at cost of combinatoric explosion: 3 prospects: $3^{\binom{3}{2}} - 1 = 26$ degrees of freedom, 5 prospects: $3^{\binom{5}{2}} - 1 = 59,048$ degrees of freedom.	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.
An experiment with 5 choice prospects and 20 observations per empirical cell corresponds to	
$20 \times \binom{5}{2} \times 2^{\binom{5}{2}} = 204,800$ binary choices or $20 \times \binom{5}{2} \times 3^{\binom{5}{2}} = 11,809,800$ ternary choices, plus fillers between blocks.	$20 \times \binom{5}{2} = 200$ binary choices or $20 \times \binom{5}{2} = 200$ ternary choices, plus distractors between choices.
Assumes each observed pattern is composed of a preference relation and *independent* errors.	Does not assume errors, but could enlarge their model (the "polytope") to accommodate *interdependent* errors, with risk of overfitting.
Can fall victim to model mimicry.	
Avoids aggregation across individuals.	
Avoids descriptive modal choice analysis.	
Uses quantitative goodness-of-fit methodology.	
Does not assume that preferences are induced by *independent* random utilities.	
Concludes overall that: **Existing evidence for intransitivity of preferences is not compelling.**	

Note. Birnbaum (1984) used similar calculations to count patterns. iid = independent and identically distributed.

RDDS would falsely diagnose as supporting transitivity while an analysis of response patterns would detect intransitivity.

What if the data are incorrectly blocked? We give a simple example using three choice alternatives where pattern counting is vulnerable to model mimicry, much as Birnbaum's (2011) thought experiment showed potential model mimicry in the RDDS approach. 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 a 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 nine responses, 000111000, when blocked, will appear as 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. 676) 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 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 Figure 1).

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 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 decision maker has considered both attributes without a conclusion, then we assume, for simplicity, that he chooses either alternative with probability one half.

Figure 1 shows Tversky's (1969) gambles and the ratios for each attribute. If the decision maker 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 C (from top). Notice the intransitive cycle $a \succ e$, $e \succ c$, $c \succ a$. The right side of Panel C shows the choice probabilities for a decision maker with just that one preference.

A

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

B

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

C

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			-	$\frac{1}{2}$

D

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

Figure 1. Tversky's (1969) gambles. Panel A shows the chance of winning and payoff value for each of the five gambles. Panel B shows the chance ratios (left) and the payoff ratios (right) as decimals. Panel C provides one of the 111 lexicographic semiorders one can obtain this way (left) and the choice probabilities of a decision maker who has only that single preference. Panel D shows the result of testing the lexicographic semiorder mixture model on Tversky's eight participants and on RDDS's 18 participants for Cash I and Cash II (with $\alpha = .05$). The last column of Panel D shows the number of simultaneous inequality constraints tested. PH denotes the priority heuristic. RDDS = Regenwetter, Dana, & Davis-Stober (2011).

For brevity, we only sketch the model and its test. The lexicographic semiorder given in Panel C of Figure 1 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 the decision maker currently strictly prefers i to j plus $1/2$ 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. We consider 111 lexicographic semiorders instead of 120 linear orders.

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. Figure 1 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, Gigerenzer, & Hertwig, 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?

Table 1 and Figure 1 summarize our findings. Both Birnbaum (Birnbaum, 2011) and Regenwetter et al. (2010, 2011) deliberately eliminated common and often undesirable assumptions in the literature. Both made 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 made blocking and independent error assumptions that RDDS did 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) aimed at classifying and dissecting the implicit or explicit assumptions made in various approaches and developing parsimonious quantitative tests. Every test makes some assumptions, so testing these assumptions is valuable. Even more valuable is to use assumptions that only need to hold 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.

References

Birnbaum, M. H. (1984). Transitivity in the Big Ten. *Bulletin of the Psychonomic Society, 22,* 351–353.

Birnbaum, M. H. (2010). Testing lexicographic semiorders as models of decision making: Priority dominance, integration, interaction, and transitivity. *Journal of Mathematical Psychology, 54,* 363–386. doi:10.1016/j.jmp.2010.03.002

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comment on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review, 118,* 674–682. doi: 10.1037/a0023852

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes, 104,* 96–112. doi:10.1016/j.obhdp.2007.02.001

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113,* 409–432. doi:10.1037/0033-295X.113.2.409

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Quantitative Psychology and Measurement, 1,* 148. doi:10.3389/fpsyg.2010.00148

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 118,* 42–56. doi:10.1037/a0021150

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76,* 31–48. doi:10.1037/h0026750

Received May 5, 2011
Revision received July 20, 2011
Accepted July 21, 2011 ■

Appendix B: Quantitative tests of the Perceived Relative Argument Model: comment on loomes (2010)

Psychological Review
2014, Vol. 121, No. 4, 696–705

0033-295X/14/$12.00 DOI: 10.1037/a0036095

COMMENT

Quantitative Tests of the Perceived Relative Argument Model: Comment on Loomes (2010)

Ying Guo and Michel Regenwetter
University of Illinois at Urbana-Champaign

Loomes (2010, *Psychological Review*) proposed the *Perceived Relative Argument Model* (PRAM) as a novel descriptive theory for risky choice. PRAM differs from models like prospect theory in that decision makers do not compare 2 prospects by first assigning each prospect an overall utility and then choosing the prospect with the higher overall utility. Instead, the decision maker determines the relative argument for one or the other prospect separately for outcomes and probabilities, before reaching an overall pairwise preference. Loomes (2010) did not model variability in choice behavior. We consider 2 types of "stochastic specification" of PRAM. In one, a decision maker has a fixed preference, and choice variability is caused by occasional errors/trembles. In the other, the parameters of the perception functions for outcomes and for probabilities are random, with no constraints on their joint distribution. State-of-the-art frequentist and Bayesian "order-constrained" inference suggest that PRAM accounts poorly for individual subject laboratory data from 67 participants. This conclusion is robust across 7 different utility functions for money and remains largely unaltered also when considering a prior unpublished version of PRAM (Loomes, 2006) that featured an additional free parameter in the perception function for probabilities.

Keywords: error model, Perceived Relative Argument Model, order-constrained inference, random preference model, quantitative testing

Supplemental materials: http://dx.doi.org/10.1037/a0036095.supp

Loomes (2010, *Psychological Review*) developed a descriptive model of individual decision making under risk, the *Perceived Relative Argument Model* (PRAM). PRAM describes how decision makers choose among lotteries in which one can win various payoffs with various probabilities. According to PRAM, the decision maker compares the perceived argument favoring one lottery based on probabilities with the perceived argument favoring the other lottery based on payoffs. She prefers one lottery over the other depending on the perceived relative argument in its favor. PRAM violates several key axioms of rational behavior, for example, independence, betweenness, and transitivity. PRAM applies to a specific domain. It models pairwise preference among two lotteries S, R of a particular form. Writing x_i for the ith monetary outcome and p_i, q_i for the probability of the ith monetary outcome in S and R, respectively, the 'safer' lottery S and the 'riskier' lottery R must satisfy the following properties:

$$\begin{array}{l} S = (x_3,\ p_3;\ x_2,\ p_2;\ x_1,\ p_1) \\ R = (x_3,\ q_3;\ x_2,\ q_2;\ x_1,\ q_1) \end{array} \text{ with } \begin{array}{c} x_3 > x_2 > x_1 \geq 0; \\ q_3 > p_3;\ q_2 < p_2;\ q_1 > p_1. \end{array} \quad (1)$$

According to PRAM, a decision maker faced with the choice between S and R evaluates the relative argument in favor of S in

Ying Guo and Michel Regenwetter, Quantitative Division, Department of Psychology, University of Illinois at Urbana-Champaign.

MATLAB computer code available at https://app.box.com/s/bqrttg9fswc6hzaft7e3

G. Loomes, H. Bleichrodt, A. Abbas, A. Popova, B. Xu, and C. Zwilling provided helpful comments on earlier drafts. We are indebted to C. Davis-Stober, S. H. Lim, and D. Cavagnaro for much advice, as well as for the core computer code used in the statistical analyses. C. Zwilling also contributed extensively to computer programming with support from the Institute for Computing in the Humanities, Arts, and Social Sciences (I-CHASS) via a 2012 *Scalable Research Challenge* award (to Michel Regenwetter and C. Zwilling). We acknowledge funding through National Science Foundation (NSF) Grants SES No. 08-20009 and SES 10-62045 (Principal Investigator [PI]: Michel Regenwetter) and an Arnold O. Beckman Award from the Research Board of the University of Illinois at Urbana-Champaign (PI: Michel Regenwetter). The nearly 60,000 CPU hours needed to carry out all quantitative analyses were made possible by Extreme Science and Engineering Discovery Environment (XSEDE) Grant SES No. 130016 (PI: Michel Regenwetter) on the *Blacklight* supercomputer at Pittsburgh Supercomputing Center, whose technical staff we thank for their support. This article is based on Ying Guo's 2012 doctoral qualifying exam in Quantitative Psychology at the University of Illinois. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and need not reflect the views of the NSF, the Pittsburgh Supercomputing Center, or the University of Illinois. The study was approved by the Institutional Review Board (IRB) of the University of Illinois under IRB No. 12632.

Correspondence concerning this article should be addressed to Michel Regenwetter, Quantitative Division, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: regenwet@illinois.edu

terms of probabilities using a *perception function for probabilities*, $\phi(b_S, b_R)$, via:

$$\phi(b_S, b_R) = \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^{\alpha}}, \text{ where } b_S = q_1 - p_1;\ b_R = q_3 - p_3. \quad (2)$$

This function has a real-valued free parameter α. The decision maker also relies on a *perception function for outcomes*, $\xi(y_R, y_S)$, to determine the relative argument in favor of R regarding the outcomes:

$$\xi(y_R, y_S) = \left(\frac{c_S}{c_R}\right)^{\delta}, \text{ where } c_S = c(x_3) - c(x_2); c_R = c(x_2) - c(x_1). \quad (3)$$

Here, $\delta \geq 1$ is a free parameter. We consider seven different utility functions $c(\ )$ for money (see Table 1).

Let $S \prec R$ denote that R is preferred to S, $S \sim R$ denote that a person is indifferent between S and R, and $S \succ R$ denote that S is preferred to R. PRAM compares the perception of probabilities with the perception of payoffs and makes the following predictions:

$$\left\langle S \begin{matrix} \prec \\ \sim \\ \succ \end{matrix} R \right\rangle \Leftrightarrow \left\langle \phi(b_S, b_R) \begin{matrix} < \\ = \\ > \end{matrix} \xi(y_R, y_S) \right\rangle. \quad (4)$$

Loomes (2010) used descriptive, across-participants, modal choice ("what did most people choose on this pair?") to provide qualitative evidence in support of PRAM's ability to explain data.[1] There are major shortcomings to both, descriptive methods, and modal choice analyses across participants (see, e.g., Regenwetter et al., 2014, for a recent discussion). Starting more or less with Thurstone's (1927) Law of Comparative Judgment and Luce's famous Choice Axiom in the 1950s (Luce, 1959, 1995), scholars have discussed how to model formally the ubiquitous variability within and between decision makers. There are essentially three classes of probabilistic models ("stochastic specifications," in the terminology of, e.g., Loomes & Sugden, 1995). One class of models typically assumes that a given decision maker has a fixed preference, possibly different from another decision maker's preference, and within-person variability is caused by occasional errors or 'trembles' (e.g., Birnbaum, 2011; Block & Marschak, 1960; Harless & Camerer, 1994; Tversky, 1969). A second class of models assumes that a decision maker has a fixed deterministic strength of preference, and that within-person variability is a monotonic function of the strength of preference, with weak/strong strength of preference entailing high/low variability (e.g., Blavatskyy & Pogrebna, 2010; Hey & Orme, 1994; Luce, 1959; Thurstone, 1927). A third class of models treats preferences and/or utilities themselves as uncertain (e.g., Block & Marschak, 1960; Regenwetter, Dana, & Davis-Stober, 2011). There is a decades-old, and ongoing, discussion about the relationships among these models and their relative merits (see, e.g., Birnbaum, 2011; Blavatskyy & Pogrebna, 2010; Carbone & Hey, 2000; Loomes & Sugden, 1995; Luce & Suppes, 1965; Regenwetter & Marley, 2001; Rieskamp, Busemeyer, & Mellers, 2006; Stott, 2006; Wilcox, 2008). We employ two very general probabilistic models that, jointly, encompass all three major notions of variability. We collect and analyze data at the individual level using novel "order-constrained" likelihood-based inference (Davis-Stober, 2009; Myung, Karabatsos, & Iverson, 2005; Regenwetter et al., 2014; Silvapulle & Sen, 2005).

Stimuli, Models, and Empirical Predictions

Our empirical analysis used 20 lottery pairs (see Table 2) that were embedded in a larger experiment. Pairs 1 to 6 are pairwise combinations of Lotteries F, G, H, and J in the left Marschak–Machina triangle in Figure 6 of Loomes (2010), and Pairs 7 to 12 are pairwise combinations of Lotteries F, G, H, and J in the right Marschak–Machina triangle in the same figure. The incentive structure of the experiment was that one choice was played for real money at the end of each session. We used lower payoffs than Loomes (2010), namely, \$20, \$10, and \$0 for reasons of budgeting and comparability with other experiments in our lab. The payoffs of Pairs 13 to 20 were \$15, \$10, and \$5. These 20 lottery pairs also lead to particularly strong predictions under PRAM. The probabilities of Pairs 13 to 20 were generated by setting $q_2 = 0$ and generating eight randomly selected stimulus pairs with fixed payouts and with the constraints that $q_3 > p_3$, $q_2 > p_2$, and $q_1 > p_1$, so as to satisfy PRAM's requirements.

Two Models of Variability

We use two probabilistic models that, jointly, incorporate three classical modeling approaches to variability in choice. The *error model* assumes that a person has a fixed preference (either $S \prec R$ or $S \succ R$) and occasionally makes errors, but not too often. Let $0 \leq \tau \leq \frac{1}{2}$ denote the upper bound on the probability of an error/tremble. The probability P_{RS} of choosing R over S and $P_{SR} = 1 - P_{RS}$ are given by

$$\left\langle \phi(b_S, b_R) \begin{matrix} < \\ > \end{matrix} \xi(y_R, y_S) \right\rangle \Leftrightarrow \left\langle S \begin{matrix} \prec \\ \succ \end{matrix} R \right\rangle \Leftrightarrow \left\langle \begin{matrix} P_{RS} \\ P_{SR} \end{matrix} \geq 1 - \tau \geq \frac{1}{2} \right\rangle. \quad (5)$$

This model is an example of an "aggregation-based" model in Regenwetter et al. (2014).[2] When $\tau = \frac{1}{2}$ this is (probabilistic, within-person) modal choice, where the decision maker may mistakenly choose the 'wrong' prospect up to 50% of the time, in each prospect pair, up to sampling variability. Most econometric specifications where the choice probability is a monotonic function of the strength of preference, $\phi(b_S, b_R) - \xi(y_R, y_S)$, imply that a decision maker chooses his/her preferred prospect more often than not, hence, they imply the error model (5) with $\tau = \frac{1}{2}$. Similarly, most models that assume fixed and equal error rates across lotteries will imply the error model. In our model, error rates can vary

[1] We understand that the empirical illustration in Loomes (2010) was more a "proof of concept" than a full-fledged empirical test. Future extensions or modifications of PRAM can be tested in similar ways as we have tested this version.

[2] The frequentist data analysis was carried out with the public-domain QTEST software of Regenwetter et al. (2014); the Bayesian analysis used a prototype of the Bayesian extension of QTEST.

Table 1
Seven Different Functional Forms for the Utility c( ) of Money and Their Predictions Under PRAM

Model	Utility for money	Functional form $c(x) =$	Range for ρ (grid search)	# of distinct preference patterns	Random preference predicts deterministic choice for certain pairs		Random preference predicts equal probabilities for certain groups of pairs
					$P_{RS} = 1$	$P_{RS} = 0$	$P_{RS} = P_{R'S'}$
PRAM 2010	Id	x	N/A	1	Pairs 1–6, 13–17, 19, 20	Pairs 7–12, 18	—
	Log	$log(\rho + x)$	$.01 \leq \rho \leq 100$	52°	—	Pairs 7–12, 18	Pairs 1 & 6; 2 & 5
	Pwr	x^ρ	$.01 \leq \rho \leq 100$	66°	—	—	Pairs 1 & 6; 2 & 5; 8 & 11; 7 & 12
	PwrA	x^ρ	$.01 \leq \rho \leq 1$	55°	—	Pairs 7–12, 18	Pairs 1 & 6; 2 & 5
	PwrS	x^ρ	$1 \leq \rho \leq 100$	12°	Pairs 1–6, 13–17, 19, 20	—	Pairs 8 & 11; 7 & 12
	Quad	$\rho x - x^2$	$.01 \leq \rho \leq 100$	89°	—	—	Pairs 1 & 6; 2 & 5; 8 & 11; 7 & 12
	Exp	$1 - e^{-\rho x}$	$.01 \leq \rho \leq 100$	50°	—	Pairs 7–12, 18	Pairs 1 & 6; 2 & 5
PRAM 2006	Id	x	N/A	17°	Pairs 1–6, 13–17, 19, 20	Pair 9	—
	Log	$log(\rho + x)$	$.01 \leq \rho \leq 100$	209°	—	Pair 9	—
	Pwr	x^ρ	$.01 \leq \rho \leq 100$	208°	—	—	—
	PwrA	x^ρ	$.01 \leq \rho \leq 1$	195°	—	Pair 9	—
	PwrS	x^ρ	$1 \leq \rho \leq 100$	30°	Pairs 1–6, 13–17, 19, 20	—	—
	Quad	$\rho x - x^2$	$.01 \leq \rho \leq 100$	453°	—	—	—
	Exp	$1 - e^{-\rho x}$	$.01 \leq \rho \leq 100$	203°	—	Pair 9	—

Note. The number of distinct preference patterns (when marked with °) was computed using a fine-grained grid search, all other results are analytical and do not depend on a grid search. In the utility functions $c(x)$, the grid search considered all values of ρ that are (where applicable) multiples of 0.01 in the range [0.01, 2] and multiples of 1 in the range [2, 100]. For the Perceived Relative Argument Model (PRAM) parameter α, the grid search considered all values in the range [−25, 25] with a step-size of 0.05. For δ, it covered the range [1, 25] with a step-size of 0.05. For the additional parameter β in the 2006 version of PRAM, it considered all multiples of 0.01 in the range [0.01, 2] and all multiples of 1 in the range [2, 100].

across respondents and across stimuli.[3] We report analyses of the error model for upper bounds $\tau = \frac{1}{2}$, $\tau = \frac{1}{4}$, and $\tau = \frac{1}{10}$ on error rates. Figure 1 gives an illustration (top three panels) of this model for Pairs 1, 2, 4, for PRAM with $c(x) = x$.

The *random preference model* allows a decision maker to waiver in his use of different parameters in c, ϕ, and ξ. Here, the parameters ρ, α, and δ of the utility and perception functions in Equations 2–3 become random variables. We place no constraints whatsoever on their joint distribution. For instance, when using the utility function $c(x) = x$, where PRAM only has two parameters α and δ, this joint distribution could look like a bivariate normal, like the illustration in the lower left hand side of Figure 1; or some other distribution, say, like the illustration in the lower right hand side of Figure 1. When the utility function is a power function, the joint distribution of ρ, α, and δ could be any trivariate distribution on the permissible range of values.[4] According to the random preference model, writing *Prob* for the probability measure governing the joint distribution of values of ρ, α, and δ, $P_{RS} = Prob(\{\text{all values } \rho, \alpha, \delta \text{ for which } \phi(b_S, b_R) < \xi(y_R, y_S)\})$.

The stimuli (see Table 2) yield restrictive predictions under the error model, and prohibitively restrictive predictions under the random preference model. In addition to the functional forms for $c( )$, the top half of Table 1 provides the number of distinct preference patterns we found for each of seven different utility functions using, in six cases, a numerical grid search. It also shows some restrictive predictions derived analytically from the random preference specification.

1) If $c(x) = x$, then, as we vary the values of α and δ, the predicted preference according to PRAM never changes in any of our lottery pairs! The table shows this by reporting only one single "Pattern." PRAM always predicts preference of the risky lottery R in Pairs 1–6, 13–17, and 19–20, and always predicts preference for the safe lottery S in Pairs 7–12 and 18, regardless of the parameters α and δ (for an analytical proof, see the online supplemental materials). Therefore, the error model predicts that R must be chosen with high probability in Pairs 1–6, 13–17, and 19–20, and S must be chosen with high probability in Pairs 7–12 and 18, no matter what parameter values for α and δ are used in the perception functions. For example, when we permit no more than 10% error, each person must choose R in each of Pairs 1–6, 13–17, and 19–20 at least 90% of the time, up to sampling variability, and the same person must choose S at least 90% of the time in each of the remaining pairs, up to sampling variability. The random preference model makes an even far stronger prediction: Even though the decision maker can waiver in his or her use of perception functions for probabilities and outcomes by randomly drawing values of α and δ from any bivariate distribution on the range of these parameters, she must always choose R in Pairs 1–6, 13–17, and 19–20, and always choose S in the remaining pairs.

As we move to other utility functions, four of the remaining six functional forms for $c( )$ still lead the random preference model to predict deterministic behavior for some pairs. The other two pre-

[3] Error rates can drift over time, within the given range, as long as the preference remains the same. The parameters ρ, α, and δ can also vary as long as they produce the same preference pattern.

[4] Note that the joint distribution over ρ, α, and δ can vary a great deal without affecting the corresponding distribution over preference patterns, because each preference pattern can be generated by many combinations of parameter values. In the statistical analysis, even a changing distribution over preference patterns is permissible because the average of several random preference models over the same collection of permissible preference states is again a random preference model.

Table 2
Lottery Pairs in the Experiment

Pair	'Safe' lottery (S)						'Risky' lottery (R)					
	x_3	p_3	x_2	p_2	x_1	p_1	x_3	q_3	x_2	q_2	x_1	q_1
1	20	0	10	1	0	0	20	0.2	10	0.75	0	0.05
2	20	0	10	1	0	0	20	0.6	10	0.25	0	0.15
3	20	0	10	1	0	0	20	0.8	10	0	0	0.2
4	20	0.2	10	0.75	0	0.05	20	0.6	10	0.25	0	0.15
5	20	0.2	10	0.75	0	0.05	20	0.8	10	0	0	0.2
6	20	0.6	10	0.25	0	0.15	20	0.8	10	0	0	0.2
7	20	0	10	1	0	0	20	0.05	10	0.75	0	0.2
8	20	0	10	1	0	0	20	0.15	10	0.25	0	0.6
9	20	0	10	1	0	0	20	0.2	10	0	0	0.8
10	20	0.05	10	0.75	0	0.2	20	0.15	10	0.25	0	0.6
11	20	0.05	10	0.75	0	0.2	20	0.2	10	0	0	0.8
12	20	0.15	10	0.25	0	0.6	20	0.2	10	0	0	0.8
13	15	0.39	10	0.33	5	0.28	15	0.68	10	0	5	0.32
14	15	0.16	10	0.47	5	0.37	15	0.56	10	0	5	0.44
15	15	0.36	10	0.50	5	0.14	15	0.76	10	0	5	0.24
16	15	0.385	10	0.404	5	0.211	15	0.70	10	0	5	0.30
17	15	0.3356	10	0.4168	5	0.2467	15	0.72	10	0	5	0.28
18	15	0.384	10	0.431	5	0.185	15	0.58	10	0	5	0.42
19	15	0.495	10	0.442	5	0.063	15	0.76	10	0	5	0.24
20	15	0.395	10	0.267	5	0.338	15	0.659	10	0	5	0.341

dict that certain choice probabilities are constant (but not necessarily 0 or 1) across certain pairs, as we see next.

2) For $c(x) = x^\rho$ with $.01 \leq \rho \leq 1$, a so called "risk averse power utility" representation for money, we found 55 different predicted preference patterns that are compatible with PRAM, depending on the values of ρ, α, and δ that we substituted in c, ϕ, and ξ, using a fine-grained grid search. However, even here, PRAM still always predicts preference for the safe lottery S in Pairs 7–12 and 18, regardless of the parameters ρ, α, and δ within their permissible (continuous) range (for an analytical proof that does not depend on a grid search, see the online supplemental materials). Even though we now permit all trivariate distributions over the values of ρ, α, and δ, the random preference model still predicts deterministic choice for Pairs 7–12 and 18. It also predicts that the probability of choosing R over S is identical in Pairs 1 and 6, as well as in Pairs 2 and 5 (for an analytical proof that does not depend on a grid search, see the online supplemental materials). In this random preference model, a single person can probabilistically use different "risk attitudes" ρ at different moments in time with the constraint that he is risk averse (including the possibility of risk neutral) at all times.

3) As we render the utility function for money even more flexible by dropping the above constraint and allowing also "risk seeking" in the "power utility" representation, that is, for $c(x) = x^\rho$ with $.01 \leq \rho \leq 100$, the random preference model no longer predicts deterministic behavior. This PRAM model, which permits a single person to fluctuate between various "risk seeking" and "risk averse" behaviors by allowing ρ to have any distribution across a very large range, nonetheless still predicts that choice probabilities must be constant across certain pairs. We derived that the probability of choosing R over S is identical in Pairs 1 and 6, in Pairs 2 and 5, in Pairs 8 and 11, as well as in Pairs 7 and 12, but no longer restricted to be either zero or one. This information is shown in the last column of Table 1 for each version of PRAM we considered.

Loomes (2010) mentioned a more general predecessor version of PRAM reported in Loomes (2006). This model, which we refer to as PRAM 2006, featured one more person-specific parameter β in the function $\phi(b_S, b_R)$ of Equation 2, in addition to α. Letting $f = 1 - \frac{p_1}{q_1}$, $g = 1 - \frac{q_2}{p_2}$, $h = 1 - \frac{p_3}{q_3}$, and $\beta \geq 0$, PRAM 2006 assumes that $\phi(b_S, b_R) = (fgh)^\beta \left(\frac{b_S}{b_R}\right)^{(b_S + b_R)^\alpha}$. The bottom half of Table 1 shows the number of predicted preference patterns (using a grid-search) as well as some analytically derived restrictive predictions for the random preference model of PRAM 2006. Two striking features stand out: On the one hand, the extra parameter β leads to a sharp increase in allowable preference states. On the other hand, for five of the seven functional forms of the utility for money, PRAM 2006 nonetheless implies deterministic behavior in random preference models, for certain stimuli (see the online supplemental materials for an analytical proof).

Experiment and Findings

Participants

Altogether, 67 adults (36 males, 31 females) responded to a campus advertisement for a paid study at the University of Illinois at Urbana-Champaign. Of these, 54 returned for a second, identical session, the next day, which served as a replication. Participants gave informed consent before proceeding with the experiment in private rooms. We analyzed the data for each session and each participant separately.

Procedures

The participants made repeated choices (20 times for each pair per session) over lottery pairs that were presented via computers using a two-alternative forced choice (2AFC) paradigm. Each lottery was displayed as a wheel of chance with

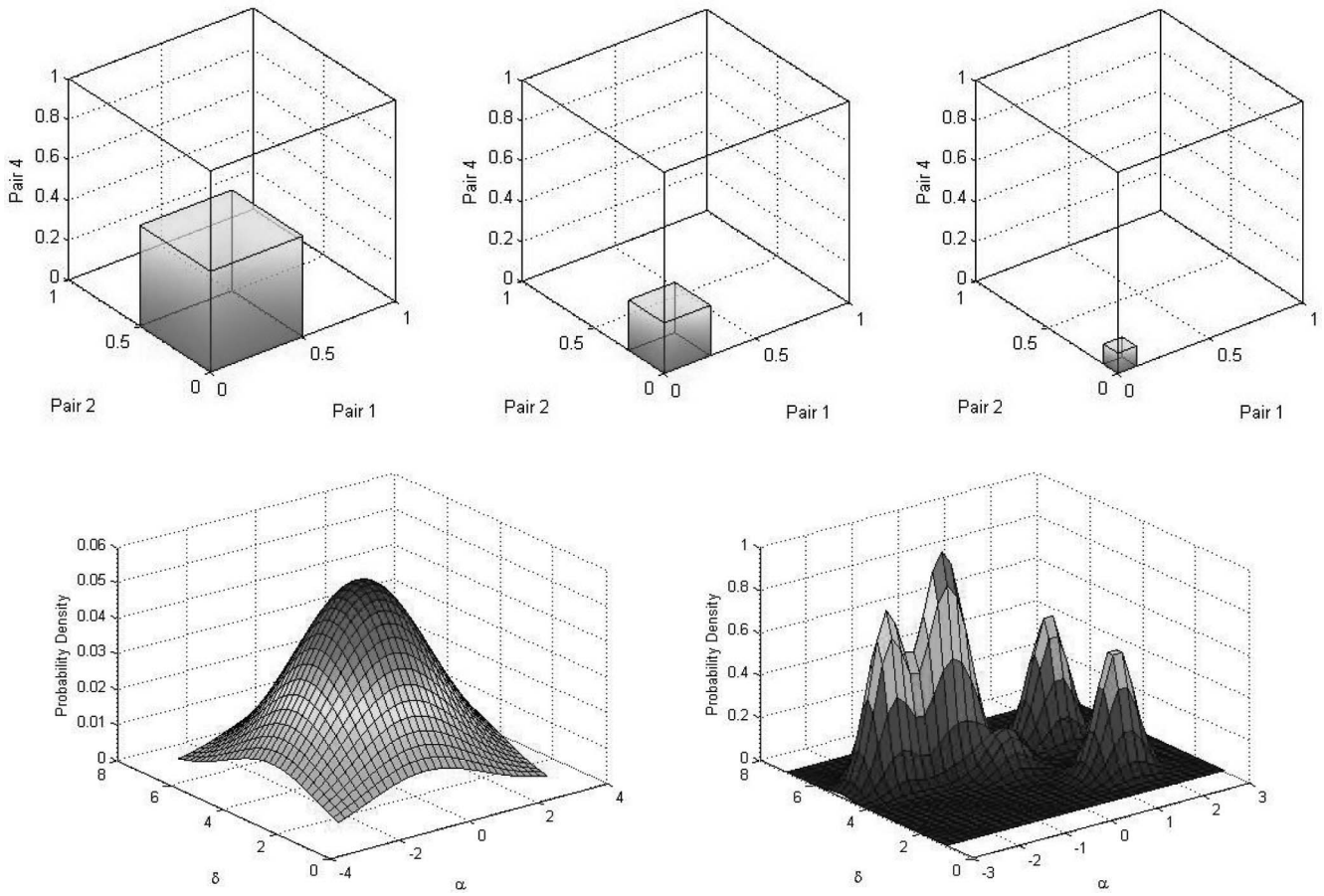

Figure 1. Probabilistic specifications. Top: Error model of the Perceived Relative Argument Model (PRAM), with $c(x) = x$, with three different upper bounds on error rates, $\tau = \frac{1}{2}, \tau = \frac{1}{4}, \tau = \frac{1}{10}$, from left to right. Each axis is a binary choice probability of choosing S in a given pair (for Pairs 1, 2, and 4 in Table 2). The highlighted regions are the choice probabilities permitted by each error model. Each graph shows a 3D projection of a 20-dimensional parameter space. Bottom: Examples of joint distributions of α and δ for a random preference model of PRAM.

colored areas to represent probabilities and numbers next to the wheels as payoffs (see Figure 2). The lottery pairs in Table 2 are only a fraction of all the stimuli used in the entire experiment. There were two other lottery sets (each with 20 pairs) designed for other purposes and serving as distractors for this study. In each session, participants made 1,600 choices, including additional distractors, in total. Repeated presentations of any given lottery pair were separated by at least 80 trials. Trials involving one same lottery (but not the same pair of lotteries) were separated by 3 or more trials.

Before starting the experiment, participants were informed that they had a chance to win a maximum of $31.43 and one of their choices was randomly selected and played for real at the end of each session. The actual payments ranged from $12.28 to $31, per session.[5] This incentivization strategy aimed to elicit true preferences from the participants. The participants first made choices for some lottery pairs in a training session, and they were prompted to ask the host if they had any questions about the task. During the experiment, the participants were encouraged to take breaks as needed. The lottery pairs were drawn randomly from different stimulus sets (that of Table 2 and three others), and the sequence of different sets remained fixed to keep prospects from a given set maximally apart. The experimental paradigm was very similar to that of Regenwetter et al. (2011). The detailed choice frequencies for each participant, lottery pair, and session are provided in the online supplemental materials.

Results

For brevity, all frequentist tests concentrate on PRAM as a Null Hypothesis, and all Bayesian analyses compare PRAM to an unconstrained model. We provide a detailed analysis for PRAM

[5] Telling them that they could win a maximum of $31.43 rather than $31 was an error of the experimenter. The payment range was a result of constraining the random draw to a uniform distribution over trials where the participant would receive at least $12.28, in order to include an implicit show up payment. We thereby circumvented telling participants beforehand that they would earn at least $12.28 no matter what their choices. Regarding these payment amounts, recall that the stimuli in Table 2 are only a fraction of all stimuli used in the experiment.

Figure 2. Screen shot of one trial.

2010 with $c(x) = x^{\rho}$, with $.01 \leq \rho \leq 100$. The online supplemental materials offer 28 different tables detailing the results for all versions. Table 3 shows our frequentist goodness-of-fit findings of the error models for 67 participants. We report results for $\tau = \frac{1}{2}$ (modal choice), for $\tau = \frac{1}{4}$ (no more than 25% errors allowed in each pair, up to sampling variability) and $\tau = \frac{1}{10}$. If a person chooses the predicted option more than 10 times out of 20 when the model predicts that the person chooses that option with probability 0.50, and this holds for every lottery pair, then the error model cannot be rejected no matter what significance level we use in the likelihood test. We call this a *perfect fit*, but it does not imply that the model holds. Perfect fits are indicated by a check mark ✓. The results for modal choice show perfect fits in 20 participants (out of 67) in the first session and 14 participants (out of 54) in the second session. Out of 54 participants who attended both sessions, six are perfectly fit by the error model with $\tau = \frac{1}{2}$ in both sessions (two have inconsistent perfect fits). Rejections at a 0.05 statistical significance level are indicated by *. When PRAM fits the data of a person for both sessions, we give the p-value in `typewriter` and/or ✓ to indicate a consistent fit. By a consistent fit of an error model, we mean that there exists a set of parameter values ρ, α, and δ of PRAM for which the error specification fits in both sessions. In that case, we also say that the two sessions *replicate* each other. When we bound error rates by 25%, 4 out of 54 participants are fit by PRAM consistently in both sessions.[6]

The next table summarizes the (frequentist) performance of the 2010 and 2006 versions of PRAM for seven utility functions, four different probabilistic specifications, and two sessions. Table 4 reports the total number of people who are fit by each PRAM model, by session (out of 67 and 54, respectively), in the top panel, and the number of people who are consistently fit by a model in both sessions (out of 54), in the bottom panel. For example, for the error model with $c(x) = x$, with $\tau = \frac{1}{2}$, that is, modal choice, 28 of 67 people in the first session and 17 of 54 people in the second session are fit at a 0.05 significance level (i.e., 39 and 37 are rejected). Ten are consistently fit across both sessions, that is, 44 are not. As we move to the much more flexible general power function $c(x) = x^{\rho}$, which we also discussed in detail in Table 3, 57 of 67 and 45 of 54 are fit by modal choice in separate analyses, but there appears to be some degree of 'over-fitting' since only 34 out of 54 replicate across sessions (see Harless & Camerer, 1994, for related warnings). Another reason why we may suspect that modal choice involves extensive 'over-fitting' is that the rejection rates leap up when we place stronger restrictions on error rates, that is, $\tau = \frac{1}{4}$ or $\tau = \frac{1}{10}$, and successful replications become extremely rare. All of these findings highlight the importance of individual subject analyses and of replications in this domain.

The Bayesian analysis uses order-constrained methods of Myung et al. (2005). For brevity, we take the simplest nontrivial approach by comparing each model to the "unconstrained" model in which there is no constraint whatsoever on the binary choice probabilities. This is also a conservative approach in that PRAM does not need to compete against alternative theories. We "discard" a PRAM model whenever its Bayesian p-value is smaller than 0.05. If the Bayesian p-value exceeds 0.05, then we compare the model against the unconstrained model according to the Deviance Information Criterion (DIC). We declare a "fit" of a PRAM model when two criteria are met: the Bayesian p-value of the model exceeds 0.05 and the model is favored over the unconstrained model by DIC. In all other cases, we "reject" the model. Table 5 shows that the Bayesian analysis is less forgiving than the frequentist analysis. Even fewer cases are classified as "fits." As we mentioned earlier, the frequentist analysis shows some evidence of 'over-fitting' for modal choice. One interpretation of the Bayesian results is that the Bayesian analysis, which takes model complexity into account, successfully 'punishes' the modal choice model for its flexibility and favors the unconstrained model more often. Overall, however, we find remarkably close alignment between frequentist and Bayesian results throughout, even though they are conceptually and computationally quite distinct (see the 28 analysis tables in the online supplemental materials). We omit further details in the interest of brevity.[7]

The random preference model specifications for PRAM vary widely in their properties, their complexity, and the availability of statistical tests, depending on the utility for money and whether we consider PRAM 2010 or PRAM 2006. As Table 1 shows, the random preference model for $c(x) = x$ predicts completely deterministic behavior. As Tables 4–5 show, this prediction does not hold empirically for even a single participant. Several random preference models predict deterministic behavior for some but not all pairs. This means that some choice is predicted to be deterministic and some is predicted to be probabilistic with order-constraints on the choice probabilities. Unfortunately, to our knowledge, neither frequentist nor Bayesian likelihood based methods are currently available for such models. To the extent that the deterministic predictions are violated, we can give upper bounds on the number of people who could be consistent with these models at best. These are marked with ★ and with ⊗ in Tables 4–5. Some random preference models for PRAM predict nondeterministic choices but imply simultaneous equality and inequality restrictions on choice probabilities. For these, the current state of order-constrained frequentist inference does not yet apply, but the Bayesian methods do apply. This is why Table 4 lists in some

[6] Note that we use advanced customized statistical methods as discussed in Davis-Stober (2009) and Myung et al. (2005).

[7] If we were to compare multiple competing models, and if two of more of these models were to account for the data well, it would be informative to supplement these analyses with Bayes factors to select the best models. We omit model competitions for brevity. Given the poor performance of PRAM, we believe that the Bayesian p-value analysis is sufficient. Computing Bayes factors for restrictive models like these in 20-dimensional space is also computationally expensive and ultimately warranted only for selection among well performing models.

Table 3
Frequentist Results for the Error Model of the Perceived Relative Argument Model (PRAM) 2010 With $c(x) = x^{\rho}$ Where $\rho \in [0.01,100]$ and With $\tau = \frac{1}{2}$ (Within-Person Modal Choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	E1	E2	E1	E2	E1	E2	#	E1	E2	E1	E2	E1	E2
1	0.62	★	★	★	★	★	39	✓	0.92	★	★	★	★
2	★	0.17	★	★	★	★	41	0.22	★	★	★	★	★
4	★	0.05	★	★	★	★	42	0.15	0.38	★	★	★	★
5	0.36	★	★	★	★	★	43	0.82	0.46	★	★	★	★
7	✓	0.49	★	★	★	★	44	★	0.09	★	★	★	★
9	0.35	0.97	★	0.17	★	★	46	0.05	0.70	★	★	★	★
11	✓	0.25	0.30	★	★	★	47	0.56	0.23	★	★	★	★
12	0.96	✓	★	0.06	★	★	48	0.31	0.40	★	★	★	★
13	★	✓	★	0.53	★	★	49	0.45	0.82	★	★	★	★
14	✓	✓	0.98	✓	0.39	0.68	50	✓	0.58	0.06	★	★	★
15	0.56	0.33	★	★	★	★	52	0.72	0.71	★	★	★	★
16	✓	✓	★	0.99	★	0.51	53	0.26	0.34	★	★	★	★
17	✓	✓	✓	✓	0.55	0.89	55	★	★	★	★	★	★
18	0.09	0.98	★	★	★	★	56	0.26	0.80	★	★	★	★
19	★	0.62	★	★	★	★	58	✓	✓	0.37	0.89	★	0.14
20	0.15	✓	★	✓	★	.999	59	0.90	0.95	★	★	★	★
21	0.10	0.22	★	★	★	★	61	0.42	0.56	★	★	★	★
22	✓	0.55	★	★	★	★	65	✓	✓	0.25	0.53	★	★
23	0.92	0.51	★	★	★	★	66	✓	★	0.81	★	★	★
24	0.88	★	★	★	★	★	67	0.87	★	★	★	★	★
25	✓	✓	★	0.35	★	★	3	0.46		★		★	
26	✓	✓	0.61	★	★	★	6	0.88		★		★	
27	0.19	✓	★	★	★	★	8	✓		✓		0.94	
28	0.33	0.79	★	★	★	★	10	✓		★		★	
29	0.74	0.61	★	★	★	★	40	0.55		★		★	
30	0.30	0.92	★	★	★	★	45	★		★		★	
31	★	★	★	★	★	★	51	0.14		★		★	
32	0.78	0.77	★	★	★	★	54	✓		★		★	
33	0.05	0.68	★	★	★	★	57	0.17		★		★	
34	0.19	0.51	★	★	★	★	60	0.66		★		★	
35	★	★	★	★	★	★	62	✓		✓		0.27	
36	✓	✓	0.97	✓	★	0.95	63	✓		0.86		0.14	
37	★	✓	★	0.59	★	★	64	✓		★		★	
38	0.19	✓	★	★	★	★							

Note. Of the 67 participants in the first session, 54 returned (# is the participant ID). Rejections at a 0.05 level are marked ★. Perfect fits are checkmarks (✓). Nonsignificant violations have their p-values listed. Successful replications across sessions are marked in typewriter and/or ✓.

cases that a full test is unavailable where Table 5 reports results. Finally, for one random preference model, the mathematical model was so complex that we did not have its mathematical description ("facet-defining inequalities," aka FDIs) computed at the time of submission to the journal. This is indicated in the tables as "Facet-defining inequalities unavailable." To us, the random preference model with the most promise is the random preference specification of PRAM 2006 with a general power utility function. It does not appear to predict equality constraints on parameters (we could not derive any such constraints analytically and a grid-search based analysis did not reveal any either), and it fits 12 of 54 participants consistently in frequentist tests. However, this is clearly a very general and flexible model: It permits 208 distinct preference states (based on our grid-search) for our stimuli. Yet, the Bayesian analysis does not punish the model severely: It also favors this model over an unconstrained model, consistently across experimental sessions, for 10 participants. Hence, it closely matches the frequentist analysis.

Conclusion

PRAM 2010 and PRAM 2006 are designed as descriptive models of risky choice. Loomes (2010) reported no quantitative predictions or statistical tests. Baillon, Bleichrodt, and Cillo (2012) previously obtained evidence against one version of PRAM using a different approach than ours. We reported a new experiment and an individual-level analysis using both frequentist and Bayesian order-constrained statistical inference. The error model of PRAM generated highly restrictive predictions that our analysis shows to have been systematically violated by the participants. We used two different core formulations of PRAM, combined with seven different utility functions for money. The most lenient error model could consistently account across two sessions for only at best 2/3 of participants in the frequentist tests, and at best about 1/3 of participants in the Bayesian analysis. Error models with error rates bounded above by $\frac{1}{4}$ could only account (with replication) for about 10% of participants. We conclude that error model analyses of PRAM

Table 4
Top Panel: Total Number of Fits for the Error and Random Preference Models of the Perceived Relative Argument Model (PRAM) by Session (E1: First Experimental Session With 67 Respondents; E2: Second Experimental Session With 54 Respondents). Bottom Panel: Total Number of People, Out of 54, Who Are Fit by PRAM in Both Sessions

	Session (# participants)							
	E1 (67)	E2 (54)	E1 (67)	E2 (54)	E1 (67)	E2 (54)	E1 (67)	E2 (54)
	Error model $\tau = \frac{1}{2}$		Error model $\tau = \frac{1}{4}$		Error model $\tau = \frac{1}{10}$		Random preference model	
	Frequentist: Fits (out of 67 or 54)							
PRAM 2010								
Id	28	17	6	1	2	1	0	0
Log	50	41	10	10	3	5	$\leq 2^{\bigstar}$	$\leq 2^{\bigstar}$
Pwr	57	45	12	12	5	6	Full test unavailable	
PwrA	50	40	10	9	3	4	$\leq 2^{\bigstar}$	$\leq 2^{\bigstar}$
PwrS	36	23	8	4	4	3	0	0
Quad	60	47	14	13	5	7	Full test unavailable	
Exp	50	41	10	10	3	5	$\leq 2^{\bigstar}$	$\leq 2^{\bigstar}$
PRAM 2006								
Id	34	21	8	2	4	2	0	0
Log	56	48	17	19	5	7	$\leq 23^{\otimes}$	$\leq 22^{\otimes}$
Pwr	58	50	17	20	5	7	23	23
PwrA	56	47	17	18	5	6	$\leq 23^{\otimes}$	$\leq 22^{\otimes}$
PwrS	36	24	8	4	4	3	0	0
Quad	61	52	18	21	5	8	Facet-defining inequalities unavailable	
Exp	56	48	17	19	5	7	$\leq 23^{\otimes}$	$\leq 22^{\otimes}$
	Error model $\tau = \frac{1}{2}$		Error model $\tau = \frac{1}{4}$		Error model $\tau = \frac{1}{10}$		Random preference model	
	Frequentist: Consistent Fits (out of 54)							
PRAM 2010								
Id	10		1		1		0	
Log	29		3		2		0	
Pwr	34		4		2		Full test unavailable	
PwrA	29		3		2		0	
PwrS	15		2		1		0	
Quad	35		4		2		Full test unavailable	
Exp	29		3		2		0	
PRAM 2006								
Id	14		2		1		0	
Log	39		7		2		$\leq 17^{\otimes}$	
Pwr	41		7		2		12	
PwrA	39		7		2		$\leq 17^{\otimes}$	
PwrS	16		2		1		0	
Quad	42		7		2		Facet-defining inequalities unavailable	
Exp	39		7		2		$\leq 17^{\otimes}$	

Note. This frequentist analysis uses order-constrained likelihood methods of Davis-Stober (2009) with a significance level of 5%. For ★ and ⊗: These constitute upper bounds. In each of these cases, a full statistical test is unavailable. ★ In Session E1 (but not E2), Participants 26 and 67 satisfied the deterministic constraints (in Pairs 7–12, 18), as did Participants 12 and 55 in Session E2 (but not E1). ⊗ The indicated number of people were consistent with the predicted deterministic choice (in Pair 9).

consistently show poor model performance, regardless of whether we use the 2006 or 2010 version, which of seven utility functions for money we use, whether we permit high or low error/tremble rates, and whether we use frequentist or Bayesian tools for evaluating performance. This error model based analysis depends on the assumption that our grid search of the parameter space identified all preference patterns of interest.

Most of our random preference model analysis did not depend on the grid-search: The deterministic choice predictions and the equal probability predictions in Table 1 were derived analytically. All deterministic choice predictions were violated by more than half of the participants. For some models, we currently lack the statistical tools for a full likelihood-based analysis. The Bayesian analysis, which is the most broadly applicable, finds that none of the random preference models for PRAM 2010 accounted consistently for more than three individuals (out of 54). We are not optimistic that the 2006 version does much better, but limitations in statistical tools currently prevent a full-fledged test. We omit partial tests for brevity.

Table 5
Top Panel: Total Number of Fits for the Error and Random Preference Models of the Perceived Relative Argument Model (PRAM) by Session (E1: First Experimental Session With 67 Respondents; E2: Second Experimental Session With 54 Respondents). Bottom Panel: Total Number of People, Out of 54, Who are Fit by PRAM in Both Sessions

	Session (# participants)							
	E1 (67)	E2 (54)	E1 (67)	E2 (54)	E1 (67)	E2 (54)	E1 (67)	E2 (54)
	Error model $\tau = \frac{1}{2}$		Error model $\tau = \frac{1}{4}$		Error model $\tau = \frac{1}{10}$		Random preference model	
	Bayesian "Fits" (out of 67 or 54)							
PRAM 2010								
Id	19	12	4	1	1	1	0	0
Log	28	23	6	7	2	5	$\leq 2^{\bigstar}$	$\leq 2^{\bigstar}$
Pwr	32	29	10	10	4	5	12	8
PwrA	26	25	6	6	2	4	$\leq 2^{\bigstar}$	$\leq 2^{\bigstar}$
PwrS	25	12	8	5	3	2	0	0
Quad	38	28	12	11	4	6	15	11
Exp	28	23	6	7	2	5	$\leq 2^{\bigstar}$	$\leq 2^{\bigstar}$
PRAM 2006								
Id	21	12	8	3	3	1	0	0
Log	37	34	12	13	4	6	$\leq 23^{\otimes}$	$\leq 22^{\otimes}$
Pwr	39	37	12	14	4	6	35	30
PwrA	38	35	12	12	4	5	$\leq 23^{\otimes}$	$\leq 22^{\otimes}$
PwrS	22	13	8	5	3	2	0	0
Quad	42	34	13	14	4	7	Facet-defining inequalities unavailable	
Exp	36	33	12	12	4	6	$\leq 23^{\otimes}$	$\leq 22^{\otimes}$

	Error model $\tau = \frac{1}{2}$	Error model $\tau = \frac{1}{4}$	Error model $\tau = \frac{1}{10}$	Random preference model
	Bayesian: Consistent "Fits" (out of 54)			
PRAM 2010				
Id	7	1	1	0
Log	12	4	2	0
Pwr	15	5	2	2
PwrA	12	4	2	0
PwrS	6	2	1	0
Quad	18	6	2	3
Exp	12	4	2	0
PRAM 2006				
Id	7	2	1	0
Log	22	5	2	$\leq 17^{\otimes}$
Pwr	23	5	2	10
PwrA	23	5	2	$\leq 17^{\otimes}$
PwrS	7	2	1	0
Quad	23	6	2	Facet-defining inequalities unavailable
Exp	21	5	2	$\leq 17^{\otimes}$

Note. This Bayesian analysis uses order-constrained likelihood based methods of Myung et al. (2005). Here, a "fit" is a case where the Bayesian p-value is ≥ 0.05, and the model wins against the unconstrained model by Deviance Information Criterion. All other cases are "rejections." For ★ and ⊗: These constitute upper bounds. In each of these cases, a full Bayesian analysis is unavailable. ★ In Session E1 (but not E2), Participants 26 and 67 satisfied the deterministic constraints (in Pairs 7–12, 18), as did Participants 12 and 55 in Session E2 (but not E1). ⊗ The indicated number of people were consistent with the predicted deterministic choice (in Pair 9).

The history of models in this domain is characterized by step-wise generalizations, for example, the gradual shift from Expected Value via Expected Utility and various other models to Cumulative Prospect Theory and beyond. Naturally, one can consider modifications to the current structure of PRAM and maintain the central idea that attributes are compared separately, and overall comparisons among prospects are secondary. Based on our findings, we are not optimistic that changes in the shape of the utility function for money alone will make much difference. We cannot tell a priori whether either an error model is more likely to accommodate future extensions than a random preference model or vice-versa. Whatever they may be, future extensions and modifications of PRAM can be tested, on an individual subject basis, quantitatively, in the same fashion. We also documented the dangers of overfitting and the need for replication.

References

Baillon, A., Bleichrodt, H., & Cillo, A. (2012). *A tailor-made test of intransitive choice*. Retrieved from http://excen.gsu.edu/fur2012/fullpapers/hbleichrodt.pdf

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comment on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review, 118,* 675–683. doi:10.1037/a0023852

Blavatskyy, P., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25,* 963–986. doi:10.1002/jae.1116

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford, CA: Stanford University Press.

Carbone, E., & Hey, J. D. (2000). Which error story is best? *Journal of Risk and Uncertainty, 20,* 161–176. doi:10.1023/A:1007829024107

Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology, 53,* 1–13. doi:10.1016/j.jmp.2008.08.003

Harless, D. W., & Camerer, C. F. (1994). The predictive value of generalized expected utility theories. *Econometrica, 62,* 1251–1289. doi:10.2307/2951749

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica, 62,* 1291–1326. doi:10.2307/2951750

Loomes, G. (2006). *The improbability of a general, rational and descriptively adequate theory of decision under risk*. Retrieved from http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/loomes/workingpapers/improbabilityasdraftedmarch2006.pdf

Loomes, G. (2010). Modeling choice and valuation in decision experiments. *Psychological Review, 117,* 902–924. doi:10.1037/a0019807

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39,* 641–648. doi:10.1016/0014-2921(94)00071-7

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology, 46,* 1–26. doi:10.1146/annurev.ps.46.020195.000245

Luce, R. D., & Suppes, P. (1965). Preference, utility and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York, NY: Wiley.

Myung, J., Karabatsos, G., & Iverson, G. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology, 49,* 205–225. doi:10.1016/j.jmp.2005.02.004

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 118,* 42–56. doi:10.1037/a0021150

Regenwetter, M., Davis-Stober, C. P., Lim, S. H., Guo, Y., Popova, A., Zwilling, C., . . . Messner, W. (2014). QTEST: Quantitative testing of theories of binary choice. *Decision, 1,* 2–34.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology, 45,* 864–912. doi:10.1006/jmps.2000.1357

Rieskamp, J., Busemeyer, J., & Mellers, B. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature, 44,* 631–661. doi:10.1257/jel.44.3.631

Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: Inequality, order, and shape restrictions*. New York, NY: Wiley.

Stott, H. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32,* 101–130. doi:10.1007/s11166-006-8289-6

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286. doi:10.1037/h0070288

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76,* 31–48. doi:10.1037/h0026750

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J. Cox & G. Harrison (Eds.), *Risk aversion in experiments* (Vol. 12, pp. 197–292). doi:10.1016/S0193-2306(08)00004-5

Received June 28, 2013
Revision received January 10, 2014
Accepted January 12, 2014 ■

Appendix C: Online Supplement Materials for Quantitative tests of the Perceived Relative Argument Model: comment on loomes (2010)

Online Supplement

A Quantitative Test of the Perceived Relative Argument Model

Commentary on Loomes (*Psychological Review*, 2010)

Ying Guo

Michel Regenwetter

Proofs of the analytic results underlying Table 1 of the manuscript.

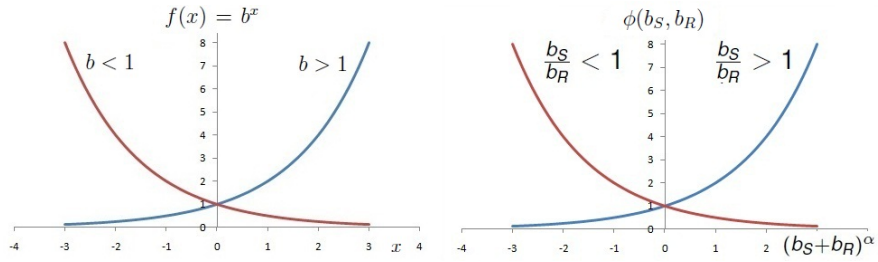

Figure 1: Exponential function b^x, for $b > 0$, as a function of x, when $b > 1$ or $b < 1$ (left graph). Exponential function $\phi(b_S, b_R)$, as a function of $(b_S + b_R)^\alpha$, for $\frac{b_S}{b_R} > 1$ or $\frac{b_S}{b_R} < 1$ (right graph).

To explain PRAM's predictions about preferences among stimuli in Table 1 of the manuscript, we review some properties of the function: $f(x) = b^x$ for $b > 0$, as illustrated by the left panel of Figure 1:

1. $f(x) > 0$;
2. when $b < 1$, $f(x)$ is decreasing;
3. when $b > 1$, $f(x)$ is increasing.

Therefore, the quantity $(b_S + b_R)^\alpha > 0$ and the perception function for probabilities yield the following properties (see also the right side of Figure 1):

$$\phi(b_S, b_R) = \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha} \begin{cases} > 1, & \text{when } \frac{b_S}{b_R} > 1, \\ = 1, & \text{when } \frac{b_S}{b_R} = 1, \\ \in (0,1), & \text{when } 0 < \frac{b_S}{b_R} < 1. \end{cases} \tag{1}$$

1 Deterministic Choice in Random PRAM 2010 with $\xi(y_R, y_S) = 1$

In the perception function of payoffs, if $c(x) = x$, then for Pairs 1 to 12, regardless of the value of δ we obtain $\xi(y_R, y_S) = \left(\frac{20-10}{10-0}\right)^\delta = 1$. For Pairs 13 to 20, $\xi(y_R, y_S) = \left(\frac{15-10}{10-5}\right)^\delta = 1$. So $\xi(y_R, y_S) = 1$ for all 20 choice pairs. Every pairwise preference in the experiment is uniquely determined, no matter what the values α and δ are. In Pairs 1 to 6, 13 to 17 and 19 to 20, $\frac{b_S}{b_R} \in (0,1)$, based on Equation 1, so $\phi(b_S, b_R) < 1$. Since $\xi(y_R, y_S) = 1$, we have $\phi(b_S, b_R) < \xi(y_R, y_S)$ and PRAM predicts the risky lotteries as preferable. By the same token, for Pairs 7 to 12 and 18, where $\frac{b_S}{b_R} > 1$, PRAM predicts the safe lotteries as preferable. ■

2 Deterministic Choice in Random PRAM 2010 with $\xi(y_R, y_S) \leq 1$

For Pairs 7 to 12 and 18, we have $\frac{b_S}{b_R} > 1$, hence, by Equation 1, $\phi(b_S, b_R) > 1$. If we can prove that $\xi(y_R, y_S) \leq 1$, then we obtain $\phi(b_S, b_R) > \xi(y_R, y_S)$. Then, PRAM predicts the safe lotteries as preferable in Pairs 7 to 12 and 18, as reported in Table 1 of the manuscript, for Log, PwrA and Exp. Since $\xi(y_R, y_S) = \left(\frac{c_S}{c_R}\right)^\delta$ with $\delta \geq 1$, we will use the fact that we have $\xi(y_R, y_S) \leq 1 \Leftrightarrow \frac{c_S}{c_R} \leq 1$.

1. $c(x) = log(\rho + x)$ with $\rho > 0$ (Log utility)

   For Pairs 7 to 12 this yields

$$\xi(y_R, y_S) = \left(\frac{log(\rho+20) - log(\rho+10)}{log(\rho+10) - log(\rho+0)}\right)^\delta = \left(\frac{log\left(\frac{\rho+20}{\rho+10}\right)}{log\left(\frac{\rho+10}{\rho}\right)}\right)^\delta = \left(\frac{log\left(1+\frac{10}{\rho+10}\right)}{log\left(1+\frac{10}{\rho}\right)}\right)^\delta.$$

   Since $\rho > 0$, we have $\frac{10}{\rho+10} < \frac{10}{\rho}$. Since $log()$ is increasing, it follows that $\frac{log\left(1+\frac{10}{\rho+10}\right)}{log\left(1+\frac{10}{\rho}\right)} \in (0,1)$, and therefore, likewise, $\xi(y_R, y_S) \in (0,1)$.

For Pair 18, we have similarly

$$\xi(y_R, y_S) = \left(\frac{log\left(1 + \frac{5}{\rho+10}\right)}{log\left(1 + \frac{5}{\rho+5}\right)} \right)^{\delta}.$$

Since $\rho > 0$, it follows that $\frac{log\left(1+\frac{5}{\rho+10}\right)}{log\left(1+\frac{5}{\rho+5}\right)} \in (0,1)$ in a similar way as the previous result, and therefore, likewise, $\xi(y_R, y_S) \in (0,1)$. ■

2. $c(x) = x^{\rho}$ with $0 < \rho \leq 1$ (PwrA utility)

For Pairs 7 to 12 this yields

$$\xi(y_R, y_S) = \left(\frac{20^{\rho} - 10^{\rho}}{10^{\rho} - 0^{\rho}} \right)^{\delta} = (2^{\rho} - 1)^{\delta}.$$

Since $\rho \in (0,1]$, $2^0 - 1 = 0, 2^1 - 1 = 1$, and since $(2^{\rho} - 1)$ is increasing , it follows that $2^{\rho} - 1 \leq 1$, and therefore, likewise, $\xi(y_R, y_S) \leq 1$.

For Pair 18.

$$\xi(y_R, y_S) = \left(\frac{15^{\rho} - 10^{\rho}}{10^{\rho} - 5^{\rho}} \right)^{\delta} = \left(\frac{15^{\rho} - 5^{\rho}}{10^{\rho} - 5^{\rho}} - 1 \right)^{\delta} = \left(\frac{3^{\rho} - 1}{2^{\rho} - 1} - 1 \right)^{\delta}.$$

Since 3^{ρ} increases faster than 2^{ρ}, the function $\frac{3^{\rho}-1}{2^{\rho}-1} - 1$ is increasing in ρ with an upper bound of $\frac{3^1-1}{2^1-1} - 1 = 1$ when $\rho \leq 1$. Therefore, $\frac{3^{\rho}-1}{2^{\rho}-1} - 1 \leq 1$ and consequently $\xi(y_R, y_S) \leq 1$. ■

3. $c(x) = 1 - e^{-\rho x}$ with $\rho > 0$ (Exp utility)

For Pairs 7 to 12 this yields

$$\xi(y_R, y_S) = \left(\frac{(1 - e^{-20\rho}) - (1 - e^{-10\rho})}{(1 - e^{-10\rho}) - (1 - e^{-0\rho})} \right)^{\delta} = \left(\frac{-e^{-20\rho} + e^{-10\rho}}{-e^{-10\rho} + e^{-0\rho}} \right)^{\delta} = \left(\frac{e^{-10\rho}(-e^{-10\rho} + 1)}{-e^{-10\rho} + 1} \right)^{\delta} = e^{-10\rho\delta}.$$

The function $e^{-10\rho}$ is decreasing in ρ with a lower bound of 0 and an upper bound of 1 when $\rho > 0$. Therefore, $e^{-10\rho} \in (0,1)$ and hence, $\xi(y_R, y_S) \in (0,1)$.

For Pair 18, by a similar argument,

$$\xi(y_R, y_S) = \left(\frac{(1 - e^{-15\rho}) - (1 - e^{-10\rho})}{(1 - e^{-10\rho}) - (1 - e^{-5\rho})} \right)^{\delta} = \left(\frac{e^{-10\rho}(e^{-5\rho} - 1)}{e^{-5\rho}(e^{-5\rho} - 1)} \right)^{\delta} = e^{-5\rho\delta},$$

and $e^{-5\rho} \in (0,1)$, since $\rho > 0$. As a result, $\xi(y_R, y_S) \in (0,1)$. ■

3 Deterministic Choice in Random PRAM 2010 with $\xi(y_R, y_S) > 1$

Based on Equation 1, for Pairs 1 to 6, 13 to 17, 19 and 20, $0 < \frac{b_S}{b_R} < 1$, hence $\phi(b_S, b_R) < 1$. If we can prove that $\xi(y_R, y_S) > 1$, then we obtain $\phi(b_S, b_R) < \xi(y_R, y_S)$. Then, PRAM predicts the risky lotteries as preferable in Pairs 1 to 6, 13 to 17, 19 and 20 as reported in Table 1 of the manuscript, for PwrS. Since $\xi(y_R, y_S) = \left(\frac{c_S}{c_R}\right)^{\delta}$ with $\delta \geq 1$, we have $\xi(y_R, y_S) > 1 \Leftrightarrow \frac{c_S}{c_R} > 1$.

We only need to consider $c(x) = x^{\rho}$ with $\rho > 1$ (PwrS utility)

For Pairs 1 to 6 this yields

$$\xi(y_R, y_S) = (2^{\rho} - 1)^{\delta}.$$

Since $2^{\rho} - 1$ is an increasing function with lower bound $2^1 - 1 = 1$ when $\rho > 1$, it follows that $2^{\rho} - 1 > 1$ and $\xi(y_R, y_S) > 1$.

By the same token, for Pairs 13 to 17, 19 and 20,

$$\xi(y_R, y_S) = \left(\frac{3^{\rho} - 1}{2^{\rho} - 1} - 1\right)^{\delta}.$$

Since 3^{ρ} increases faster than 2^{ρ}, the function $\frac{3^{\rho}-1}{2^{\rho}-1} - 1$ is increasing in ρ with a lower bound of $\frac{3^1-1}{2^1-1} - 1 = 1$ when $\rho > 1$. Therefore, $\frac{3^{\rho}-1}{2^{\rho}-1} - 1 > 1$ and $\xi(y_R, y_S) > 1$. ■

4 Equal Probabilities in Random PRAM 2010

We show four other strong implications of the random preference model for PRAM 2010 with arbitrary utility function, listed in Table 1 of the manuscript. For both Pair 1 and Pair 6, $b_S = .05$ and $b_R = .2$. Therefore,

$$\phi(b_S, b_R) = \left(\frac{b_S}{b_R}\right)^{(b_S + b_R)^{\alpha}} = \left(\frac{.05}{.2}\right)^{(.05+.2)^{\alpha}} = \left(\frac{1}{4}\right)^{.25^{\alpha}}.$$

Both lottery pairs have the same payoffs: $x_3 = \$20$, $x_2 = \$10$ and $x_1 = \$0$, thus they give the same value in the perception function for payoffs $\xi(y_R, y_S)$, regardless of the value of δ and regardless of the utility function for money. Therefore, in Pairs 1 and 6, PRAM 2010 uses the same values for $\phi(b_S, b_R)$ and $\xi(y_R, y_S)$, regardless of the values of α, δ and regardless of the utility function. Hence, for any joint distribution of the values of the parameters α, δ, ρ in PRAM 2010, the random preference model of PRAM predicts the same

probability of choosing S over R for Pairs 1 and 6, no matter what the utility function is. Likewise, for Pairs 2 and 5, PRAM uses the same $\phi(b_S, b_R) = \left(\frac{1}{4}\right)^{.75^\alpha}$ and the same $\xi(y_R, y_S)$. Therefore, it predicts the same choice probability for Pairs 2 and 5. Analogous reasoning applies for Pairs 8 and 11 (with $\phi(b_S, b_R) = 4^{.75^\alpha}$), as well as for Pairs 7 and 12 (with $\phi(b_S, b_R) = 4^{.25^\alpha}$). ■

5 Random PRAM 2006

The 2006 version of PRAM has one more person-specific parameter β in the function $\phi(b_S, b_R)$, in addition to α. Let $f = 1 - \frac{p_1}{q_1}$, $g = 1 - \frac{q_2}{p_2}$, $h = 1 - \frac{p_3}{q_3}$ and $\beta \geq 0$. In this version,

$$\phi(b_S, b_R) = (fgh)^\beta \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha}.$$

In Pair 9, PRAM 2006 reduces to PRAM 2010, regardless of β:

$$(fgh)^\beta \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha} = \left[\left(1 - \frac{0}{.8}\right)\left(1 - \frac{0}{1}\right)\left(1 - \frac{0}{.2}\right)\right]^\beta \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha} = \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha}.$$

Hence, for Pair 9, all the above results of Random PRAM 2010 carry over to Random PRAM 2006.

For Pairs 1-6, 13-17, 19 and 20, we proved that $\left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha} < 1$. Since PRAM requires that $q_3 > p_3, q_2 < p_2, q_1 > p_1$ we have $0 < f, g, h \leq 1$, and therefore, likewise, $0 < fgh \leq 1$. Thus, $(fgh)^\beta \left(\frac{b_S}{b_R}\right)^{(b_S+b_R)^\alpha} < 1$. As a consequence, for Pairs 1-6, 13-17, 19 and 20, $\phi(b_S, b_R) < \xi(y_R, y_S)$, following the same proof as that used in Section 3 for Random PRAM 2010. ■

Data.

Table 1: Data from the experiment. The number in each cell is the observed frequency of choosing S over R. The rows are the participants and the columns are gamble pairs. Each gamble pair appears twice, the left instance refers to the first session, the right instance refers to the second session. The last 13 participants in the table only attended the first session. The total number of repetitions for each gamble pair is 20 per session.

	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18	18	19	19	20	20
1	13	18	18	19	13	17	13	14	12	4	7	3	19	20	20	20	20	20	19	19	20	20	20	20	7	1	9	5	9	8	11	9	8	2	16	16	18	16	2	0
2	5	8	4	8	3	9	8	9	5	8	6	4	5	5	18	12	16	12	15	15	14	16	17	13	4	4	4	8	5	6	5	7	6	7	17	11	4	7	5	4
4	3	4	5	9	6	4	19	15	19	14	5	8	18	16	18	18	19	20	18	18	20	18	20	20	8	7	7	6	15	9	17	11	8	5	17	17	17	15	3	9
5	4	0	1	2	0	1	1	0	2	0	1	1	6	2	19	20	20	20	20	18	20	18	20	19	3	2	4	6	2	0	4	3	2	3	18	8	5	3	1	4
7	10	13	12	8	15	13	12	11	14	12	19	15	12	17	19	18	19	20	15	12	19	16	19	18	13	13	11	8	18	6	17	11	13	8	19	14	15	14	15	10
9	1	3	7	15	19	20	15	19	19	20	9	5	19	20	20	20	20	20	20	18	20	19	20	20	1	1	7	1	18	17	14	11	2	1	20	20	20	19	1	1
11	7	7	3	5	2	6	4	10	2	8	3	5	18	5	19	12	18	10	19	12	18	11	18	14	4	9	4	10	3	6	7	4	3	9	14	9	6	7	2	12
12	8	10	9	14	13	18	8	12	15	11	2	0	19	20	20	20	20	20	20	20	20	20	20	20	4	2	7	2	11	14	11	9	4	2	20	20	20	20	1	0
13	2	0	1	1	3	2	0	5	2	3	1	5	2	3	3	1	4	0	3	6	5	6	16	9	1	2	1	3	1	1	0	5	2	0	1	6	0	2	1	5
14	1	2	2	4	3	1	0	1	0	0	2	0	20	20	20	20	20	20	19	19	20	20	20	19	0	0	0	2	0	0	0	2	0	0	13	18	2	5	0	0
15	4	9	8	16	11	15	11	18	13	17	9	6	14	14	15	16	14	15	16	16	17	14	16	17	7	4	8	6	10	14	9	11	10	7	11	11	15	15	6	8
16	6	18	11	20	10	20	11	20	11	19	5	19	14	20	13	20	16	18	16	20	13	18	17	19	6	19	4	19	4	20	9	20	7	19	12	20	11	20	9	14
17	16	20	19	19	19	20	18	18	20	19	18	18	18	20	18	18	18	20	19	18	19	19	18	19	16	16	20	20	19	17	18	20	18	19	19	18	19	20	16	17
18	19	17	19	20	19	20	18	20	14	20	13	18	20	20	20	20	20	20	20	20	20	19	20	19	4	6	5	4	5	8	4	7	4	9	4	9	8	12	5	7
19	3	16	7	13	5	17	12	13	12	15	11	11	5	14	4	17	2	11	11	11	10	12	12	14	10	11	8	11	14	10	12	9	14	14	17	11	13	9	8	13
20	0	0	0	0	0	0	0	1	2	0	1	0	0	1	0	1	2	0	9	0	9	0	16	2	1	0	1	0	1	0	0	0	1	0	0	0	2	1	1	0
21	0	1	0	0	0	0	1	1	0	1	1	4	1	1	6	1	9	1	7	2	7	5	17	15	0	1	0	0	0	1	1	1	0	0	1	0	0	1	0	0
22	10	12	7	7	7	9	6	11	6	9	5	7	14	12	17	10	17	12	16	8	13	9	15	11	6	8	6	6	4	8	5	8	3	9	12	7	10	7	3	11
23	12	10	11	13	12	11	10	11	10	10	10	9	12	14	11	17	19	15	12	13	11	7	14	11	11	12	8	14	9	11	9	12	10	13	10	13	13	14	9	10
24	11	19	20	20	19	20	12	12	11	6	5	0	20	20	20	19	20	20	18	20	18	19	12	16	3	4	3	2	2	1	4	1	2	0	11	8	9	5	3	0
25	3	10	18	18	18	20	10	20	15	20	5	13	19	20	20	20	20	20	19	15	20	19	20	19	9	20	9	18	8	15	10	18	7	19	19	18	11	15	7	18
26	18	15	19	20	19	20	20	19	20	20	17	20	20	20	20	20	20	20	20	20	20	20	20	20	3	10	5	9	9	10	13	12	5	6	20	17	18	16	2	9
27	5	6	3	5	3	6	5	9	4	5	4	8	8	11	16	11	15	15	15	12	17	14	16	14	4	10	7	6	4	6	5	6	4	6	5	7	6	5	1	8
28	12	14	13	14	10	12	10	10	12	10	7	11	7	13	12	8	13	9	11	11	6	12	11	14	8	4	13	11	9	10	10	9	7	8	5	14	7	11	10	10
29	13	14	12	17	13	18	10	8	11	12	9	16	16	19	14	18	11	18	12	10	14	16	13	19	13	10	14	19	12	15	12	14	11	14	14	11	11	10	10	14
30	9	10	8	9	10	11	7	9	3	10	8	11	11	10	9	9	14	11	16	16	15	9	15	10	12	7	12	9	13	11	7	11	12	9	14	9	11	11	7	8
31	18	17	19	16	18	16	2	5	2	3	0	6	19	17	16	19	19	17	18	16	19	14	18	10	3	2	5	2	3	4	6	5	4	4	7	7	4	7	4	4
32	0	14	0	0	0	0	0	0	4	2	0	0	6	17	20	20	20	20	20	20	20	20	20	20	2	0	2	1	1	1	0	1	2	0	17	18	10	13	1	0
33	18	19	18	20	19	20	13	19	16	20	3	6	20	20	20	20	20	20	20	19	20	20	20	20	5	0	8	0	10	12	6	6	3	2	18	15	16	15	0	1
34	17	10	14	6	14	9	6	9	9	12	10	12	18	11	17	8	16	11	16	13	17	11	16	13	6	9	7	8	8	7	6	12	7	14	10	12	13	11	9	9
35	0	2	0	3	1	5	17	11	14	10	10	4	4	2	20	19	20	20	18	16	19	18	20	16	0	4	6	2	5	4	4	4	3	5	8	5	12	6	4	3

Continued on next page

Table 1 – continued from previous page

	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18	18	19	19	20	20
36	20	20	20	20	18	20	19	19	20	20	17	20	20	20	19	20	19	20	19	20	19	20	20	19	16	18	15	16	18	20	16	20	15	19	17	19	15	19	6	2
37	0	15	0	17	0	17	20	16	19	14	19	17	0	16	0	17	0	16	19	16	19	18	17	14	18	16	19	18	19	16	19	17	19	13	19	15	19	16	19	14
38	8	3	4	3	6	3	8	6	12	4	7	3	10	17	18	18	18	18	17	18	18	19	15	16	11	3	13	9	11	7	14	10	5	4	16	17	12	17	6	2
39	12	10	16	11	11	13	12	12	13	9	11	9	13	11	12	13	12	11	11	12	13	13	12	12	11	10	12	11	12	10	13	14	17	10	10	11	16	10	13	9
41	15	18	11	19	14	19	6	11	9	6	13	4	17	20	15	20	18	19	16	13	20	20	17	19	6	9	6	4	5	2	12	7	6	8	9	9	9	12	8	5
42	15	18	18	20	20	19	11	7	6	12	1	7	20	20	20	18	20	20	20	19	20	20	20	20	1	2	0	1	2	6	2	6	0	5	7	12	5	13	1	3
43	4	6	11	4	6	6	4	8	7	3	11	9	14	8	16	8	17	7	14	9	17	10	17	15	10	9	6	2	4	4	7	12	3	8	10	9	10	9	9	7
44	2	4	2	1	0	0	4	3	3	0	6	2	0	4	18	17	19	19	18	19	20	18	18	20	3	0	10	0	6	1	7	5	5	2	14	11	8	3	5	1
46	6	2	5	0	1	1	14	6	15	6	8	4	13	9	19	20	20	19	17	16	20	18	19	15	9	4	8	2	13	3	13	2	9	3	18	4	15	2	9	3
47	12	19	6	13	10	13	8	11	4	8	6	8	20	14	18	16	19	16	19	17	18	17	19	11	3	6	4	9	5	11	4	5	3	13	16	9	9	7	3	8
48	11	12	15	9	13	12	16	6	14	11	7	10	16	15	17	10	16	11	13	15	12	10	14	12	13	11	11	12	7	10	14	9	8	14	9	9	12	7	9	9
49	4	10	5	8	5	7	9	11	6	9	9	8	8	5	8	8	12	10	5	9	14	6	11	11	10	8	9	9	4	10	7	8	6	7	7	6	8	7	5	7
50	13	12	12	10	17	16	15	9	14	10	10	11	18	13	15	12	16	12	15	11	15	11	16	14	15	7	14	12	13	11	13	9	14	10	17	15	13	10	12	13
52	11	10	8	8	8	11	10	10	10	10	10	10	8	10	9	11	10	14	12	12	16	6	13	9	7	15	13	10	9	10	10	10	9	9	8	12	12	9	10	9
53	5	8	4	4	8	9	4	11	4	12	10	14	3	12	11	9	14	9	15	10	14	13	13	8	6	9	6	13	7	12	7	6	6	11	11	11	5	8	4	15
55	2	4	0	19	18	20	0	16	7	19	0	1	20	20	20	20	20	20	19	20	20	20	19	20	1	2	2	1	14	15	19	19	0	0	20	20	19	20	1	0
56	11	15	7	15	12	15	8	10	12	9	11	13	17	10	5	10	11	15	14	10	9	10	13	10	14	10	11	10	8	9	9	7	12	13	8	8	10	10	11	8
58	4	0	0	1	2	1	0	1	3	0	2	0	20	20	20	19	20	20	20	20	19	19	19	20	7	3	4	1	5	1	7	2	2	0	9	7	6	1	4	3
59	10	10	6	10	8	7	5	8	10	9	10	9	10	5	9	7	7	12	10	10	13	11	8	10	10	7	9	8	9	9	11	10	8	11	12	7	9	10	10	8
61	7	12	17	9	16	17	14	9	15	9	8	9	18	18	17	12	20	14	18	7	20	12	18	10	7	9	11	9	12	11	15	13	10	11	17	8	16	11	4	11
65	7	0	5	2	6	1	1	1	3	0	5	1	18	11	16	20	15	20	16	16	17	20	16	20	3	2	8	0	7	2	3	3	6	0	15	19	8	14	3	1
66	0	0	0	1	0	0	0	11	0	15	1	2	20	20	20	20	20	20	20	19	20	20	20	20	0	2	0	11	1	15	0	18	0	2	11	20	7	20	0	0
67	13	18	3	15	2	4	7	13	8	8	0	1	20	20	20	20	20	20	20	20	20	20	20	20	2	2	2	0	4	4	11	10	0	1	20	19	19	15	0	2
3	13		10		14		9		8		9		13		11		16		16		13		15		4		8		6		5		5		8		5		5	
6	0		0		1		0		1		8		0		0		0		0		2		12		2		2		6		4		1		6		3		4	
8	0		0		1		0		0		4		19		20		20		17		19		20		0		0		0		0		0		2		1		0	
10	16		17		16		16		14		12		19		19		17		12		15		16		11		14		14		15		10		14		14		10	
40	9		14		9		7		9		11		10		13		14		9		13		14		12		8		11		9		7		10		8		13	
45	1		12		17		17		19		16		17		20		20		19		18		19		15		15		19		20		15		20		20		15	
51	4		10		8		9		10		7		8		7		7		9		7		9		14		6		14		8		9		9		7		10	
54	2		0		0		2		3		6		14		17		18		17		15		19		6		10		4		8		4		12		10		8	
57	8		9		3		6		11		12		10		5		11		12		14		12		6		7		8		6		7		8		8		9	
60	9		5		9		5		8		9		14		10		11		12		13		12		8		10		6		12		11		13		7		7	
62	0		0		0		0		1		0		17		20		20		18		19		15		1		2		1		2		1		4		5		0	
63	0		0		1		1		1		4		18		20		20		19		20		20		2		7		1		2		1		17		2		0	
64	8		6		6		5		10		5		14		18		14		12		16		18		8		4		5		9		5		13		10		7	

Frequentist and Bayesian error model analyses of PRAM 2010 and PRAM 2006 with seven utility functions for money.

Table 2: Frequentist results for the error model of PRAM 2010 with $c(x) = x$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	0.29	$\star$	$\star$	$\star$	$\star$
2	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	0.35	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.27`	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.66	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.12`	0.30	$\star$	$\star$	$\star$	47	0.49	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	0.34	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.67`	50	$\star$	0.55	$\star$	$\star$	$\star$	$\star$
15	0.73	$\star$	$\star$	$\star$	$\star$	$\star$	52	`0.57`	`0.20`	$\star$	$\star$	$\star$	$\star$
16	0.71	$\star$	$\star$	$\star$	$\star$	$\star$	53	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	0.61	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	`0.95`	`0.90`	0.12	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	`0.71`	`0.12`	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.33`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	`0.57`	0.25	$\star$	$\star$	$\star$
23	0.55	$\star$	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	0.65		$\star$		$\star$	
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	0.50	$\star$	$\star$	$\star$	$\star$	8	0.53		$\star$		$\star$	
28	$\star$	0.19	$\star$	$\star$	$\star$	$\star$	10	$\star$		$\star$		$\star$	
29	0.84	$\star$	$\star$	$\star$	$\star$	$\star$	40	0.18		$\star$		$\star$	
30	`0.28`	`0.91`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	`0.77`	`0.62`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	0.24		$\star$		$\star$	
34	$\star$	0.13	$\star$	$\star$	$\star$	$\star$	60	0.57		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	0.15		$\star$		$\star$	
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	$\surd$		0.16		0.15	
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.11	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 3: Frequentist results for the error model of PRAM 2010 with $c(x) = log(\rho + x)$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.62	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\surd$	`0.92`	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.06	$\star$	$\star$	$\star$	$\star$	41	0.22	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	`0.15`	`0.38`	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.31`	$\star$	$\star$	$\star$	$\star$
7	$\surd$	`0.49`	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.07	$\star$	$\star$	$\star$	$\star$
9	`0.72`	$\surd$	$\star$	0.56	$\star$	$\star$	46	0.05	$\star$	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.12`	0.30	$\star$	$\star$	$\star$	47	`0.56`	`0.15`	$\star$	$\star$	$\star$	$\star$
12	`0.96`	$\surd$	$\star$	0.06	$\star$	$\star$	48	`0.24`	`0.37`	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\surd$	`0.58`	0.06	$\star$	$\star$	$\star$
15	`0.56`	$\surd$	$\star$	$\star$	$\star$	$\star$	52	`0.72`	`0.71`	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	$\star$	0.34	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	`0.55`	`0.89`	55	$\star$	$\surd$	$\star$	$\surd$	$\star$	0.46
18	`0.09`	`0.98`	$\star$	$\star$	$\star$	$\star$	56	`0.12`	`0.80`	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	`0.95`	`0.78`	0.10	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	`0.74`	`0.16`	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\surd$	`0.56`	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.38`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	`0.92`	`0.51`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	0.88	$\star$	$\star$	$\star$	$\star$	$\star$	67	0.87	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	0.24		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	0.19	$\star$	$\star$	$\star$	$\star$	8	0.05		$\star$		$\star$	
28	$\star$	0.79	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	`0.74`	`0.61`	$\star$	$\star$	$\star$	$\star$	40	0.52		$\star$		$\star$	
30	`0.30`	`0.92`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.09		$\star$		$\star$	
32	`0.78`	`0.77`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	`0.05`	`0.68`	$\star$	$\star$	$\star$	$\star$	57	0.06		$\star$		$\star$	
34	`0.19`	`0.51`	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	0.15		$\star$		$\star$	
36	$\surd$	$\surd$	`0.97`	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	`0.19`	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 4: Frequentist results for the error model of PRAM 2010 with $c(x) = x^\rho$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.62	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\surd$	0.92	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.17	$\star$	$\star$	$\star$	$\star$	41	0.22	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	0.15	0.38	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	0.82	0.46	$\star$	$\star$	$\star$	$\star$
7	$\surd$	0.49	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	0.35	0.97	$\star$	0.17	$\star$	$\star$	46	0.05	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	0.25	0.30	$\star$	$\star$	$\star$	47	0.56	0.23	$\star$	$\star$	$\star$	$\star$
12	0.96	$\surd$	$\star$	0.06	$\star$	$\star$	48	0.31	0.40	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\surd$	$\star$	0.53	$\star$	$\star$	49	0.45	0.82	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	0.98	$\surd$	0.39	0.68	50	$\surd$	0.58	0.06	$\star$	$\star$	$\star$
15	0.56	0.33	$\star$	$\star$	$\star$	$\star$	52	0.72	0.71	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	0.26	0.34	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	0.55	0.89	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	0.09	0.98	$\star$	$\star$	$\star$	$\star$	56	0.26	0.80	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	0.37	0.89	$\star$	0.14
20	0.15	$\surd$	$\star$	$\surd$	$\star$	1.00	59	0.90	0.95	$\star$	$\star$	$\star$	$\star$
21	0.10	0.22	$\star$	$\star$	$\star$	$\star$	61	0.42	0.56	$\star$	$\star$	$\star$	$\star$
22	$\surd$	0.55	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	0.92	0.51	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	0.88	$\star$	$\star$	$\star$	$\star$	$\star$	67	0.87	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	0.46		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	0.88		$\star$		$\star$	
27	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	0.33	0.79	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	0.74	0.61	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	0.30	0.92	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.14		$\star$		$\star$	
32	0.78	0.77	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	0.05	0.68	$\star$	$\star$	$\star$	$\star$	57	0.17		$\star$		$\star$	
34	0.19	0.51	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 5: Frequentist results for the error model of PRAM 2010 with $c(x) = x^\rho$ where $\rho \in [.01, 1]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.62	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\surd$	`0.92`	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.06	$\star$	$\star$	$\star$	$\star$	41	0.22	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	0.15	0.38	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.31`	$\star$	$\star$	$\star$	$\star$
7	$\surd$	`0.49`	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.07	$\star$	$\star$	$\star$	$\star$
9	`0.35`	`0.97`	$\star$	0.17	$\star$	$\star$	46	0.05	$\star$	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.12`	0.30	$\star$	$\star$	$\star$	47	`0.56`	`0.23`	$\star$	$\star$	$\star$	$\star$
12	`0.96`	$\surd$	$\star$	0.06	$\star$	$\star$	48	`0.31`	`0.40`	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\surd$	`0.58`	0.06	$\star$	$\star$	$\star$
15	`0.56`	`0.33`	$\star$	$\star$	$\star$	$\star$	52	`0.72`	`0.71`	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	$\star$	0.34	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	`0.55`	`0.89`	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	`0.09`	`0.98`	$\star$	$\star$	$\star$	$\star$	56	`0.12`	`0.80`	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	`0.95`	`0.78`	0.10	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	`0.74`	`0.16`	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	`0.42`	`0.56`	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.38`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	`0.92`	`0.51`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	0.88	$\star$	$\star$	$\star$	$\star$	$\star$	67	0.87	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	0.46		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	0.19	$\star$	$\star$	$\star$	$\star$	8	0.05		$\star$		$\star$	
28	$\star$	0.79	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	`0.74`	`0.61`	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	`0.30`	`0.92`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.09		$\star$		$\star$	
32	`0.78`	`0.77`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	`0.05`	`0.68`	$\star$	$\star$	$\star$	$\star$	57	0.08		$\star$		$\star$	
34	`0.19`	`0.51`	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	0.15		$\star$		$\star$	
36	$\surd$	$\surd$	`0.97`	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	`0.19`	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 6: Frequentist results for the error model of PRAM 2010 with $c(x) = x^\rho$ where $\rho \in [1.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	0.36	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.17	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.46`	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	0.25	0.30	$\star$	$\star$	$\star$	47	0.50	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	0.38	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\surd$	$\star$	0.53	$\star$	$\star$	49	`0.45`	`0.82`	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\star$	0.06	$\star$	$\star$	$\star$	$\star$
15	0.09	$\star$	$\star$	$\star$	$\star$	$\star$	52	`0.61`	`0.36`	$\star$	$\star$	$\star$	$\star$
16	0.71	$\star$	$\star$	$\star$	$\star$	$\star$	53	0.26	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	0.26	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	`0.37`	`0.89`	$\star$	0.14
20	`0.15`	$\surd$	$\star$	$\surd$	$\star$	1.00	59	0.90	0.95	$\star$	$\star$	$\star$	$\star$
21	`0.10`	`0.22`	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.55`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	`0.56`	0.25	$\star$	$\star$	$\star$
23	0.56	$\star$	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	0.07		$\star$		$\star$	
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	0.88		$\star$		$\star$	
27	`0.19`	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	`0.33`	`0.19`	$\star$	$\star$	$\star$	$\star$	10	$\star$		$\star$		$\star$	
29	0.10	$\star$	$\star$	$\star$	$\star$	$\star$	40	0.18		$\star$		$\star$	
30	`0.28`	`0.92`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.14		$\star$		$\star$	
32	`0.78`	`0.62`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	0.17		$\star$		$\star$	
34	$\star$	0.20	$\star$	$\star$	$\star$	$\star$	60	0.62		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	$\surd$		0.86		0.14	
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.11	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 7: Frequentist results for the error model of PRAM 2010 with $c(x) = \rho x - x^2$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.62	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\surd$	`0.92`	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.17	$\star$	$\star$	$\star$	$\star$	41	0.22	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	`0.15`	`0.38`	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.46`	$\star$	$\star$	$\star$	$\star$
7	$\surd$	`0.49`	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	`0.72`	$\surd$	$\star$	0.56	$\star$	$\star$	46	`0.05`	`0.70`	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.26`	0.30	$\star$	$\star$	$\star$	47	`0.56`	`0.23`	$\star$	$\star$	$\star$	$\star$
12	`0.97`	$\surd$	$\star$	0.06	$\star$	$\star$	48	`0.29`	`0.61`	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\surd$	$\star$	0.53	$\star$	$\star$	49	`0.45`	`0.82`	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\surd$	`0.58`	0.06	$\star$	$\star$	$\star$
15	`0.61`	$\surd$	$\star$	$\star$	$\star$	$\star$	52	`0.81`	`0.71`	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	0.26	0.22	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	`0.55`	`0.89`	55	$\surd$	$\surd$	0.92	$\surd$	$\star$	0.46
18	$\surd$	`0.98`	0.72	$\star$	$\star$	$\star$	56	`0.26`	`0.84`	$\star$	$\star$	$\star$	$\star$
19	0.06	0.62	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	`0.37`	`0.89`	$\star$	0.14
20	`0.15`	$\surd$	$\star$	$\surd$	$\star$	1.00	59	`0.90`	`0.95`	$\star$	$\star$	$\star$	$\star$
21	`0.10`	`0.22`	$\star$	$\star$	$\star$	$\star$	61	$\surd$	`0.56`	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.55`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	`0.92`	`0.51`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	0.51	0.81	$\star$	$\star$	$\star$
24	0.88	$\star$	$\star$	$\star$	$\star$	$\star$	67	0.87	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	0.52		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	0.88		$\star$		$\star$	
27	`0.19`	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	`0.33`	`0.79`	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	`0.74`	`0.61`	$\star$	$\star$	$\star$	$\star$	40	0.52		$\star$		$\star$	
30	`0.37`	`0.95`	$\star$	$\star$	$\star$	$\star$	45	0.07		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.17		$\star$		$\star$	
32	`0.78`	`0.77`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	`0.05`	`0.68`	$\star$	$\star$	$\star$	$\star$	57	0.17		$\star$		$\star$	
34	`0.19`	`0.51`	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	`0.97`	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	`0.45`	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 8: Frequentist results for the error model of PRAM 2010 with $c(x) = 1 - e^{-\rho x}$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\underline{\surd}$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.62	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\underline{\surd}$	`0.92`	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.06	$\star$	$\star$	$\star$	$\star$	41	0.22	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	0.15	0.38	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.31`	$\star$	$\star$	$\star$	$\star$
7	$\underline{\surd}$	`0.49`	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.07	$\star$	$\star$	$\star$	$\star$
9	`0.72`	$\underline{\surd}$	$\star$	0.56	$\star$	$\star$	46	0.05	$\star$	$\star$	$\star$	$\star$	$\star$
11	$\underline{\surd}$	`0.12`	0.30	$\star$	$\star$	$\star$	47	`0.56`	`0.15`	$\star$	$\star$	$\star$	$\star$
12	`0.96`	$\underline{\surd}$	$\star$	0.06	$\star$	$\star$	48	`0.24`	`0.37`	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	$\underline{\surd}$	$\underline{\surd}$	`0.98`	$\underline{\surd}$	`0.39`	`0.68`	50	$\underline{\surd}$	`0.58`	0.06	$\star$	$\star$	$\star$
15	`0.56`	$\underline{\surd}$	$\star$	$\star$	$\star$	$\star$	52	`0.74`	`0.71`	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	$\star$	0.34	$\star$	$\star$	$\star$	$\star$
17	$\underline{\surd}$	$\underline{\surd}$	$\underline{\surd}$	$\underline{\surd}$	`0.55`	`0.89`	55	$\star$	$\surd$	$\star$	$\surd$	$\star$	0.46
18	`0.09`	`0.98`	$\star$	$\star$	$\star$	$\star$	56	`0.12`	`0.80`	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	`0.95`	`0.78`	0.10	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	`0.74`	`0.16`	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\underline{\surd}$	`0.56`	$\star$	$\star$	$\star$	$\star$
22	$\underline{\surd}$	`0.38`	$\star$	$\star$	$\star$	$\star$	65	$\underline{\surd}$	$\underline{\surd}$	0.25	0.53	$\star$	$\star$
23	`0.92`	`0.51`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	0.88	$\star$	$\star$	$\star$	$\star$	$\star$	67	0.87	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	0.24		$\star$		$\star$	
26	$\underline{\surd}$	$\underline{\surd}$	0.61	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	0.19	$\star$	$\star$	$\star$	$\star$	8	0.05		$\star$		$\star$	
28	$\star$	0.79	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	`0.74`	`0.61`	$\star$	$\star$	$\star$	$\star$	40	0.52		$\star$		$\star$	
30	`0.30`	`0.92`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.09		$\star$		$\star$	
32	`0.78`	`0.77`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	`0.05`	`0.68`	$\star$	$\star$	$\star$	$\star$	57	0.06		$\star$		$\star$	
34	`0.19`	`0.51`	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	0.15		$\star$		$\star$	
36	$\underline{\surd}$	$\underline{\surd}$	`0.97`	$\underline{\surd}$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	`0.19`	$\underline{\surd}$	$\star$	$\star$	$\star$	$\star$							

Table 9: Frequentist results for the error model of PRAM 2006 with $c(x) = x$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\underline{\surd}$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	0.34	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.17	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.29`	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.24`	0.30	$\star$	$\star$	$\star$	47	0.50	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	0.34	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	`0.45`	`0.81`	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\star$	0.07	$\star$	$\star$	$\star$	$\star$
15	0.09	$\star$	$\star$	$\star$	$\star$	$\star$	52	`0.61`	`0.44`	$\star$	$\star$	$\star$	$\star$
16	0.71	$\star$	$\star$	$\star$	$\star$	$\star$	53	0.26	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	0.23	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	`0.37`	`0.89`	$\star$	0.14
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	`0.71`	`0.95`	$\star$	$\star$	$\star$	$\star$
21	0.10	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.62`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	`0.56`	0.25	$\star$	$\star$	$\star$
23	`0.62`	`0.07`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	0.07	$\star$	$\star$	$\star$	$\star$	$\star$
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
27	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$	$\star$	$\surd$	$\star$	0.94	$\star$
28	`0.35`	`0.19`	$\star$	$\star$	$\star$	$\star$	10	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
29	0.08	$\star$	$\star$	$\star$	$\star$	$\star$	40	0.18	$\star$	$\star$	$\star$	$\star$	$\star$
30	`0.28`	`0.99`	$\star$	$\star$	$\star$	$\star$	45	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.10	$\star$	$\star$	$\star$	$\star$	$\star$
32	`0.78`	`0.62`	$\star$	$\star$	$\star$	$\star$	54	$\surd$	$\star$	$\star$	$\star$	$\star$	$\star$
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	0.14	$\star$	$\star$	$\star$	$\star$	$\star$
34	$\star$	0.16	$\star$	$\star$	$\star$	$\star$	60	0.57	$\star$	$\star$	$\star$	$\star$	$\star$
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$	$\star$	$\surd$	$\star$	0.27	$\star$
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	$\surd$	$\star$	0.86	$\star$	0.14	$\star$
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	$\surd$	$\star$	$\star$	$\star$	$\star$	$\star$
38	0.11	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 10: Frequentist results for the error model of PRAM 2006 with $c(x) = log(\rho + x)$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	`0.88`	$\surd$	$\star$	0.30	$\star$	$\star$	39	$\surd$	`0.92`	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.92	$\star$	$\star$	$\star$	$\star$	41	`0.36`	`0.81`	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	$\surd$	$\surd$	0.63	0.31	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.31`	$\star$	$\star$	$\star$	$\star$
7	$\surd$	`0.49`	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	`0.72`	$\surd$	$\star$	0.56	$\star$	$\star$	46	`0.05`	`0.70`	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.24`	0.30	$\star$	$\star$	$\star$	47	$\surd$	`0.57`	$\star$	$\star$	$\star$	$\star$
12	`0.96`	$\surd$	$\star$	0.06	$\star$	$\star$	48	`0.24`	`0.43`	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	`0.45`	`0.81`	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\surd$	`0.62`	0.06	$\star$	$\star$	$\star$
15	`0.56`	$\surd$	$\star$	$\star$	$\star$	$\star$	52	`0.82`	`0.71`	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	`0.26`	`0.22`	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	`0.55`	`0.89`	55	$\star$	$\surd$	$\star$	$\surd$	$\star$	0.46
18	$\surd$	$\surd$	`0.72`	`0.08`	$\star$	$\star$	56	`0.24`	`0.85`	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	`0.37`	`0.89`	$\star$	0.14
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	`0.74`	`0.95`	$\star$	$\star$	$\star$	$\star$
21	0.10	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\surd$	`0.56`	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.77`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	`0.92`	`0.51`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\surd$	$\surd$	$\star$	0.58	$\star$	$\star$	67	`0.99`	$\surd$	0.11	0.20	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	$\surd$		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	`0.19`	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	`0.49`	`0.81`	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	`0.74`	`0.61`	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	`0.32`	`0.99`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\surd$	$\surd$	`0.51`	`0.07`	$\star$	$\star$	51	0.10		$\star$		$\star$	
32	`0.78`	$\surd$	$\star$	0.97	$\star$	0.07	54	$\surd$		$\star$		$\star$	
33	$\surd$	$\surd$	`0.19`	`0.89`	$\star$	$\star$	57	0.14		$\star$		$\star$	
34	$\surd$	`0.51`	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	`0.97`	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	`0.23`	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 11: Frequentist results for the error model of PRAM 2006 with $c(x) = x^\rho$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.88	$\surd$	$\star$	0.30	$\star$	$\star$	39	$\surd$	0.92	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.92	$\star$	$\star$	$\star$	$\star$	41	0.36	0.81	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	$\surd$	$\surd$	0.63	0.31	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	0.82	0.46	$\star$	$\star$	$\star$	$\star$
7	$\surd$	0.49	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	0.35	0.97	$\star$	0.17	$\star$	$\star$	46	0.05	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	0.24	0.30	$\star$	$\star$	$\star$	47	$\surd$	0.57	$\star$	$\star$	$\star$	$\star$
12	0.96	$\surd$	$\star$	0.06	$\star$	$\star$	48	0.24	0.43	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\surd$	$\star$	0.53	$\star$	$\star$	49	0.45	0.82	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	0.98	$\surd$	0.39	0.68	50	$\surd$	0.58	0.06	$\star$	$\star$	$\star$
15	0.56	0.47	$\star$	$\star$	$\star$	$\star$	52	0.82	0.71	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	0.26	0.22	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	0.55	0.89	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\surd$	$\surd$	0.72	0.08	$\star$	$\star$	56	0.26	0.85	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	0.37	0.89	$\star$	0.14
20	0.15	$\surd$	$\star$	$\surd$	$\star$	1.00	59	0.90	0.95	$\star$	$\star$	$\star$	$\star$
21	0.10	0.22	$\star$	$\star$	$\star$	$\star$	61	0.42	0.56	$\star$	$\star$	$\star$	$\star$
22	$\surd$	0.77	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	0.92	0.51	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\surd$	$\surd$	$\star$	0.58	$\star$	$\star$	67	0.99	$\surd$	0.11	0.20	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	$\surd$		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	0.88		$\star$		$\star$	
27	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	0.49	0.81	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	0.74	0.61	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	0.32	0.99	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\surd$	$\surd$	0.51	0.07	$\star$	$\star$	51	0.14		$\star$		$\star$	
32	0.78	$\surd$	$\star$	0.97	$\star$	0.07	54	$\surd$		$\star$		$\star$	
33	$\surd$	$\surd$	0.19	0.89	$\star$	$\star$	57	0.17		$\star$		$\star$	
34	$\surd$	0.51	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.23	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 12: Frequentist results for the error model of PRAM 2006 with $c(x) = x^\rho$ where $\rho \in [.01, 1]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.88	$\surd$	$\star$	0.30	$\star$	$\star$	39	$\surd$	0.92	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.92	$\star$	$\star$	$\star$	$\star$	41	0.36	0.81	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	$\surd$	$\surd$	0.63	0.31	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	0.82	0.31	$\star$	$\star$	$\star$	$\star$
7	$\surd$	0.49	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	0.35	0.97	$\star$	0.17	$\star$	$\star$	46	0.05	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	0.24	0.30	$\star$	$\star$	$\star$	47	$\surd$	0.57	$\star$	$\star$	$\star$	$\star$
12	0.96	$\surd$	$\star$	0.06	$\star$	$\star$	48	0.24	0.43	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	0.45	0.81	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	0.98	$\surd$	0.39	0.68	50	$\surd$	0.58	0.06	$\star$	$\star$	$\star$
15	0.56	0.47	$\star$	$\star$	$\star$	$\star$	52	0.82	0.71	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	0.26	0.22	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	0.55	0.89	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\surd$	$\surd$	0.72	0.08	$\star$	$\star$	56	0.24	0.85	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	0.37	0.89	$\star$	0.14
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	0.74	0.95	$\star$	$\star$	$\star$	$\star$
21	0.10	$\star$	$\star$	$\star$	$\star$	$\star$	61	0.42	0.56	$\star$	$\star$	$\star$	$\star$
22	$\surd$	0.77	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	0.92	0.51	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\surd$	$\surd$	$\star$	0.58	$\star$	$\star$	67	0.99	$\surd$	0.11	0.20	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	$\surd$		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	0.49	0.81	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	0.74	0.61	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	0.32	0.99	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\surd$	$\surd$	0.51	0.07	$\star$	$\star$	51	0.10		$\star$		$\star$	
32	0.78	$\surd$	$\star$	0.97	$\star$	0.07	54	$\surd$		$\star$		$\star$	
33	$\surd$	$\surd$	0.19	0.89	$\star$	$\star$	57	0.14		$\star$		$\star$	
34	$\surd$	0.51	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.23	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 13: Frequentist results for the error model of PRAM 2006 with $c(x) = x^\rho$ where $\rho \in [1.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	0.36	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.17	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	`0.82`	`0.46`	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	`0.24`	0.30	$\star$	$\star$	$\star$	47	0.50	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	0.38	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\surd$	$\star$	0.53	$\star$	$\star$	49	`0.45`	`0.82`	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	`0.98`	$\surd$	`0.39`	`0.68`	50	$\star$	0.07	$\star$	$\star$	$\star$	$\star$
15	0.09	$\star$	$\star$	$\star$	$\star$	$\star$	52	`0.61`	`0.44`	$\star$	$\star$	$\star$	$\star$
16	0.71	$\star$	$\star$	$\star$	$\star$	$\star$	53	0.26	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	0.26	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	`0.37`	`0.89`	$\star$	0.14
20	`0.15`	$\surd$	$\star$	$\surd$	$\star$	1.00	59	`0.90`	`0.95`	$\star$	$\star$	$\star$	$\star$
21	`0.10`	`0.22`	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	$\surd$	`0.62`	$\star$	$\star$	$\star$	$\star$	65	$\surd$	`0.56`	0.25	$\star$	$\star$	$\star$
23	`0.62`	`0.07`	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	0.07		$\star$		$\star$	
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	0.88		$\star$		$\star$	
27	`0.19`	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	`0.35`	`0.19`	$\star$	$\star$	$\star$	$\star$	10	$\star$		$\star$		$\star$	
29	0.10	$\star$	$\star$	$\star$	$\star$	$\star$	40	0.18		$\star$		$\star$	
30	`0.28`	`0.99`	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	0.14		$\star$		$\star$	
32	`0.78`	`0.62`	$\star$	$\star$	$\star$	$\star$	54	$\surd$		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	0.17		$\star$		$\star$	
34	$\star$	0.20	$\star$	$\star$	$\star$	$\star$	60	0.62		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	$\surd$		0.86		0.14	
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.11	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 14: Frequentist results for the error model of PRAM 2006 with $c(x) = \rho x - x^2$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.88	$\surd$	$\star$	0.30	$\star$	$\star$	39	$\surd$	0.92	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.92	$\star$	$\star$	$\star$	$\star$	41	0.36	0.81	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	0.99	$\surd$	0.24	0.31	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	0.82	0.46	$\star$	$\star$	$\star$	$\star$
7	$\surd$	0.49	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	0.72	$\surd$	$\star$	0.56	$\star$	$\star$	46	0.05	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	0.26	0.30	$\star$	$\star$	$\star$	47	$\surd$	0.57	$\star$	$\star$	$\star$	$\star$
12	0.97	$\surd$	$\star$	0.06	$\star$	$\star$	48	0.29	0.65	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\surd$	$\star$	0.53	$\star$	$\star$	49	0.45	0.82	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	0.98	$\surd$	0.39	0.68	50	$\surd$	0.62	0.06	$\star$	$\star$	$\star$
15	0.61	$\surd$	$\star$	$\star$	$\star$	$\star$	52	0.81	0.93	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	0.26	0.22	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	0.55	0.89	55	$\surd$	$\surd$	0.92	$\surd$	$\star$	0.46
18	$\surd$	$\surd$	0.72	0.08	$\star$	$\star$	56	0.26	0.85	$\star$	$\star$	$\star$	$\star$
19	0.06	0.67	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	0.37	0.89	$\star$	0.14
20	0.15	$\surd$	$\star$	$\surd$	$\star$	1.00	59	0.90	0.95	$\star$	$\star$	$\star$	$\star$
21	0.10	0.22	$\star$	$\star$	$\star$	$\star$	61	$\surd$	0.59	$\star$	$\star$	$\star$	$\star$
22	$\surd$	0.77	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	0.92	0.55	$\star$	$\star$	$\star$	$\star$	66	$\surd$	0.51	0.81	$\star$	$\star$	$\star$
24	$\surd$	0.92	$\star$	0.05	$\star$	$\star$	67	0.99	$\surd$	0.11	0.20	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	$\surd$		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	0.88		$\star$		$\star$	
27	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	0.49	0.81	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	$\surd$	0.61	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	0.37	0.99	$\star$	$\star$	$\star$	$\star$	45	0.07		$\star$		$\star$	
31	$\surd$	$\surd$	0.51	0.07	$\star$	$\star$	51	0.17		$\star$		$\star$	
32	0.78	$\surd$	$\star$	0.97	$\star$	0.07	54	$\surd$		$\star$		$\star$	
33	$\surd$	$\surd$	0.19	0.89	$\star$	$\star$	57	0.17		$\star$		$\star$	
34	$\surd$	0.34	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.45	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 15: Frequentist results for the error model of PRAM 2006 with $c(x) = 1 - e^{-\rho x}$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# 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. Successful replications across sessions are marked in `typewriter` and/or $\surd$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	0.88	$\surd$	$\star$	0.30	$\star$	$\star$	39	$\surd$	0.92	$\star$	$\star$	$\star$	$\star$
2	$\star$	0.92	$\star$	$\star$	$\star$	$\star$	41	0.36	0.81	$\star$	$\star$	$\star$	$\star$
4	$\star$	0.05	$\star$	$\star$	$\star$	$\star$	42	0.99	$\surd$	0.24	0.31	$\star$	$\star$
5	0.36	$\star$	$\star$	$\star$	$\star$	$\star$	43	0.82	0.31	$\star$	$\star$	$\star$	$\star$
7	$\surd$	0.49	$\star$	$\star$	$\star$	$\star$	44	$\star$	0.09	$\star$	$\star$	$\star$	$\star$
9	0.72	$\surd$	$\star$	0.56	$\star$	$\star$	46	0.05	0.70	$\star$	$\star$	$\star$	$\star$
11	$\surd$	0.24	0.30	$\star$	$\star$	$\star$	47	$\surd$	0.57	$\star$	$\star$	$\star$	$\star$
12	0.96	$\surd$	$\star$	0.06	$\star$	$\star$	48	0.24	0.43	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	0.45	0.81	$\star$	$\star$	$\star$	$\star$
14	$\surd$	$\surd$	0.98	$\surd$	0.39	0.68	50	$\surd$	0.62	0.06	$\star$	$\star$	$\star$
15	0.56	$\surd$	$\star$	$\star$	$\star$	$\star$	52	0.82	0.71	$\star$	$\star$	$\star$	$\star$
16	$\surd$	$\surd$	$\star$	0.99	$\star$	0.51	53	0.26	0.22	$\star$	$\star$	$\star$	$\star$
17	$\surd$	$\surd$	$\surd$	$\surd$	0.55	0.89	55	$\star$	$\surd$	$\star$	$\surd$	$\star$	0.46
18	$\surd$	$\surd$	0.72	0.08	$\star$	$\star$	56	0.24	0.85	$\star$	$\star$	$\star$	$\star$
19	$\star$	0.62	$\star$	$\star$	$\star$	$\star$	58	$\surd$	$\surd$	0.37	0.89	$\star$	0.14
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	0.74	0.95	$\star$	$\star$	$\star$	$\star$
21	0.10	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\surd$	0.56	$\star$	$\star$	$\star$	$\star$
22	$\surd$	0.77	$\star$	$\star$	$\star$	$\star$	65	$\surd$	$\surd$	0.25	0.53	$\star$	$\star$
23	0.92	0.51	$\star$	$\star$	$\star$	$\star$	66	$\surd$	$\star$	0.81	$\star$	$\star$	$\star$
24	$\surd$	0.92	$\star$	0.05	$\star$	$\star$	67	0.99	$\surd$	0.11	0.20	$\star$	$\star$
25	$\surd$	$\surd$	$\star$	0.35	$\star$	$\star$	3	$\surd$		$\star$		$\star$	
26	$\surd$	$\surd$	0.61	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	0.19	$\surd$	$\star$	$\star$	$\star$	$\star$	8	$\surd$		$\surd$		0.94	
28	0.49	0.81	$\star$	$\star$	$\star$	$\star$	10	$\surd$		$\star$		$\star$	
29	0.74	0.61	$\star$	$\star$	$\star$	$\star$	40	0.55		$\star$		$\star$	
30	0.32	0.99	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\surd$	$\surd$	0.51	0.07	$\star$	$\star$	51	0.10		$\star$		$\star$	
32	0.78	$\surd$	$\star$	0.97	$\star$	0.07	54	$\surd$		$\star$		$\star$	
33	$\surd$	$\surd$	0.19	0.89	$\star$	$\star$	57	0.14		$\star$		$\star$	
34	$\surd$	0.51	$\star$	$\star$	$\star$	$\star$	60	0.66		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\surd$		$\surd$		0.27	
36	$\surd$	$\surd$	0.97	$\surd$	$\star$	0.95	63	$\surd$		0.86		0.14	
37	$\star$	$\surd$	$\star$	0.59	$\star$	$\star$	64	$\surd$		$\star$		$\star$	
38	0.23	$\surd$	$\star$	$\star$	$\star$	$\star$							

Table 16: Bayesian results for the error model of PRAM 2010 with $c(x) = x$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
11	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	47	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	PRAM	PRAM	PRAM	PRAM	PRAM	PRAM	50	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
15	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	52	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	53	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$
23	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	$\star$		$\star$		$\star$	
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	8	$\star$		$\star$		$\star$	
28	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	10	$\star$		$\star$		$\star$	
29	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	40	PRAM		$\star$		$\star$	
30	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	PRAM		$\star$		$\star$	
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		$\star$		$\star$	
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	PRAM		PRAM		$\star$	
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 17: Bayesian results for the error model of PRAM 2010 with $c(x) = log(\rho + x)$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
11	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	47	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	PRAM	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	PRAM	$\star$	PRAM
18	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	8	$\star$		$\star$		$\star$	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		$\star$		$\star$	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 18: Bayesian results for the error model of PRAM 2010 with $c(x) = x^\rho$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	$\star$		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	PRAM		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	PRAM		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	PRAM		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 19: Bayesian results for the error model of PRAM 2010 with $c(x) = x^\rho$ where $\rho \in [0.01, 1]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	46	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
11	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	47	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	PRAM	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	$\star$		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	8	$\star$		$\star$		$\star$	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		$\star$		$\star$	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 20: Bayesian results for the error model of PRAM 2010 with $c(x) = x^\rho$ where $\rho \in [1.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	PRAM	PRAM	PRAM	PRAM	PRAM	50	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
15	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	52	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$
23	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	PRAM		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	10	$\star$		$\star$		$\star$	
29	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	40	PRAM		$\star$		$\star$	
30	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	PRAM		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	PRAM		PRAM		PRAM	
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	PRAM		PRAM		$\star$	
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 21: Bayesian results for the error model of PRAM 2010 with $c(x) = \rho x - x^2$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM
18	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	PRAM		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	PRAM		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	PRAM		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 22: Bayesian results for the error model of PRAM 2010 with $c(x) = 1 - e^{-\rho x}$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
11	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	47	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
14	PRAM	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	PRAM	$\star$	PRAM
18	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	8	$\star$		$\star$		$\star$	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		$\star$		$\star$	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 23: Bayesian results for the error model of PRAM 2006 with $c(x) = x$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id).Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	PRAM	PRAM	PRAM	PRAM	PRAM	50	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
15	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	52	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$
23	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM	$\star$	PRAM	$\star$	PRAM	$\star$
28	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	10	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
29	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	40	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
30	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
34	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$	$\star$	PRAM	$\star$	PRAM	$\star$
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
38	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 24: Bayesian results for the error model of PRAM 2006 with $c(x) = log(\rho + x)$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
12	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	48	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	PRAM	$\star$	PRAM
18	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	67	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	54	PRAM		$\star$		$\star$	
33	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 25: Bayesian results for the error model of PRAM 2006 with $c(x) = x^\rho$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	67	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	PRAM		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	51	PRAM		$\star$		$\star$	
32	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	54	PRAM		$\star$		$\star$	
33	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 26: Bayesian results for the error model of PRAM 2006 with $c(x) = x^\rho$ where $\rho \in [0.01, 1]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	67	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	54	PRAM		$\star$		$\star$	
33	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 27: Bayesian results for the error model of PRAM 2006 with $c(x) = x^\rho$ where $\rho \in [1.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	39	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
12	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	48	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	PRAM	PRAM	PRAM	PRAM	PRAM	50	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
15	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	52	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	55	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
18	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	56	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
19	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$
23	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	67	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
25	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	3	$\star$		$\star$		$\star$	
26	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	6	PRAM		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	10	$\star$		$\star$		$\star$	
29	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	51	PRAM		$\star$		$\star$	
32	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	54	PRAM		$\star$		$\star$	
33	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		PRAM		PRAM	
36	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	63	PRAM		PRAM		$\star$	
37	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$							

Table 28: Bayesian results for the error model of PRAM 2006 with $c(x) = \rho x - x^2$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
12	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	48	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM
18	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	PRAM		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	54	PRAM		$\star$		$\star$	
33	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$							

Table 29: Bayesian results for the error model of PRAM 2006 with $c(x) = 1 - e^{-\rho x}$ where $\rho \in [.01, 100]$ and with $\tau = \frac{1}{2}$ (within-person modal choice), $\tau = \frac{1}{4}$ and $\tau = \frac{1}{10}$. Of the 67 participants in the first session, 54 returned (# is the participant id). Here, each case listed as "PRAM" is a case where the Bayesian p value is ≥ 0.05 and the model wins against the unconstrained model by DIC. All other cases are "rejections," marked with $\star$.

	$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$			$\tau = \frac{1}{2}$		$\tau = \frac{1}{4}$		$\tau = \frac{1}{10}$	
#	S1	S2	S1	S2	S1	S2	#	S1	S2	S1	S2	S1	S2
1	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	39	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
2	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	41	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
4	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	42	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
5	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	43	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
7	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	44	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
9	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	46	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$
11	PRAM	$\star$	PRAM	$\star$	$\star$	$\star$	47	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
12	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	48	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
13	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	49	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
14	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	50	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
15	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	52	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
16	PRAM	$\star$	$\star$	PRAM	$\star$	PRAM	53	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
17	$\star$	$\star$	PRAM	PRAM	PRAM	PRAM	55	$\star$	$\star$	$\star$	PRAM	$\star$	PRAM
18	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	56	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$
19	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	58	PRAM	$\star$	PRAM	PRAM	$\star$	$\star$
20	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	59	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
21	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	61	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
22	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	65	PRAM	PRAM	PRAM	PRAM	$\star$	$\star$
23	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	66	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$
24	PRAM	$\star$	$\star$	$\star$	$\star$	$\star$	67	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$
25	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	3	PRAM		$\star$		$\star$	
26	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	6	$\star$		$\star$		$\star$	
27	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	8	PRAM		PRAM		PRAM	
28	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	10	PRAM		$\star$		$\star$	
29	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	40	$\star$		$\star$		$\star$	
30	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$	45	$\star$		$\star$		$\star$	
31	PRAM	PRAM	PRAM	$\star$	$\star$	$\star$	51	$\star$		$\star$		$\star$	
32	$\star$	PRAM	$\star$	PRAM	$\star$	PRAM	54	PRAM		$\star$		$\star$	
33	PRAM	PRAM	$\star$	PRAM	$\star$	$\star$	57	$\star$		$\star$		$\star$	
34	PRAM	PRAM	$\star$	$\star$	$\star$	$\star$	60	PRAM		$\star$		$\star$	
35	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	62	$\star$		PRAM		PRAM	
36	PRAM	$\star$	PRAM	PRAM	$\star$	PRAM	63	PRAM		PRAM		$\star$	
37	$\star$	PRAM	$\star$	PRAM	$\star$	$\star$	64	PRAM		$\star$		$\star$	
38	$\star$	PRAM	$\star$	$\star$	$\star$	$\star$							

Appendix D: Heterogeneity and Parsimony in Intertemporal Choice

Decision
2018, Vol. 5, No. 2, 63–94

2325-9965/18/$12.00 http://dx.doi.org/10.1037/dec0000069

Heterogeneity and Parsimony in Intertemporal Choice

Michel Regenwetter
University of Illinois at Urbana–Champaign

Daniel R. Cavagnaro
California State University, Fullerton

Anna Popova
Dell Research Labs, Round Rock, Texas

Ying Guo and Chris Zwilling
University of Illinois at Urbana–Champaign

Shiau Hong Lim
IBM Research, Singapore, Singapore

Jeffrey R. Stevens
Max Planck Institute for Human Development, Berlin, Germany, and University of Nebraska–Lincoln

Behavioral theories of intertemporal choice involve many moving parts. Most descriptive theories model how time delays and rewards are perceived, compared, and/or combined into preferences or utilities. Most behavioral studies neglect to spell out how such constructs translate into heterogeneous observable choices. We consider several broad models of transitive intertemporal preference and combine these with several mathematically formal, yet very general, models of heterogeneity. We evaluate 20 probabilistic models of intertemporal choice using binary choice data from two large-scale experiments. Our analysis documents the interplay between heterogeneity and parsimony in accounting for empirical data: We find evidence for heterogeneity across individuals and across stimulus sets that can be accommodated with transitive models of varying complexity. We do not find systematic violations of transitivity in our data. Future work should continue to tackle the complex trade-off between parsimony and heterogeneity.

Keywords: heterogeneity, intertemporal choice, noise, random preference, transitivity of preferences

Supplemental materials: http://dx.doi.org/10.1037/dec0000069.supp

This article was published Online First January 12, 2017.

Michel Regenwetter, Department of Psychology, University of Illinois at Urbana–Champaign; Daniel R. Cavagnaro, Department of Information Systems and Decision Sciences, California State University, Fullerton; Anna Popova, Dell Research Labs, Round Rock, Texas; Ying Guo and Chris Zwilling, Department of Psychology, University of Illinois at Urbana–Champaign; Shiau Hong Lim, IBM Research, Singapore, Singapore; Jeffrey R. Stevens, Max Planck Institute for Human Development, Berlin, Germany, and Department of Psychology and Center for Brain, Biology & Behavior, University of Nebraska–Lincoln.

This work was supported by the National Science Foundation (SES-10-62045 and SES-14-59699, PI: Michel Regenwetter), the Alexander von Humboldt Foundation (TransCoop grant, PI: Jeffrey R. Stevens), and XSEDE (SES-130016, PI: Michel Regenwetter). We are grateful to Muye Chen for comments on a draft and to Gregor Caregnato for testing respondents at the Max Planck Institute for Human Development. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of their colleagues, their funding agencies, or their universities and research labs.

Correspondence concerning this article should be addressed to Michel Regenwetter, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: regenwet@illinois.edu

A dieter must choose between the immediate gratification of a waistline-expanding piece of cake or the longer term health benefits of fruit. A business manager must choose between developing projects with "low-hanging fruit" or investing time, personnel, and money into achieving long-term goals of the firm. From diet choices to large-scale organizational decisions, all such *intertemporal choices* involve options available at different points in time (Read, 2004). In this article, we consider binary choice between one delayed reward and another that is larger in size but also requires a longer wait. Such pairwise choices are highly heterogeneous in that they vary across decision makers and within a given decision maker over repeated decisions within even short time periods.

Economists and psychologists have developed dozens of models for intertemporal choices aimed at understanding how decision makers trade off between smaller, sooner and larger, later rewards. Most of these are *temporal discounting* models that generate a subjective present value for an option discounted by the time delay to receiving the reward. For instance, $100 in a year is less valuable than $100 in a week, which, in turn, is still less valuable than $100 today. Discounting models that map rewards and time delays to numerical subjective values of time-delayed rewards, such as exponential and hyperbolic discounting, imply transitive preferences according to which a person preferring x to y and y to z must prefer x to z (see, e.g., Doyle, 2013; Doyle & Chen, 2012; Ebert & Prelec, 2007; Frederick, Loewenstein, & O'Donoghue, 2002; Green & Myerson, 2004; Killeen, 2009; Laibson, 1997; Loewenstein & Prelec, 1992; Mazur, 1987; McClure, Ericson, Laibson, Loewenstein, & Cohen, 2007; Samuelson, 1937).[1]

The study of the fundamental nature of intertemporal preferences faces a profound challenge. Existing tests of intertemporal choice theories rarely account explicitly for heterogeneity in behavior within and between people. It may not be possible to select a 'good' theory of intertemporal choice unless this theory jointly accounts for core preferences and heterogeneity in behavior. In our view, if we are to understand intertemporal choices, we should develop a rigorous approach that incorporates individual differences, variability in choices, and generalizability across stimuli. Therefore, rather than attend to the specifics of core preferences, such as the functional form of discounting curves, and rather than seek out a 'best' theory, we focus in this article on the complicated interplay between parsimony and empirical variability. We also concentrate on transitive intertemporal preference and how it manifests itself in probabilistic choice. Combining transitivity of preferences with the trade-off between parsimony and variability fills a gap in the existing literature in intertemporal choice by zooming out to a broad class of theories while zooming in to the sources and types of heterogeneity.

Accounting for heterogeneity comes at the cost of reducing model parsimony. Intuitively, an excessively parsimonious model may only account for one choice made by one person at one time point for one particular stimulus. Such an overly specific model is unlikely to generalize to other stimuli presented to the same person, to other occasions on which the same person is presented with the same stimulus, to other individuals, and/or to other stimuli. At the other end of the spectrum, a model that universally accounts for the behavior of all of humanity, at all times, and over all conceivable intertemporal stimuli may have to be overly flexible. Clearly, we need to aim for some sort of middle ground. It is therefore not surprising that much of the literature in decision research, and intertemporal choice in particular, aims merely at modeling the prototypical decision maker or at documenting trends and significant effects. Though this may be useful, it could also be inherently misleading in that almost no actual person might act like that 'prototypical' decision maker. We unpack the intimate connection between models of heterogeneity in preferences and in responses for transitive theories of intertemporal preference. We also explore how

[1] Other models, such as the "similarity" and "tradeoff" models, permit intransitive preferences (see, e.g., Leland, 2002; Manzini & Mariotti, 2006; Read, 2001; Rubinstein, 2003; Scholten & Read, 2006, 2010; Stevens, 2016). Here, a person may prefer x to y and y to z, yet prefer z to x for some x, y, z. A separate article tests nontransitive heuristic models on different stimuli and different respondents.

adequate theoretical accounts may vary with the stimuli used. We believe that careful attention to the nature and sources of heterogeneity is essential to advancing our understanding of intertemporal choice.

Without a good theory of heterogeneity, scholars risk making too many modifications in the functional forms of core theories in an effort to accommodate "discrepancies" between theory and data, when, instead, they should model the sources of heterogeneity of behavior more explicitly. This article provides a roadmap for accomplishing the latter by formally spelling out two major sources of heterogeneity: probabilistic responses and probabilistic preferences. We then show that these sources of heterogeneity can be incorporated into theories of intertemporal choice at an abstract level. We take a big-picture perspective and tackle intertemporal choice at a somewhat abstract level. We consider general classes of core models that share one or more of the features that (a) preferences are transitive linear orders, (b) choice options are represented by numerical utilities, and (c) strengths of preferences are consistent with transitive preferences. Likewise, we consider general classes of probabilistic mechanisms for pairwise choice, namely (a) aggregation-based models that encompass various response error models as special cases and (b) distribution-free random preference, random function and random utility models that model the preferences themselves as uncertain. This approach to heterogeneity is conceptually and mathematically different from the common approach that aims to accommodate individual differences through refining the core functional form of a theory, for example, by adding extra parameters that permit specific kinds of flexibility in the core theory. Instead, our approach resembles the literature on axiom testing in decision making in that we consider the general axiom of transitivity together with general classes of probabilistic specifications.

A major strength of our approach is that it allows triage of entire classes of theories. Nonetheless, even within this general and abstract paradigm of transitivity of intertemporal preference, the number of models to consider is substantial, and different models differ dramatically in their parsimony. Furthermore, investigating the tradeoff between parsimony and heterogeneity is computationally costly. Because we consider 20 probabilistic models separately for 61 individual decision makers on six different stimulus sets, because we employ both frequentist and Bayesian analysis methods, and because many of our analyses utilize either grid search or Monte Carlo sampling methods, our analyses necessitated the use of supercomputing resources.[2]

We first discuss how to spell out a model of binary choice behavior for a person with transitive preferences. We emphasize that, in contrast to the risky choice literature, the intertemporal choice literature has largely neglected modeling the sources and types of uncertainty that underlie probabilistic behavioral data. We fill this gap by introducing eight types of probabilistic choice models of transitive intertemporal preference. After we review suitable statistical analysis methods and two experiments, we give an in-depth report on quantitative analyses at the individual and group level. We particularly highlight how parsimony trades off with accounting for within- and between-person heterogeneity. In contrast to previous such projects, we concentrate on intertemporal choice.

Transitive Intertemporal Preference and Choice

In behavioral science, it is crucial not to mistake models of hypothetical constructs for models of observable behavior. The literature on intertemporal choice engages in a thorough discussion about hypothetical constructs such as preference or utility, while usually omitting a detailed model of observable behavior such as choice. We review probabilistic choice models aimed at formally representing the uncertainty that is inherent in overt behavior. We then walk through the step-by-step approach to design and test an explicitly specified theory of pairwise intertemporal choice. Because any real collec-

[2] We ran the most computationally expensive analyses on Pittsburgh Supercomputer Center's *Blacklight* and *Greenfield* supercomputers, as an *Extreme Science and Engineering Discovery Environment* project (see also Towns et al., 2014). The analyses in this article expended about 24,000 CPU hr on the supercomputer and more than 1,000 hr on the PC.

tion of experiments can only utilize finitely many stimuli, we assume throughout, and without much loss of generality, that the set of all choice alternatives under consideration is finite. We also concentrate on the common experimental paradigm of pairwise choice between a larger reward available with a longer delay and a smaller reward available with less delay.

Preference

Many models of binary preference between a larger, later reward L and a smaller, sooner reward S characterize a three-component cognitive process: They specify implicitly or explicitly how a decision maker (a) subjectively perceives time, (b) subjectively perceives rewards, and (c) subjectively perceives the interaction between time and rewards. This permits them to define such hypothetical constructs as the pairwise preference among choice options, the subjective value of an option, or the subjective strength of preference among pairs of options. In addition, to actually predict or explain behavior, a model must specify how hypothetical constructs such as subjective values or preferences translate into something one can observe, such as overt choice behavior. Before discussing choice, we start by reviewing models of transitive intertemporal preference.

A broad class of theories for intertemporal preference uses numerical functions and operations on numbers to model either subjective values of options or subjective strengths of preference among options. Suppose that x is the option of receiving a monetary or nonmonetary reward A after a time delay $t \geq 0$ (with $t = 0$ denoting an immediate reward). Many numerical models, especially many discounting models, assume that reward A is mapped into a numerical value via some value function v, that time delay t is mapped into a numerical value via some time weighting function Ψ, and that these numerical values are combined into an overall numerical value for x via some mathematical operation $\odot$, to yield an overall subjective numerical value $u(x)$ for option x as

$$u(x) = v(A) \odot \Psi(t). \tag{1}$$

Using this representation, many models of intertemporal preference model the preference $\succ$ as

$$L \succ S \Leftrightarrow u(L) > u(S), \tag{2}$$

where $L \succ S$ denotes that L is strictly preferred to S (see also Doyle, 2013, for similar formulations). Such a binary preference relation $\succ$ is *transitive* in that, for any options x, y, z, whenever $x \succ y$ and $y \succ z$, it follows from the right hand side of Condition 2 that $x \succ z$ as well. The general approach (1) – (2) encompasses the vast majority of theories for intertemporal choice, including the bulk of discounting models. Different implementations of such theories vary in their assumptions about the specific functional forms of v and Ψ and the operation $\odot$: Different theories use different functions $v(A)$, oftentimes focusing on quantitative rewards $A \in \mathbb{R}^+$, such as money,

$$v(A) = \begin{cases} \alpha A & \text{(often with } \alpha = 1\text{, Samuelson, 1937; Mazur, 1984),} \\ A^{\alpha} & \text{(Killeen, 2009),} \\ \dots, & \end{cases} \tag{3}$$

different functions $\Psi(t)$,

$$\Psi(t) = \begin{cases} \delta^t & \text{(Samuelson, 1937),} \\ \delta t^{\beta} & \text{(Killeen, 2009),} \\ \dfrac{1}{1+\delta t} & \text{(Mazur, 1984),} \\ \dfrac{1}{1+\delta t^{\beta}} & \text{(Mazur, 1987),} \\ \dfrac{1}{(1+\delta t)^{\beta/\delta}} & \text{(Loewenstein \& Prelec, 1992; Green \& Myerson, 2004),} \\ e^{-(\delta t)^{\beta}} & \text{(Ebert \& Prelec, 2007),} \\ \omega e^{-\delta t} + (1-\omega) e^{-\beta t} & \text{(McClure et al., 2007),} \\ \dots, & \end{cases} \tag{4}$$

and different operations $\odot$,

$$v(A) \odot \Psi(t) = \begin{cases} v(A) \times \Psi(t) & \text{(Samuelson, 1937; Laibson, 1997; Mazur, 1984),} \\ v(A) - \Psi(t) & \text{(Killeen, 2009, Doyle \& Chen, 2012),} \\ \dots. & \end{cases} \tag{5}$$

(The cited articles also provide permissible ranges for the parameters α, β, δ, and ω in these functions.)

Even the two examples of v in Equation 3, seven examples of Ψ in Equation 4, and two operators $\odot$ in Equation 5 permit $2 \times 7 \times 2 = 28$ different combinations. The intertemporal choice literature has generated a panoply of such models for preferences, subjective values, or strengths of preferences. Most studies stop with the derivation of these constructs and do not specify response mechanisms that convert hypothetical constructs into predictions about heterogeneous overt choice behavior. Some scholars have recently started to incorporate stochastic specifications of response processes into theories of intertemporal choice (Arfer & Luhmann, 2015; Dai & Busemeyer, 2014; Ericson, White, Laibson, & Cohen, 2015).

The fact that most theories of intertemporal choice are silent about the response mechanism is problematic. Scholars in other domains, most notably in risky choice, have warned not to think of response mechanisms as a mere optional add-on that one selects based on convenience or subjective taste of what constitutes an elegant model (Carbone & Hey, 2000; Hey, 2005; Hey & Orme, 1994; Loomes, Moffatt, & Sugden, 2002; Loomes & Sugden, 1995; Luce, 1959, 1995; Luce & Narens, 1994; Luce & Suppes, 1965; McCausland & Marley, 2014). Misspecification of response processes substantially affects conclusions about parameter values and readily distorts the functional form of the underlying core algebraic model (Blavatskyy & Pogrebna, 2010; Stott, 2006; Wilcox, 2008). Mis- and overspecification also compromise one's ability to predict future choices based on best-fitting parameter values in a current study. An additional formidable challenge, compounded with the suitable selection of response models, often lies in finding suitable statistical methods (Davis-Stober, 2009; Iverson & Falmagne, 1985; Myung, Karabatsos, & Iverson, 2005). Our models and methods tackle these challenges at a high level of generality. Rather than look for a 'best' model, we focus on the interplay between heterogeneity and parsimony.

Preference and Choice

We now review major model classes of probabilistic choice. We assume throughout the rest of the article that there are only finitely many choice options under consideration; hence, we always only consider finitely many binary choice probabilities.

Tremble models build on the hypothetical construct of binary preference. They start from the premise that the decision maker has a fixed "true" preference $\succ$, and that choice probabilities reflect a tendency to make occasional errors in revealing the underlying hypothetical construct. In a tremble model, it is usually assumed that the error rate for a given pair of options (x, y) is a free parameter ϵ_{xy} (Birnbaum, 2008; Birnbaum & Navarrete, 1998; Harless & Camerer, 1994), so that the probability P_{xy} of choosing x over y is

$$P_{xy} = \begin{cases} 1 - \epsilon_{xy} & \text{if } x \succ y, \\ \epsilon_{xy} & \text{if } y \succ x, \end{cases} \quad \text{with, usually,} \quad 0 < \epsilon_{xy} \leq \frac{1}{2}.$$

Similarly, *Fechnerian models* are based on the notion that a decision maker has a fixed "true" utility function, but because of random noise, the decision maker reveals the underlying hypothetical construct only probabilistically. In contrast to tremble models, Fechnerian models explicitly model error rates as a monotonically decreasing function of the strength of preference, $\mathbb{S}_{xy}$, with choices for strongly preferred options (large values of $|\mathbb{S}_{xy}|$) being close to deterministic and choices for extremely weakly preferred options (small values of $|\mathbb{S}_{xy}|$) resembling the toss of a fair coin (Hey & Orme, 1994; Manski & McFadden, 1981; McFadden, 2001; Thurstone, 1927). According to a Fechnerian model, the binary choice probability is given by

$$P_{xy} = F(\mathbb{S}_{xy}),$$

with F a cumulative distribution function

$$\text{and } F(0) = \frac{1}{2}.$$

A logistic cumulative distribution function (CDF) yields the well-known *logit* model and a normal CDF yields the *probit* model, respectively.[3]

The strength of preference $\mathbb{S}_{xy}$, in turn, is another hypothetical construct, often derived from u using another operation, $\ominus$, via $\mathbb{S}_{xy} = u(x) \ominus u(y)$.

[3] One can also derive binary logit and probit models within a random utility framework, discussed below, by assuming that random utilities have extreme value or normal distributions, respectively.

Examples include $\mathbb{S}_{xy} = u(x) - u(y)$ or, for $u > 0$, $\mathbb{S}_{xy} = ln\left(\frac{u(x)}{u(y)}\right)$. The latter is used in a historically prominent Fechnerian model called *Luce's choice axiom* (Luce, 1959; Yellott, 1977), together with a unit-scaled logistic CDF, $F(x) = \frac{1}{1+e^{-x}}$, giving

$$P_{xy} = \frac{u(x)}{u(x) + u(y)}, \quad \text{with } u(x), u(y) > 0.$$

These two response models, tremble and Fechner, treat the decision maker's hypothetical constructs (preference, utility, strength of preference) as deterministic, and they create response probabilities through the introduction of various concepts of "error." Conceptually, they model heterogeneity in responses but not in preferences. The Fechnerian models, because they are quite specific, work most naturally with a theory that is, likewise, highly specific in its mathematical form, that is, a model in which every component is spelled out in its full and precise functional form. They also are only well defined if they are given a numerical hypothetical construct as input, such as the function u or the strength of preference $\mathbb{S}$ we have discussed above. Tremble models are less specific and require no numerical input; binary preference relations suffice. In that sense, tremble models are more flexible.[4]

The response models we reviewed so far have been generalized to a single broader class of "aggregation-based" specifications, according to which binary choice probabilities yield the hypothetical core deterministic preference at a suitably defined aggregate level (Regenwetter et al., 2014), such as "majority" (modal choice) or "supermajority" aggregation. Here, a hypothetical construct is only describing aggregate behavior, not necessarily every single choice made by a person. The key feature is that one or both of the following equivalences hold in tremble and Fechner models:

$$x \succ y \quad \Leftrightarrow \quad P_{xy} > \frac{1}{2} \quad \Leftrightarrow \quad u(x) > u(y). \tag{6}$$

A person is more likely to choose what he prefers than what he does not prefer. In the most general case where we consider all possible one-to-one functions u and, equivalently, all linear orders $\succ$, this representation is called the *weak utility model* (Luce & Suppes, 1965). It is equivalent to

$$\left[P_{xy} > \frac{1}{2}\right] \wedge \left[P_{yz} > \frac{1}{2}\right] \Rightarrow \left[P_{xz} > \frac{1}{2}\right]$$
$$\text{(for all distinct options } x, y, z), \tag{7}$$

labeled *weak stochastic transitivity*, because the right hand side of Condition 6 forces $\succ$ in the left hand side to be transitive, and therefore Condition 7 must hold for the central term of Condition 6. Regarding the right hand side equivalence of Condition 6, it is worth noting that it only requires that one specify the function u up to a monotonic transformation. Hence, for testing, the weak utility model (6) is very general and inclusive. But for estimation and prediction, it is not sufficiently specific to uniquely identify the function u used in most theories.

Another class of models, whose predictions overlap with, yet also differ from, aggregation-based specifications, and which is built on different conceptual and theoretical primitives, are "random preference," "random utility," and "random function" models (Becker, DeGroot, & Marschak, 1963; Block & Marschak, 1960; Loomes & Sugden, 1995; Marschak, 1960; Regenwetter & Marley, 2001). These follow from the premise that the preferences and utilities, rather than the responses, are probabilistic.

In a random preference model, one considers the collection $\mathcal{R}$ of all permissible preference relations, say, for instance, $\mathcal{R}$ might denote the collection of all binary preference relations $\succ$ that are consistent with Equation 1 and Condition 2 using some core family of functions v, Ψ, and some core operation $\odot$, such as, say, $v(A) = A^{\alpha}$, $\Psi(t) = \frac{1}{1+\delta t}$, and $\times$ for $\odot$. According to such a *random preference model*, there exists a probability measure $\mathbb{P}$ on the set of all parameter values for α and δ, such that, for x giving A with time delay t and y giving B with time delay s,

$$P_{xy} = \mathbb{P}\left(\{\alpha, \delta \mid u(x) > u(y)\}\right)$$
$$= \mathbb{P}\left(\left\{\alpha, \delta \,\middle|\, \frac{A^{\alpha}}{1+\delta t} > \frac{B^{\alpha}}{1+\delta s}\right\}\right). \tag{8}$$

[4] This makes them compatible with simple nonnumeric heuristics, for which Fechnerian models are ill-defined.

The most natural interpretation of a random preference model is that the decision maker, while fully consistent with a given core theory, is uncertain about her preferences and acts in accordance with a probability distribution over preference states that are consistent with that core theory, say, by sampling discount rates from a latent distribution. The formulation in Equation 8 makes it clear that this model can also be interpreted as a *random function model* (Regenwetter & Marley, 2001), because Equation 8 effectively makes $\mathbb{P}$ a probability measure on an appropriately defined measurable space of utility functions.

To see how much random preference models differ from tremble and Fechner models, consider, for a moment, the unusual choice between a larger, sooner and a smaller, later reward, a type of stimulus that is sometimes inserted into a study for quality control. If the respondent does not select the larger, sooner reward, this is sometimes interpreted as suggesting that he is not being attentive. Indeed, the random preference model predicts deterministic behavior in such a case because, no matter what the specific parameter values α and δ, the larger, sooner reward is preferred to the smaller, later reward: When $A > B$, $t < s$ in Equation 8, then the random preference model in Equation 8 yields $P_{xy} = 1$, regardless of the joint distribution on the values of α and δ. However, neither tremble nor Fechner models predict deterministic choice for such stimuli. Simply put, whereas a Fechner model derives probabilistic choice predictions from deterministic hypothetical constructs, a random preference model may, in certain cases, derive deterministic choice predictions from probabilistic hypothetical constructs.

A closely related *random utility model* specifies that the subjective values assigned to options x and y are uncertain. It captures this formally by defining jointly distributed random variables $\mathbf{U}_x$, $\mathbf{U}_y$ to denote the random utilities of options x and y. Using $\mathbb{P}$ to denote the probability measure governing the joint distribution of the random variables $\mathbf{U}_x$ (over all options x), assuming $\mathbb{P}(\mathbf{U}_x = \mathbf{U}_y) = 0$, $\forall x \neq y$, according to the random utility model,

$$P_{xy} = \mathbb{P}(\mathbf{U}_x > \mathbf{U}_y). \tag{9}$$

If, at every sample point of the underlying sample space, the joint realization of these random variables satisfies Conditions 1–2 with $\mathbf{U}_x$ substituted for $u(x)$, using a core family of functions $v(A) = A^{\alpha}$, $\Psi(t) = \frac{1}{1+\delta t}$, and $\times$ for $\odot$, then the choice probabilities in Equations 8 and 9 are the same. In particular, in such a random utility model, Equation 9 gives $P_{xy} = 1$ when x is a larger, sooner reward.

Just like many discounting models in the literature specify particular functions v and Φ, so do many random preference and random utility models specify properties of the probability measures $\mathbb{P}$ and/or the joint distribution of the random utilities. For example, the most commonly used random utility models assume multivariate normal distributions (probit) or extreme value distributions (logit), oftentimes for mathematical and statistical convenience. In both cases, $P_{xy} < 1$ in 'quality control' stimuli where x is a larger sooner reward. For very 'similar' stimuli, P_{xy} can, in fact, be 'close' to $\frac{1}{2}$. As we have seen earlier, these parametric random utility models are also Fechner models. However, the fully general class of random utility models makes no distributional assumptions.

Interplay Between Preference, Choice, and Heterogeneity

Even just within the paradigm of models of the form $u(x) = v(A) \odot \Psi(t)$ of Equation 1, we face a combinatorial explosion of possible models. A fully specified model of binary choice probabilities for this paradigm states the permissible functions v and Ψ and their permissible parameter values, as well as the permissible operations $\odot$, if it is to fully detail the deterministic core hypothetical constructs. In addition, one needs to consider a suitable response mechanism, such as, for example, upper bounds on permissible error rates ϵ_{xy}, an operation $\ominus$, a distribution function F. Or, if considering a probabilistic generalization of its core hypothetical constructs, it may need to spell out distributional assumptions about random preferences or random utilities.[5] The full range of these considerations has received

[5] For prior examples of such research programs, see Stott (2006) or Blavatskyy and Pogrebna (2010). These articles considered various combinations of core theory and probabilistic specification in the domain of risky choice.

little attention in intertemporal choice research because the latter has primarily focused on the algebraic core only.

For example, for monetary rewards, and $u(x) = v(A) \odot \Psi(t)$, $v(A) = A$, letting $\odot$ be the $\times$ operation, $\Psi(t) = \delta^t$, letting $\ominus$ be the $-$ operation, and F a normal CDF Φ with mean 0, we obtain a Thurstonian (probit) model of exponential discounting. Writing A_L, A_S for the larger and smaller rewards of L and S respectively, and t_L, t_S for the corresponding longer and shorter time delays, preference among L and S is deterministic, and responses probabilistic via

$$P_{LS} = \Phi\left(A_L\delta^{t_L} - A_S\delta^{t_S}\right). \tag{10}$$

In a random preference model, on the other hand, using the same deterministic core (but leaving out $\ominus$, which it does not use), preferences are probabilistic, and responses deterministic, via

$$P_{LS} = \mathbb{P}\left(\left\{\delta \mid A_L\delta^{t_L} > A_S\delta^{t_S}\right\}\right), \tag{11}$$

possibly with some constraints on the distribution of values of δ, say, a truncated normal distribution. Even though they are both grounded in standard exponential discounting, these two models have very different motivations: One is derived from assuming deterministic preference and probabilistic responses; the other is derived from deterministic responses based on probabilistic preferences. These models also feature drastically different mathematical properties; hence, they make distinctly different predictions about behavior. In other words, not only do they make different assumptions about the source and substantive meaning of heterogeneity, they also generate different predictions about the type of heterogeneity of behavior one may observe.

Here, we are particularly interested in the types of heterogeneity different models permit. A probability mixture of models each satisfying Equation 10 need not, itself, satisfy Equation 10: Consider $0 \le p_1, p_2, \ldots, p_k \le 1$ with $\sum_{i=1}^{k} p_i = 1$ and let $\delta_1, \delta_2, \ldots, \delta_k$ be distinct parameter values. Then, there generally does not exist a parameter value δ such that

$$\Phi\left(A_L\delta^{t_L} - A_S\delta^{t_S}\right) = \sum_{i=1}^{k} p_i\, \Phi\left(A_L\delta_i^{t_L} - A_S\delta_i^{t_S}\right),$$

which means that tests of this model cannot let choice probabilities change/drift excessively within a person over the course of an experiment, and one cannot safely pool data across respondents who differ in their core preferences. In contrast, mixtures of models, each satisfying the distribution-free form of Equation 11, do, in turn, satisfy Equation 11: Consider $0 \le p_1, p_2, \ldots, p_k \le 1$ with $\sum_{i=1}^{k} p_i = 1$ and let $\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_k$ be distinct probability measures. Then there always exists a probability measure $\mathbb{P}$ such that

$$\mathbb{P}\left(\left\{\delta \mid A_L\delta^{t_L} > A_S\delta^{t_S}\right\}\right) = \sum_{i=1}^{k} p_i\, \mathbb{P}_i\left(\left\{\delta \mid A_L\delta^{t_L} > A_S\delta^{t_S}\right\}\right),$$

namely, $\mathbb{P} = \sum_{i=1}^{k} p_i\mathbb{P}_i$. This means that these models permit high degrees of heterogeneity within and across individuals. On the other hand, distribution-free models like the one in Equation 11 can be mathematically intractable and most distribution-free random preference models require "order-constrained" statistical methods (Regenwetter, Dana, & Davis-Stober, 2011; Regenwetter et al., 2014).

There is, however, also much potential for model mimicry among models that are, like these, derived even from very different conceptual and mathematical primitives: While different probabilistic choice models make different predictions, it is important to note that some of their predictions usually overlap. For example, both Equation 10 and Equation 11 predict near-certain choice of L if $A_L\delta^{t_L} - A_S\delta^{t_S}$ is very large in Equation 10 and if Equation 11 places nearly all probability mass on δ-values for which $A_L\delta^{t_L} - A_S\delta^{t_S}$ is positive. In general, however, neither Equation 10 implies Equation 11 nor vice versa, that is, neither model is a special case of the other.

The literature on discounting models has made it quite clear that every detail about v, Ψ, and $\odot$ matters, and many articles are dedicated to discussing the details of the deterministic core structure (Doyle, 2013; Frederick et al., 2002). The literature on probabilistic response mechanisms, much of which has operated in

empirical paradigms outside intertemporal choice, has likewise highlighted that every detail about probabilistic response mechanisms matters, because misspecified response mechanisms lead to distortions of the deterministic core in statistical tests and in statistical estimation. Many articles are, in turn, dedicated to discussing the details of response mechanisms, primarily in risky choice (Birnbaum, 2011; Blavatskyy, 2011; Blavatskyy & Pogrebna, 2010; Hey, 2005; Iverson, 1990; Loomes et al., 2002; Luce, 1997; Stott, 2006; Wilcox, 2008). The intertemporal choice literature has much to gain from taking a similarly comprehensive look at sources of heterogeneity and how to model them beyond just refined deterministic cores.

Using the framework we provided above, one can select one or several specifications of hypothetical constructs, and one or several probabilistic specifications, to construct a collection of competing models of pairwise choice probabilities. One can then evaluate these competing models on suitably designed stimuli using the appropriate statistical methods. Exploring, testing, and statistically estimating every possible combination of fully specified deterministic and probabilistic components, even among a modest collection of cases like those we reviewed in the previous two subsections, poses formidable challenges. (a) Because of the many moving parts in a fully explicit theory, there can easily be thousands of combinations one may need to consider in a comprehensive analysis. (b) Models grounded in different or similar conceptual primitives need not imply the analogous similarities and differences in their probabilistic and statistical properties. (c) Different models differ strongly in their a priori flexibility to accommodate potential empirical data. (d) Parsimony in the model of hypothetical constructs can be completely disconnected from parsimony of the resulting choice model: Models with a larger number of parameters in the deterministic core need not be more flexible in their full probabilistic formulation. In fact, they can easily be more parsimonious in the space of permissible probabilistic responses. Hence, the standard approach of evaluating the parsimony of a theory by counting the number of parameters used by its deterministic functional specification is only a coarse heuristic. (e) Allowing for individual differences compounds the complexity and computational cost of reconciling preference, choice, and heterogeneity.

In light of these challenges, we proceed in a manner different from typical model selection approaches. Instead of considering specific functional forms for preferences, as is common in the literature, we abstract away to a core property shared by a large class of models for intertemporal preferences: transitivity of intertemporal preference. In other words, we follow a long tradition of axiom testing as a method to triage viable theories. Instead of considering specific functional forms of probabilistic response mechanisms, we abstract away to broad classes of probabilistic choice models. We create a collection of 20 models of pairwise choice probabilities by (a) varying whether we allow for one, some, or all transitive preferences; (b) varying whether we consider preferences, choices, or both to be probabilistic; and (c) varying the upper bounds on error probabilities where applicable. Applying these 20 models to several different stimulus sets and investigating their performance at both the individual and collective level allows us to document in detail the tradeoff between heterogeneity and parsimony.

Probabilistic Choice Models of Transitive Intertemporal Preference

We consider 20 probabilistic choice models of transitive intertemporal preference at various levels of parsimony (see also Figure 1). These 20 models form eight distinct model types. Four of these model types build on the theoretical premise that preferences, utilities, or strengths of preference are deterministic and that responses are probabilistic. These are the noisy-$\mathcal{P}$ (noisy patience), noisy-$\mathcal{I}$ (noisy impatience), noisy-$\mathcal{PI}$ (noisy patience-impatience), and noisy-$\mathcal{LO}$ (noisy linear order) models, each of which we consider with three different bounds on error rates. Two model types treat preferences as probabilistic and model responses as deterministic reflections of those preferences. These are the random-$\mathcal{LO}$ (random linear order) and the random-$\mathcal{LOT}$ (random linear order with tradeoffs) models. The other two model types are hybrids derived from the assumption that both preferences and responses are probabilistic. These are the noisy-$\mathcal{PI}$-mix (noisy patience-

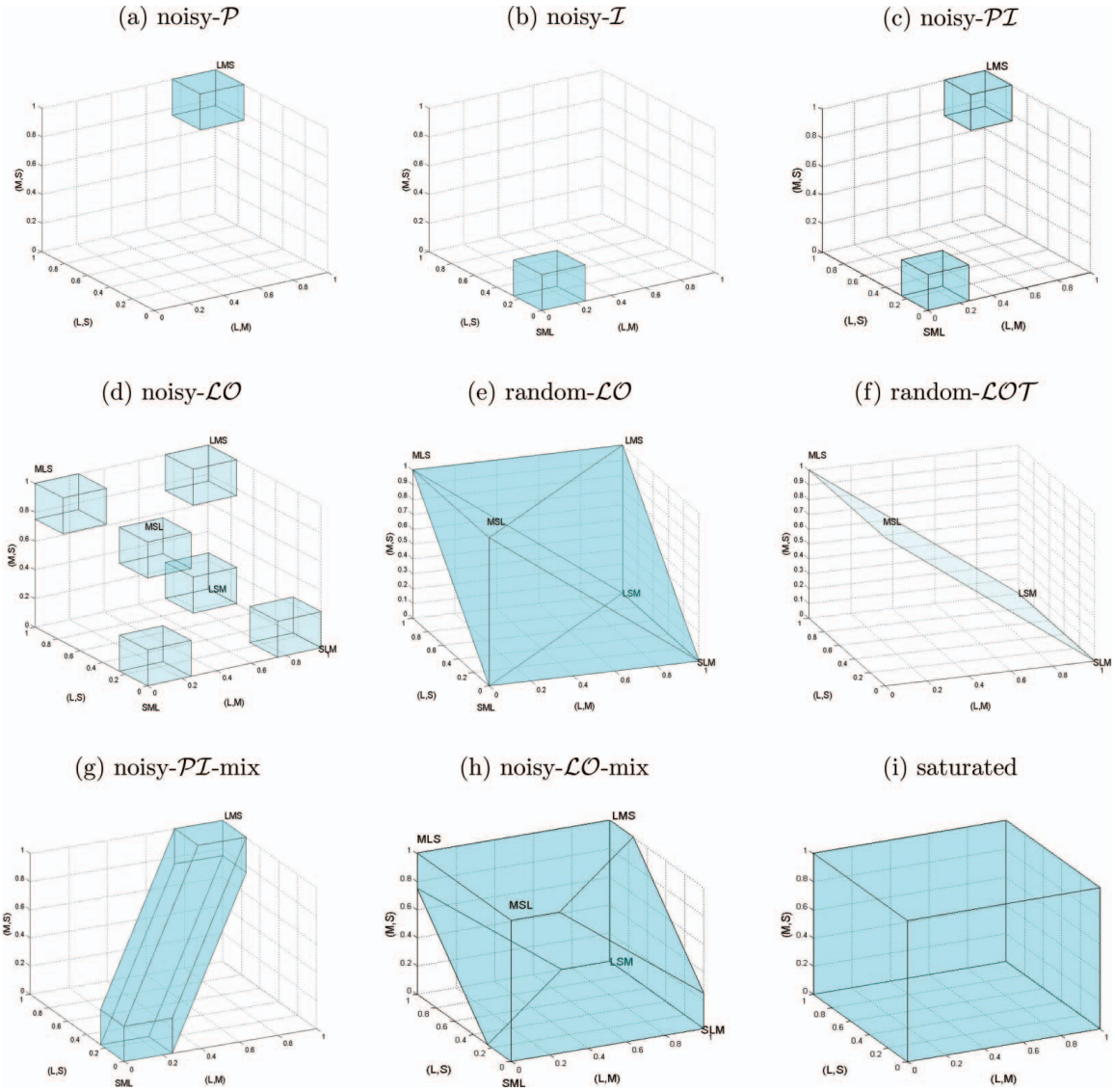

Figure 1. Eight types of probabilistic choice models for linear order intertemporal preferences and the saturated model. The coordinates are the choice probabilities P_{LM}, P_{LS}, P_{MS}. The shaded regions are the permissible choice probabilities for each model. The figure shows the case when $\tau = 0.25$ in (a)–(d), (g)–(h). Considering different upper bounds on error rates yields additional models in these cases. See the online article for the color version of this figure.

impatience mixture), and the noisy-$\mathcal{LO}$-mix (noisy linear order mixture) models, each of which we consider with three different bounds on error rates.

Deterministic Preferences Revealed Through a Probabilistic Response Process

We first consider a simple model in which a decision maker's preference corresponds to the linear order $\succ_A$ that rank orders the choice alternatives from most to least desirable reward, no matter the time delay. A possible reason for this could be that the differences in time delays used in a given study might be perceived as negligible, compared to the relative attractiveness of the rewards. Hence, this preference ordering could derive from a more highly structured mathematical model like the general class of models, shown in Equations 1 and 2, that we reviewed earlier: For example, the functions v

and Ψ of $u = v \odot \Psi$ might yield the linear order $\succ_A$ on the stimuli used in the study. For one collection of stimuli in our experimental study (our "Set 5" stimuli), this is the case, for example, when $v(A) = A$, $\Psi(t) = \frac{1}{1+\delta t}$ and $\odot = \times$, regardless of the discount parameter $\delta > 0$: Hyperbolic discounting makes very restrictive predictions about preferences for our Set 5. Alternatively, it could capture a simple "larger is better, no matter when" heuristic on some domain of stimuli. It is natural to suspect that the model may be limited to idiosyncratic data, that is, it may only perform well for certain stimuli and certain respondents.

The noisy-$\mathcal{P}$ model. Suppose that possible rewards are linearly ordered. An example would be distinct cash rewards, ordered from largest to smallest amounts. The *noisy-$\mathcal{P}$ model* (noisy patience model) states that the decision maker facing L versus S chooses the larger, later reward, L, regardless of time delay, up to random error. Formally, writing $\succ_A$ for the ordering of the options from best to worst reward and setting $0 < \tau \leq \frac{1}{2}$ as *upper bound* on the permissible error rate,

$$P_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ_A y, \\ \leq \tau & \text{if } y \succ_A x. \end{cases} \tag{12}$$

Special cases of noisy-$\mathcal{P}$. One possibility is a tremble model of $\succ_A$, according to which a decision maker has fixed preference $\succ_A$ and fixed probabilities ϵ_{xy} of making an error, with each $0 < \epsilon_{xy} < \tau$. The noisy-$\mathcal{P}$ model is more general in that only the upper bound τ on error rates is fixed, and error rates are permitted to vary, subject to the upper bound constraint. Hence, the error rates are not assumed to be statistically identifiable, nor are they assumed to be constant over time or across respondents. Alternatively, for monetary rewards, the decision maker might have a (fixed) utility function $u = v \odot \Psi$, which, when constrained to the options used in the study, happens to be monotonically increasing in the magnitude of the rewards. If L involves receiving A_L and S involves receiving only A_S, with $u > 0$, a specific Fechnerian (probit) model could state

$$P_{xy} = \Phi\left(ln\left(\frac{A_L^{\alpha}}{A_S^{\alpha}}\right)\right),$$

where Φ is a cumulative normal with mean zero. Here, the core theory models a decision maker consistent with a concave exponential utility function for money with exponent $\alpha < 1$, whose strength of preference is the ratio of subjective utilities. This model is also nested in the noisy-$\mathcal{P}$ model with $\tau = \frac{1}{2}$.

In sum, there are many possible ways to construct examples of the noisy-$\mathcal{P}$ model from either very specific or rather abstract assumptions about the subjective perception of rewards, the perception of time, the perception of the interplay between rewards and time, as well as a multitude of response mechanisms. No matter the details of such a construction, the model describes a patient decision maker with a deterministic core preference $\succ_A$ and noisy responses.

The noisy-$\mathcal{I}$ model. The *noisy-$\mathcal{I}$ model* (noisy impatience model) states that the decision maker chooses the smaller, sooner reward, S, regardless of the reward magnitude, up to random error. Formally, writing $\succ_t$ for the ordering of the options from soonest to latest, and setting $0 < \tau \leq \frac{1}{2}$ as *upper bound* on the permissible error rate,

$$P_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ_t y, \\ \leq \tau & \text{if } y \succ_t x. \end{cases} \tag{13}$$

Note that, for any S and L pair, we have $S \succ_t L$ and $L \succ_A S$. As was the case for the noisy-$\mathcal{P}$ model, the noisy-$\mathcal{I}$ model includes a multitude of nested submodels and, hence, abstracts away from a multitude of models about subjective perceptions of rewards, time, their interaction, and response mechanisms. Despite these abstractions, this model is rather restrictive in that it only permits one single core deterministic preference relation.

The noisy-$\mathcal{PI}$ model. The noisy-$\mathcal{P}$ model and the noisy-$\mathcal{I}$ model are extreme cases where either only the linear order of the options along the dimension of the reward or the dimension of time matters. A slight generalization, allows either $\succ_A$ or $\succ_t$ to be the underlying core deterministic preference, that is, it has a free pa-

rameter $\succ$ that may take two 'values,' namely either $\succ_A$ or $\succ_t$.

The *noisy-$\mathcal{PI}$ model* (noisy patience or impatience model) states that the decision maker is either consistently patient or consistently impatient, for a given stimulus set. More precisely, she either consistently prefers L to S, regardless of the time delays, or consistently prefers S to L, regardless of the monetary values, and chooses the preferred option up to random error. Setting $0 < \tau \leq \frac{1}{2}$ as *upper bound* on the permissible error rate,

$$\exists \succ \in \{\succ_A, \succ_t\} \quad \text{such that}$$
$$P_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ y, \\ \leq \tau & \text{if } y \succ x. \end{cases}$$

One attraction of this model is its potential to account for different stimulus sets in a very parsimonious fashion: A person may be patient for all stimuli in some stimulus sets and impatient for all stimuli in other stimulus sets. For example, for four of our stimulus sets, this model is a natural abstraction of hyperbolic discounting, that is, $\Psi(t) = \frac{1}{1+\delta t}$, $v(A) = A$ and $\odot = \times$. For our experimental stimulus collections labeled "Set 1" through "Set 4," hyperbolic discounting makes very restrictive predictions: In each case, regardless of the discount parameter δ, the resulting preference is either $\succ_A$ or $\succ_t$. However, one can specify a multitude of other models that would predict either $\succ_A$ or $\succ_t$, besides hyperbolic discounting.

The noisy-$\mathcal{LO}$ model. Moving beyond patience and impatience, we also consider richer models that permit true trade-offs among reward and time. The first model of this kind permits every linear order as a core preference (or, equivalently, permits every one-to-one utility function u). Like the noisy-$\mathcal{P}$ and noisy-$\mathcal{I}$ models, it features a free parameter τ that can be interpreted as the maximal permissible error rate. With the most generous choice of error bound, $\tau = \frac{1}{2}$, this model becomes the weak utility model (6), one of the staple probabilistic models used for testing transitivity of preferences in the literature (Tversky, 1969).

The *noisy-$\mathcal{LO}$ model* (noisy linear order model) states that there exists a fixed linear order $\succ$ of the options, such that the decision maker chooses in accordance with $\succ$, up to random error. The linear order in question is unknown to the experimenter and must be inferred from the data. Formally, writing $\mathcal{LO}$ for the collection of all linear orders of the options, and setting $0 < \tau \leq \frac{1}{2}$ as *upper bound* on the permissible error rate,

$$\exists \succ \in \mathcal{LO} \quad \text{such that}$$
$$P_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ y, \\ \leq \tau & \text{if } y \succ x. \end{cases}$$

The noisy-$\mathcal{P}$ model and the noisy-$\mathcal{I}$ model are both nested in the noisy-$\mathcal{LO}$ model: Because $\succ_A \in \mathcal{LO}$ and $\succ_t \in \mathcal{LO}$, if a person satisfies the noisy-$\mathcal{P}$ model or the noisy-$\mathcal{I}$ model then she also satisfies the noisy-$\mathcal{LO}$ model. The noisy-$\mathcal{P}$ model with $\tau = \frac{1}{2}$ is called "weak stochastic transitivity" (7) and the "weak utility model" (6) in the literature (Becker et al., 1963; Block & Marschak, 1960; Luce & Suppes, 1965; Marschak, 1960). Weak stochastic transitivity requires advanced order-constrained statistical methods (Iverson & Falmagne, 1985; Myung et al., 2005)[6] for a direct test. Tsai and Böckenholt (2006) tested a probabilistic intertemporal choice model on data of Roelofsma and Read (2000) and obtained choice probability estimates consistent with weak stochastic transitivity.[7] Dai (2014) tested weak stochastic transitivity directly using order-constrained Bayesian methods and found it to be well supported in an intertemporal choice task.

The noisy-$\mathcal{LO}$ model is clearly far less parsimonious than the noisy-$\mathcal{P}$ model, the noisy-$\mathcal{I}$ model, or the noisy-$\mathcal{PI}$ model because it is flexible enough to permit *any linear order* as deterministic core preferences (and *any* one-to-one utility function u). On the flip-side, this may enable us to model more respondents and more types of stimuli. At the same time, however, it is important to note that this model is highly sensitive to heterogeneity: Put simply, if we randomly select decision makers who each satisfy

[6] As Regenwetter et al. (2011) discuss in the context of risky choice, there are many published articles with inadequate tests of weak stochastic transitivity.

[7] Roelofsma and Read (2000) had interpreted their findings as evidence for intransitivity. Our R&R stimulus set uses stimuli similar to those of Roelofsma and Read (2000) to bring all 20 of our models to bear on that debate.

weak stochastic transitivity, and we let them make intertemporal choices, then their overall combined (pooled) choice probabilities typically violate weak stochastic transitivity.[8] In any probabilistic choice model with deterministic core preferences, heterogeneity across individuals and/or across time is a recipe for havoc. The same problem applies to the special cases in which linear orders are derived from functional forms: If a person's parameter values within a fixed functional form for, say, a discounting model, drift over the course of an experiment, then the person's overall choice probabilities may violate the noisy-$\mathcal{LO}$ model, even though every individual choice may have originated from that model. The same applies to interindividual differences: If two decision makers satisfy, say, probit models of hyperbolic discounting (i.e., models that satisfy weak stochastic transitivity), but they use different discount rates, then their averaged choice probabilities need not satisfy a probit model of hyperbolic discounting at all, and typically do not even satisfy weak stochastic transitivity.[9]

Probabilistic Preferences Revealed Through a Deterministic Response Process

Random preference and certain distribution-free random utility models start from fundamentally different premises than the four models we have just discussed. Here, the decision maker is uncertain about which option is preferable, yet, no matter which sample point of the underlying sample space is realized, the core theory is fully satisfied. Conditional on the momentary preference, the response is error-free.

The random-$\mathcal{LO}$ model. Binary choice probabilities satisfy the *random-$\mathcal{LO}$ model* (random linear order model) if there exists a probability distribution over linear orders such that the binary choice probability of choosing L over S is the total probability of those linear orders in which L is preferred to S. Formally, let $\mathcal{LO}$ denote the collection of all linear orders on a given set of choice options. Binary choice probabilities satisfy the random-$\mathcal{LO}$ model if there exists a probability distribution $\mathbb{P}$ on $\mathcal{LO}$, that is, $0 \leq \mathbb{P}(\succ) \leq 1, \forall \succ \in \mathcal{LO}$ and $\sum_{\succ \in \mathcal{LO}} \mathbb{P}(\succ) = 1$, such that

$$P_{xy} = \sum_{\substack{\succ \in \mathcal{LO} \\ x \succ y}} \mathbb{P}(\succ) \quad \text{(for all distinct options } x, y\text{)}.$$

This model is mathematically equivalent to the distribution-free random utility model (9) in that binary choice probabilities satisfy one model if and only if they satisfy the other (Block & Marschak, 1960).

The random-$\mathcal{LOT}$ model. We consider one more random preference model, namely the case in which all linear orders, except $\succ_A$ and $\succ_t$ are permissible preferences states. This model rules out the extreme cases of completely patient or completely impatient preference states. Let $\mathcal{LOT}$ denote the collection of all linear orders on a given set of choice options, except $\succ_A$ and $\succ_t$, that is, $\mathcal{LOT} = \mathcal{LO} \setminus \{\succ_A, \succ_t\}$. Binary choice probabilities satisfy the *random-$\mathcal{LOT}$ model* (random linear order with tradeoffs model) if there exists a probability distribution $\mathbb{P}$ on $\mathcal{LOT}$, such that

$$P_{xy} = \sum_{\substack{\succ \in \mathcal{LOT} \\ x \succ y}} \mathbb{P}(\succ) \quad \text{(for all distinct options } x, y\text{)}. \tag{14}$$

This model can also be restated in random utility terms. Binary choice probabilities satisfy Equation 14 if and only if there exist jointly distributed random variables, with $\mathbf{U}_x$ denoting the random utility of option x and $\mathbb{P}$ denoting the probability measure governing the joint distribution, with $\mathbb{P}(\mathbf{U}_x = \mathbf{U}_y) = 0, \forall x \neq y$, such that $\mathbb{P}\left(\bigcap_{r \succ_A s}^{r,s} \mathbf{U}_r > \mathbf{U}_s\right) = 0$ and $\mathbb{P}\left(\bigcap_{v \succ_t w}^{v,w} \mathbf{U}_v > \mathbf{U}_w\right) = 0$.

Probabilistic Preferences Compounded With Probabilistic Responses

We now consider a hybrid between the noisy-$\mathcal{P}$ model and the noisy-$\mathcal{I}$ model, and a hybrid of the random-$\mathcal{LO}$ model and the noisy-$\mathcal{LO}$ model. They follow from the general theoretical premise that preferences and responses are both probabilistic. Within an individual, this premise

[8] The weak utility model's sensitivity to heterogeneous populations is historically known as the famous *Condorcet paradox* of social choice theory (Condorcet, 1785).

[9] These observations follow trivially from the convexity or nonconvexity of various probability spaces.

can capture the idea that the individual is both uncertain about his preference and responds in a noisy fashion. At the group level, these models describe a heterogeneous population of up to three types of decision makers: those with deterministic preferences who respond in a noisy fashion, those with uncertain preferences who respond in a deterministic fashion, and those with uncertain preferences who also respond noisily. We limit ourselves to the two extreme cases where either only the two preferences $\succ_A$ and $\succ_t$ are permissible, or where all linear orders are permissible.

The noisy-$\mathcal{PI}$-mix model. Let $0 < \tau \leq \frac{1}{2}$ be an *upper bound* on the permissible error rate. Let P^A_{xy} denote the binary choice probabilities according to the noisy-$\mathcal{P}$ model (12) and let P^t_{xy} denote the binary choice probabilities according to the noisy-$\mathcal{I}$ model (13). According to the *noisy-$\mathcal{PI}$-mix model* (noisy patience-impatience mixture model), there exists a mixture probability p such that, in any given pairwise choice between L and S the person chooses according to the noisy-$\mathcal{I}$ model with probability p and according to the noisy-$\mathcal{P}$ model otherwise.[10]

$$\exists p \in [0,1] \quad \text{such that} \quad P_{xy} = pP^t_{xy} + (1-p)P^A_{xy},$$

$$\text{where } P^t_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ_t y, \\ \leq \tau & \text{if } y \succ_t x, \end{cases} \text{ and}$$

$$P^A_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ_A y, \\ \leq \tau & \text{if } y \succ_A x, \end{cases}$$

(for all distinct options x, y).

This model could, for example, model a population consisting of patient and impatient individuals only, with each decision maker also potentially making errors in his choices. Within person, it can model an individual who, for example, waivers between being patient and impatient, compounded with errors in her choices. This model is particularly interesting in that it does not connect to, say, discounting models, as easily as others. To satisfy this model, a population would have to consist of individuals whose discount rates are consistent with only the two preference rankings $\succ_A$ and $\succ_t$ on a given set of stimuli. As a discounting model of an individual, this would only allow discount rates according to which the individual either has preference $\succ_A$ or $\succ_t$. In our stimulus sets Set 1 to Set 5 (but not R&R), this is indeed the case for hyperbolic discounting: As we have seen earlier, hyperbolic discounting predicts $\succ_A$ or $\succ_t$ regardless of discount rate in those five stimulus sets. Other discounting models predict a larger variety of preferences.

The noisy-$\mathcal{LO}$-mix model. Our most complex (i.e., least statistically parsimonious) model permits a probability distribution over all possible linear order core preferences, compounded with noisy responses. Let $0 < \tau \leq \frac{1}{2}$ be an *upper bound* on the permissible error rate, and $\forall \succ \in \mathcal{LO}$, let $p_\succ$ denote the probability of making choices according to a noisy process with $\succ$ as core preference. Then the *noisy-$\mathcal{LO}$-mix model* (noisy linear order mixture model) states that

$$P_{xy} = \sum_{\succ \in \mathcal{LO}} p_\succ P^\succ_{xy} \quad \text{with } P^\succ_{xy}\begin{cases} \geq 1-\tau & \text{if } x \succ y, \\ \leq \tau & \text{if } y \succ x, \end{cases}$$

(for all distinct options x, y).

The noisy-$\mathcal{PI}$-mix model is a nested submodel of the noisy-$\mathcal{LO}$-mix model, in which $p_{\succ_A} = p = 1 - p_{\succ_t}$ and $P^{\succ_A} = P^A$, as well as $P^{\succ_t} = P^t$.

Summary of models. Figure 1 visualizes some of the similarities and differences between these models. Suppose that L is larger and later than M, which is, in turn, larger and later than S. The coordinates of the three-dimensional (3D) figure show binary choice probabilities P_{MS} on the vertical axis marked (M, S), P_{LM} on the axis marked (L, M) from the origin to the right, and P_{LS} on the axis marked (L, S) from the origin to the left. The deterministic core preferences correspond to corners (binary choice probabilities equaling 0 or 1) of the 3D cube. Despite being based on similar core premises about the hypothetical constructs of preferences or utilities, the models differ dramatically in their behavioral predictions. At the same time, probabilistic choice models that are built on different underlying premises overlap in complex ways. While

[10] Note that our formulation of this model does not permit p to vary with xy. However, because it forms a convex set, the model does allow *some* variation of p over time, including some degree of variation over repeated observations. Likewise, viewed as a model of a population, it allows for interindividual heterogeneity in the value of p.

the figure shows correctly which models are nested (such as noisy-$\mathcal{PI}$ in noisy-$\mathcal{PI}$-mix), it is important not to overinterpret the 3D visualization with respect to the parsimony of these models. Some of the models that appear to be relatively large in Figure 1 (such as random-$\mathcal{LO}$) rapidly become very restrictive in higher dimensions (i.e., they become more parsimonious when there are more than three choice probabilities). Likewise, some models that are very restrictive on just three choice probabilities may be less so in higher dimensions (e.g., random-$\mathcal{LOT}$ is only slightly more restrictive than random-$\mathcal{LO}$ in higher dimensions).

Table 1 summarizes our models from a different perspective. The first column lists the model names, whereas the second column shows the set of core preference states permitted by the core theory in each model. In addition to the eight models above, we also consider a *saturated* model that places no constraints whatsoever on binary choice probabilities. Its core theory is unconstrained in that it allows all (asymmetric) binary preference relations as preference states. We denote the set of all such binary preferences by $\mathcal{B}$. Columns 4 and 5 of Table 1 summarize whether preferences and responses are each deterministic or probabilistic. The last column gives each model a label that we use in our data analyses below. Models derived from probabilistic core preferences are shaded with a gray background. Models with deterministic response processes are marked in bold.

Model Specification for Bayesian Statistical Analysis

The premise of this article is threefold. (a) There are many moving parts to a fully specified model of intertemporal binary choice behavior, with much prior work discussing only unobservable hypothetical constructs in detail. (b) Different transitive models of observable intertemporal choice behavior vary in their parsimony. (c) We expect a tradeoff between the parsimony of a model and the variety of individuals and stimuli for which it can account, with the most parsimonious models likely working only for specific individuals and specific stimuli, and a universal model for all individuals and stimuli likely requiring extreme flexibility. In line with these conceptual expectations, we analyze our data from multiple perspectives. In contrast with most of the literature, our analyses are custom-designed to account formally for various levels and types of heterogeneity and parsimony.

We report all our analyses in Bayesian terms here and provide frequentist (hypothesis testing) analyses in the supplementary materials.[11] We use Bayesian p values (Gelman, Meng, & Stern, 1996) to assess model viability, Bayes factors (Kass & Raftery, 1995) to compare models at the level of each individual respondent, and group Bayes factors (GBFs; Stephan, Weiskopf, Drysdale, Robinson, & Friston, 2007) to aggregate Bayes factors across respondents. The magnitude of the Bayes factor between two models is the degree of evidence in favor of one model over the other. Our application of these methods to behavioral data follows similar recent analyses in the context of risky choice (Cavagnaro & Davis-Stober, 2014; Davis-Stober, Brown, & Cavagnaro, 2015; Guo & Regenwetter, 2014). In those studies, as in ours, models were defined through systems of linear inequality constraints on binary choice probabilities. Because Bayesian model selection requires that, in addition to constraints on choice probabilities such as those visualized in Figure 1, the models be cast via a likelihood function and a prior, we reformulate each set of inequality constraints using a prior distribution with support over only those probability vectors that are consistent with the model in question (see also Myung et al., 2005).

Formally, let $\mathcal{C}$ denote a collection of d distinct unordered pairs of choice options. For each pair $\{x, y\} \in \mathcal{C}$, let P_{xy} denote the binary choice probability of x being chosen from $\{x, y\}$, and let $\vec{P} = \{P_{xy}\}_{\{x, y\} \in \mathcal{C}}$ denote a binary choice probability vector (because each $P_{xy} = 1 - P_{yx}$, we only use/count one of these two probabilities for each pair $\{x, y\}$). Then, for each model q defined above, let $\Lambda_q \subseteq [0,1]^d$ denote the subset of

[11] Wherever both statistical approaches are applicable, our Bayesian and frequentist analyses are well aligned in the scientific conclusions that they support. The Bayesian approach is advantageous here: It naturally handles a situation like ours, in which some but not all models are nested within each other, and some models differ strongly in their parsimony despite having the same number of free parameters (here each model is characterized by 10 binomials).

Table 1
Summary and Notational Convention for the Models Under Consideration

Name	Fig. 1	Core Theory	Preferences	Response Process	Label
noisy-$\mathcal{P}$	(a)	$\{\succ_A\}$	Deterministic	Probabilistic	$\succ_A$
noisy-$\mathcal{I}$	(b)	$\{\succ_t\}$	Deterministic	Probabilistic	$\succ_t$
noisy-$\mathcal{PI}$	(c)	$\{\succ_A, \succ_t\}$	Deterministic	Probabilistic	$\succ_A \vee \succ_t$
noisy-$\mathcal{LO}$	(d)	$\mathcal{LO}$	Deterministic	Probabilistic	$\mathcal{LO}$
random-$\mathcal{LO}$	(e)	$\mathcal{LO}$	Probabilistic	**Deterministic**	$\boldsymbol{\mathcal{LO}}$
random-$\mathcal{LOT}$	(f)	$\mathcal{LO} \setminus \{\succ_{\mathcal{A}}, \succ_t\}$	Probabilistic	**Deterministic**	$\boldsymbol{\mathcal{LOT}}$
noisy-$\mathcal{PI}$-mix	(g)	$\{\succ_A, \succ_t\}$	Probabilistic	Probabilistic	$\succ_A \vee \succ_t$
noisy-$\mathcal{LO}$-mix	(h)	$\mathcal{LO}$	Probabilistic	Probabilistic	$\mathcal{LO}$
saturated	(i)	$\mathcal{B}$	–	–	

binary choice probability vectors $\vec{P}$ satisfying the inequality constraints that characterize model q, and let v_q denote the Lebesgue measure (i.e., volume) of Λ_q. We construct the Bayesian model M_q with a uniform prior over the model, that is, with the order-constrained prior distribution

$$\pi(\vec{P}\,|\,M_q) = \begin{cases} \dfrac{1}{v_q} & \text{if } \vec{P} \in \Lambda_q, \\ 0 & \text{otherwise,} \end{cases}$$

(for all $\vec{P} \in [0, 1]^d$).

Fully specified Bayesian models follow naturally by combining each order-constrained prior with a likelihood function, defined as follows. Let N_{xy} denote the number of times that the pair of delayed rewards $\{x, y\}$ is presented to the decision maker, let n_{xy} denote the number of times that x was chosen from $\{x, y\}$, and let $\vec{n} = \{n_{xy}\}_{\{x,y\}\in\mathcal{C}}$. Assuming that repeated choices from each option pair are identically distributed and that all choices are mutually independent,[12] the likelihood function f for a set of responses $\vec{n}$ takes the following product-of-binomials form:

$$f(\vec{P}\,|\,\vec{n}) = \prod_{x,y\in\mathcal{C}} \binom{N_{xy}}{n_{xy}} P_{xy}^{n_{xy}} (1 - P_{xy})^{N_{xy} - n_{xy}}. \tag{15}$$

In addition to the models we have already described, we also define a "saturated" model to serve as a common baseline against which to compare each substantive model. This model puts no constraints on binary choice probabilities, so it is defined by the prior $\pi(\vec{P}|\text{ saturated model}) = 1, \vec{P} \in [0, 1]^d$; that is, a uniform prior over the entire space of all choice probability vectors. This model is vacuous in the sense that it is guaranteed to fit any set of data perfectly. In model selection analyses that penalize for complexity, this model will receive the largest penalty because it is maximally complex. The saturated model provides a common benchmark for measuring the degree of evidence supporting or contradicting each substantive model. It also lets us define what it means for a substantive model to fail: If a model's Bayes factor against the saturated model is less than 1.0, then we are better off using the saturated model (i.e., no model) than the substantive model. If the Bayes factors of all our substantive models were less than 1.0, this would suggest that the data violated a fundamental assumption shared by these models, such as, for example, transitivity.

[12] In a Bayesian framework, the same likelihood function can be derived from different theoretical primitives about the data generating process and the interpretation of P_{xy}. In particular, one may assume that repeated choices on the same option pair are infinitely exchangeable and that choices on different choices pairs are independent. See Bernardo (1996) for discussion.

Experiments

We ran two studies aimed at evaluating the eight types of probabilistic choice models of transitive intertemporal preference. Decision makers made pairwise choices between larger, later and smaller, sooner options. The experiments were run in two locations: Urbana–Champaign, Illinois, USA and Berlin, Germany. In each location, we used six different stimulus sets to cover a range of different stimuli. One experiment collected enough repeated choices for the same stimuli from each person (mixed with a large number of distractors) to permit individual subject analyses. The other experiment drastically simplified the task by asking each respondent to make each pairwise choice only once. Hence, the second experiment does not provide enough data from each respondent for individual-level analyses.

Respondents

Respondent recruitment and testing took place at both the University of Illinois at Urbana–Champaign (UIUC) and the Max Planck Institute for Human Development (MPI). UIUC respondents were university students and local residents. MPI respondents attended a German university and chose to participate through their institute's experimental respondent pool. All respondents received monetary rewards based on choices they made during the experiment and they only learned their reward amount after completion of the experiment. In accord with payment standards at the University of Illinois, UIUC respondents also received an additional base payment ($12 for Experiment 1 and $8 for Experiment 2).

Before experimental testing, we selected a subset of trials from which all rewards would be paid. These preselected trials all had relatively high reward amounts, thus ensuring sufficient remuneration. Each respondent's particular reward was determined by randomly selecting one of these preselected trials. Respondents were not informed about the mechanism by which we selected stimuli that were used for payment and whether this selection was made before or after data collection. Respondents were explicitly instructed at the beginning of the experiment to make choices based on their true preferences because they would receive one of their chosen time-delayed rewards as a real payment. We paid UIUC respondents with the exact delay specified (even if the date fell on a weekend or holiday) by implementing a payment system via an agreement between the university and Amazon.com. After the experiment was over, respondents provided an email address to which an electronic Amazon gift code (matching the U.S. dollar value of their chosen reward) was sent on the specified calendar day in the future (matching the delay of their reward). The MPI offered respondents two options at the end of the study. If the real reward was an option that included a positive time delay, respondents could opt to receive 85% of the amount in cash immediately instead of waiting for the delayed full reward. Respondents were not told that they could substitute this immediate payment until they had completed all choices. If they opted for the full delayed reward, they received it after the specified delay through a bank transfer in euros.[13]

For Experiment 1 (at UIUC), we tested 31 respondents (14 males, 17 females) from June to October 2012 with a mean ± *SD* age of 20.8 ± 2.4 years (range 18–28). At MPI, we tested 30 respondents (16 males, 14 females) from June to July 2012 with a mean ± *SD* age of 25.6 ± 3.7 years (range 20–34). For Experiment 2 (at UIUC), we tested 34 respondents from September to November 2013. Age and gender of these respondents was not recorded. At MPI, we tested 30 respondents (10 males, 20 females) from November to December 2013 with a mean ± *SD* age of 25.3 ± 2.6 years (range 20–30).

Experimental Procedure

The UIUC Institutional Review Board and the Ethics Committee of the MPI reviewed and approved both experiments.[14]

Procedure. Respondents completed the experiments on computers. UIUC respondents saw English text and U.S. dollars for currency, whereas MPI respondents saw German text and euros for currency but identical numbers as did the U.S. respondents (not currency-converted values). Respondents could first provide their

[13] The supplementary materials provide the instructions to respondents and the stimuli used for real payment.

[14] University of Illinois at Urbana–Champaign Institutional Review Board approval #11427.

age, gender, and occupation.[15] They then read one set of instructions, completed 10 practice trials, and then read a final set of instructions before beginning the actual trials. This final instruction set informed each respondent that their reward at the end of the experiment would be determined by one of the choices made during the study. For each trial, the respondents used a computer mouse to select one of two options presented on the screen, each characterized by a specified reward amount and a time delay. At the end of the experiment, respondents were then shown the reward that they were going to receive.[16]

Experiment 1 consisted of two sessions with 1,006 trials each (including six warm-up trials). At UIUC, respondents completed the two sessions of Experiment 1 on 2 different days. MPI respondents completed the two sessions for Experiment 1 on a single day, separated by a 5- to 15-min break. Each session of Experiment 1 took respondents 30–90 min to complete. While Experiment 1 was designed to elicit enough information from each person to permit within-respondent data analyses, Experiment 2 was aimed at collecting the same kind of data with a much smaller number of trials, for a joint analysis of all respondents combined. It had a single session with 106 questions (six of them warm-up trials) and took respondents 10–30 min to complete. Respondents in Experiment 2 saw the same questions as respondents in Experiment 1, except that none of the items were repeated.

Stimuli. We created six option sets. Sets 1–5 each consisted of five intertemporal options (top of Table 2). The sets varied in the magnitude and spread of monetary amounts (stated in $ and €) and in the magnitude and spread of time delays (stated in days). For each set of five options, we created all 10 possible pairwise combinations of options to create 10 different *option pairs* per option set. Across all five sets of stimuli, this resulted in a total of 50 option pairs. We also used an additional collection of nine option pairs. We adapted one triple from stimuli in Roelofsma and Read's (2000) study of intransitivity in intertemporal choice, and two additional such triples were similar but varied and expanded the range of reward amounts. This sixth stimulus set of nine option pairs is labeled R&R (bottom of Table 2).

For Experiment 1, to permit within-respondent statistical analysis, respondents saw each of the 59 option pairs 20 times,[17] yielding 1,180 experimental trials. These 1,180 trials were mixed with another 832 pairs of stimuli, some of which were designed to test other hypotheses while others served as distractors. The 2,012 pairs of stimuli were divided into *blocks*, each consisting of a series of five consecutive option pairs. Within each block, we randomized the order of presentation across respondents. The order of the blocks was constant across respondents. Each block contained two or three experimental pairs, but never from the same stimulus set. We placed option pairs from the same set in alternating blocks, so respondents saw 5–13 other pairs between experimental pairs from the same stimulus set. Respondents were shown 95–103 option pairs before experiencing a repetition of the same pair.

It is natural to question whether making in excess of 1,000 decisions per session could bias a respondent's behavior and yield unrealistic data. We tested this concern empirical by running a second experiment with the same stimuli, but with a small number of individual trials per person. Hence, for Experiment 2, where we did not aim to carry out individual respondent statistical analyses, respondents saw each of the 59 option pairs exactly once. They were also given another 47 distractor pairs. The method of sequencing the presentations of these option pairs matched that of Experiment 1.

Results of Experiment 1

We tested all eight model types, as illustrated three-dimensionally in Figure 1. For noisy-$\mathcal{P}$,

[15] This step was accidentally omitted by the person administering Experiment 2 at UIUC.

[16] The experimental software used was a custom-made program called *Disk'n'Risk*, developed by Uwe Cziens-kowski at MPI. The supplementary materials give further experimental details.

[17] We repeated each option pair 20 times to accommodate a frequentist analysis. If we only ran the Bayesian analysis, we could cut this by a factor of 3. For example, Davis-Stober et al. (2015) used eight repetitions per option pair in a ternary choice experiment. Some parametric models, such as logit and probit models work without repetitions of the same stimuli and, instead, use many different stimuli for statistical convergence.

Table 2
Six Stimulus Sets

Set 1 options		Set 2 options		Set 3 options		Set 4 options		Set 5 options	
Money	Days	Money	Days	Money	Days	Money	Days	Money	Days
3	4	1	1	14	23	1	1	9	80
5	28	5	21	15	27	3	4	11	83
7	52	9	41	16	31	5	7	13	86
9	76	13	61	17	35	7	10	15	89
11	100	17	81	18	39	9	13	17	92

R&R pairs				
S			*L*	
Money	Days	Versus	Money	Days
7	7	vs.	8	14
7	7	vs.	10	49
8	14	vs.	10	49
10	16	vs.	12	18
10	16	vs.	15	25
12	18	vs.	15	25
4	13	vs.	5	16
4	13	vs.	11	22
5	16	vs.	11	22

Note. In Sets 1–5, we considered all 10 possible distinct *S* vs. *L* pairs among the five listed options. In R&R, we considered the nine listed *S* vs. *L* pairs.

noisy-$\mathcal{I}$, noisy-$\mathcal{PI}$, noisy-$\mathcal{PI}$-mix, noisy-$\mathcal{LO}$, noisy-$\mathcal{LO}$-mix, we furthermore used three different bounds τ on error rates: $\tau = 0.5$ (modal choice, which contains Fechnerian models, such as logit and probit specifications, as special cases), $\tau = 0.25$ (whose maximum error rate is considered adequate by some scholars, e.g., Harless & Camerer, 1994), and $\tau = 0.1$ (according to which errors are not a major component of the response process). All in all, therefore, we tested 20 different transitive probabilistic models of intertemporal choice. All of our analyses require order-constrained statistical inference, implemented in the public domain software QTest, programmed for multiple computing platforms[18] (Regenwetter et al., 2014).

Are Transitive Models Viable?

We first assess the overall viability of each model for each respondent and stimulus set by computing the Bayesian *p* values (Gelman et al., 1996). The Bayesian *p* value is a posterior predictive check of the descriptive adequacy of each model. It is computationally inexpensive and relatively easy to interpret. Essentially, the Bayesian *p* value is computed by comparing the observed data to the posterior predictive distribution of the model. If the observed data are consistent with the posterior predictive distribution, then the Bayesian *p* value is high; otherwise, it is low (see Myung et al., 2005, for details on computation). A standard approach is to declare an *adequate fit* of a model to the data whenever the Bayesian *p* value exceeds 0.05. The Bayesian *p* value does not indicate the probability that a model is correct. Bayesian *p* values cannot be compared across models. We use Bayesian *p* values only to determine the proportion of respondents for whom each model provides at least an adequate fit, and we leave model selection for later.

We computed the Bayesian *p* value of each model separately for each respondent and stimulus set. Figure 2 shows, for each model and stimulus set, the proportion of respondents for

[18] The original (frequentist only) release of QTest is available at www.regenwetterlab.org. A new multicore compatible version with Bayesian capabilities is available from the authors while it is being prepared for public release.

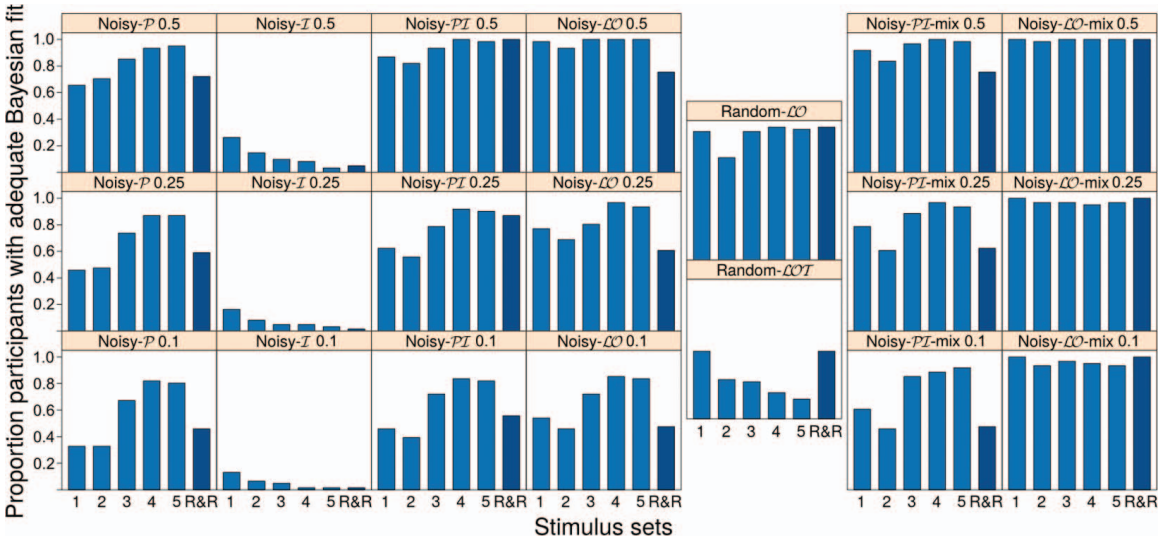

Figure 2. Bayesian p values in Experiment 1. Each panel shows the results for one model, with the level of τ indicated in the header after the model name (where applicable). Each panel reports the proportion of respondents (out of 61) with adequate fits (Bayesian p value > 0.05), on the vertical axis, separately for the six stimulus sets. See the online article for the color version of this figure.

whom that model provided an adequate fit (frequentist fits are available in Figure S1 in the supplementary materials). Overall, there seem to be several transitive models that provide adequate fits for most respondents and most stimulus sets. The most complex model, in which all linear orders are permissible preference states and in which responses can be maximally noisy, the noisy-$\mathcal{LO}$-mix model with $\tau = \frac{1}{2}$, provides an adequate fit for nearly every respondent in every stimulus set. On the one hand, this means that transitive models can account almost universally for our data across respondents and stimulus sets. On the other hand, the three instances of the noisy-$\mathcal{LO}$-mix are among the most statistically complex of the models we have tested, and the Bayesian p value does not penalize models for complexity.

In contrast, the noisy-$\mathcal{I}$ models at all noise bounds were inadequate for all but a few respondents in each stimulus set, casting doubt on this model's viability as an explanation of the data at any level of the error bounds. However, because this model is especially parsimonious relative to the others, especially at the 0.1 noise bound, it is possible that a noisy-$\mathcal{I}$ model could provide the best explanation for those respondents and stimulus sets in which its Bayesian p value exceeded 0.05.

The random-$\mathcal{LO}$ model fits a large proportion of respondents. When the $\succ_A$ and $\succ_t$ options are removed in the random-$\mathcal{LOT}$ model, however, the fit drops dramatically. The large decrease in fit caused by the removal of these preference states suggests that linear orders based exclusively on either amount or time played a key role in the good performance of the random-$\mathcal{LO}$ model.

The noisy-$\mathcal{P}$ models seem to show the greatest interaction across stimulus sets, especially at the 0.25 and 0.1 noise bounds, as they are adequate for most respondents in Stimulus Sets 3, 4, and 5, but fewer than half of the respondents in Sets 1, 2, and R&R. Similar patterns of interaction emerge for the noisy $\mathcal{LO}$ models, noisy-$\mathcal{PI}$ models, and noisy-$\mathcal{PI}$-mix models, especially those with lower error bounds τ. These results raise the question whether respondents' behavior may be best described by different models in different stimulus sets, with an overall model across stimulus sets requiring some flexibility. To answer this question more conclusively, we proceed to the model selection analysis.

Model Selection Results: Individual Level Analyses

Our next goal is to identify the best model at the individual level, before we proceed to the

group level. Our criterion for model selection is the Bayes factor (Kass & Raftery, 1995), defined as the ratio of the marginal likelihoods of two models, derived from Bayesian updating. The Bayes factor accounts for both goodness-of-fit and complexity/parsimony. It selects among models based on generalizability (Pitt & Myung, 2002), in that the model with the highest Bayes factor is the one deemed to most accurately predict future data samples from the same process that generated the currently observed sample (see, e.g., Liu & Aitkin, 2008).

To identify the best model at the individual level, we computed the Bayes factor of each model, relative to the saturated model, separately for each respondent and stimulus set.[19] With 20 models, 61 respondents, and six stimulus sets in our study, this analysis yielded a total of 7,320 respondent-level Bayes factors. Our Bayes factors varied across many orders of magnitude (the Bayes factors for each model, respondent, and stimulus set are available in a spreadsheet that is part of the supplementary materials). Many Bayes factors were quite large and, hence, provided strong evidence in favor of the model in question. However, likewise, in many cases, the evidence against a given model was quite strong: Of the nearly 3,000 Bayes factors that were smaller than 1.0, nearly half (1,450) had log_{10} values between -10 and -200. Of these, 984 were for the noisy-$\mathcal{I}$, 223 were for the noisy-$\mathcal{P}$, 131 were for the noisy-$\mathcal{PI}$, 58 were for the noisy-$\mathcal{PI}$-mix and 54 were for the noisy-$\mathcal{LO}$. Table 3 summarizes the results by reporting key features of the best model for each respondent and stimulus set. The features are identified using the labels introduced in Table 1. For example, the best model for Respondent 1 in Set 1 in the UIUC sample is noisy-$\mathcal{PI}$-mix, which assumes probabilistic preferences and choices. So, the corresponding cell is shaded to indicate probabilistic preferences and it shows the core theory $\{\succ_A, \succ_I\}$ in plain text (rather than bold) to indicate probabilistic choices. For simplicity, the table uses the same label for all models with the same core theory, preferences, and response process, regardless of error bound (e.g., noisy-$\mathcal{PI}$-mix with $\tau = 0.5$ and noisy-$\mathcal{PI}$-mix with $\tau = 0.1$).

Perhaps the most prominent aspect of Table 3 is the apparent heterogeneity across respondents and stimulus sets. No single core theory, type of preference, or type of response process was robust across all respondents and stimulus sets. In fact, not only was there heterogeneity in terms of the best model, there was also heterogeneity in terms of which models were adequate. That is to say, no model had a Bayes factor greater than 1.0 for every respondent and stimulus set, meaning that every model failed on at least one respondent and stimulus set, relative to the saturated model (see the spreadsheet in the supplementary materials for the Bayes factor of each model, respondent, and stimulus set). This does *not* mean all of the models failed overall, as there were only very few cases (eight out of 366 respondent-by-stimulus combinations, indicated by the black shaded boxes in Table 3) in which none of the 20 models had a Bayes factor greater than 1.0. Nevertheless, the eight cases in which the saturated model was favored represent instances in which transitivity (a core assumption shared by all 20 models under consideration) may have been violated. In the current modeling framework, a violation of transitivity means that the core theory of the best model includes one or more intransitive preferences. Interestingly, four of the apparent violations involved just two respondents: UIUC Respondent 14 and MPI Respondent 22; and six of them involved just one stimulus set: Set 2. This clustering of apparent violations within certain experimental conditions and respondents is consistent with the findings of Cavagnaro and Davis-Stober (2014) and suggests that the violations may represent robust individual differences.

[19] In general, Bayes factors of inequality constrained models cannot be obtained analytically. However, in this particular case, we were able to obtain analytical solutions for the Bayes factors of noisy-$\mathcal{P}$, noisy-$\mathcal{I}$, noisy-$\mathcal{PI}$, and noisy-$\mathcal{LO}$, relative to the saturated model. This is because the inequality constraints are orthogonal within each of these models, and the priors on each dimension are independent and conjugate to the likelihood function. We obtained respondent-level Bayes factors for the remaining models, in which the order constraints are not orthogonal, using Monte Carlo integration. To compute PBFs, we used a specialized procedure developed by Klugkist and Hoijtink (2007). In short, this algorithm yields the Bayes factor for an order-constrained model versus the saturated model by drawing a large sample from the posterior distribution of the saturated model and computing the proportion of the sample that satisfies the order constraints of the nested model (see Cavagnaro & Davis-Stober, 2014, for additional details).

Table 3
Experiment 1: Core Theory, Preference Structure, and Response Process Structure of the Best Model for Each Respondent in Each Stimulus Set

	University of Illinois at Urbana-Champaign sample							Max Planck Institute for Human Development sample						
Respondent	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Set 5**	**R&R**	**Respondent**	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Set 5**	**R&R**
1	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-1-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-2-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-2-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-3-	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_t$	3	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-4-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-4-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-5-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	5	$\mathcal{LO}$	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$
-6-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-6-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
7	$\succ_t$	$\succ_t$	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	7	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
8	$\mathcal{LO}$	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	8	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$
-9-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-9-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
10	$\succ_A \vee \succ_t$	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	-10-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
11	$\succ_t$	$\mathcal{LO}$	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	-11-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
12	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	12	$\succ_t$	$\mathcal{LO}$	$\succ_t$	$\succ_A$	$\succ_A \vee \succ_t$	$\mathcal{LO}$
-13-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-13-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
14			$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	14	$\mathcal{LO}$		$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$
15	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-15-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
16	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_A$	$\mathcal{LO}$	-16-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
17	$\mathcal{LO}$		$\mathcal{LOT}$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	-17-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
18	$\succ_t$	$\mathcal{LO}$	$\mathcal{LO}$	$\succ_A$	$\mathcal{LO}$	$\mathcal{LO}$	18	$\mathcal{LO}$		$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-19-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-19-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
20	$\succ_t$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	20	$\mathcal{LO}$		$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-21-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	21	$\succ_t$	$\mathcal{LO}$	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$	$\mathcal{LO}$
22	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$	22			$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$
23	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_t$	$\succ_t$	$\mathcal{LO}$	-23-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
24	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	-24-	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$	$\succ_A \vee \succ_t$
25	$\succ_A$	$\mathcal{LO}$	$\succ_A \vee \succ_t$	$\succ_A$	$\mathcal{LO}$	$\succ_A$	25	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-26-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-26-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
27	$\mathcal{LO}$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-27-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-28-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-28-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
-29-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	-29-	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$	$\succ_A$
30	$\succ_t$	$\succ_t$	$\succ_A \vee \succ_t$	$\mathcal{LOT}$	$\succ_A$	$\mathcal{LO}$	30	$\succ_A \vee \succ_t$	$\mathcal{LOT}$	$\succ_A \vee \succ_t$	$\succ_A$	$\succ_A$	$\succ_A$
31	$\succ_t$	$\succ_t$	$\succ_A$	$\succ_A$	$\succ_A$	$\mathcal{LO}$	-	-	-	-	-	-	-

Note. Those 31 cases where the best model matches across all six stimulus sets are enclosed in hyphens, such as -2-.

Although no core theory was best across the board, $\succ_A$ most frequently performed best (264 of 366 entries in Table 3), indicating that most respondents seem to prefer the option with the highest amount, regardless of the time delay. This was especially the case in Stimulus Sets 4 and 5, in which all but eight respondents were best described by a model assuming core theory $\succ_A$. In contrast, fewer than two thirds of respondents were best described by $\succ_A$ in Stimulus Sets 1, 2, and R&R. Despite these variations across stimulus sets, we found that about half (31 out of 61) respondents were best described by the same core in all six stimulus sets (these are marked in Table 3 with respondent numbers enclosed in hyphens, e.g., -2-). This consistency suggests that the best core theory may be somewhat robust across stimulus sets, within some respondents.

Like any model selection analysis on experimental data, our analysis is specific to the models, participants, and stimuli considered. The fact that $\succ_A$ accounts well for some stimulus sets but not others suggests that it is worthwhile considering core theories that agree with $\succ_A$ on some stimulus sets but not others. In our Roadmap section, we discuss how to evaluate a variety of core theories using the same general approach, and with appropriate stimuli.

Model Selection Results: Group-Level Analyses

To select among models at the group level, we use two measures: the GBF (Stephan et al., 2007) and the pooled Bayes factor (PBF). Both select among models at the group level, but they differ in the mechanism by which respondent-level results are aggregated: the PBF aggregates *data* across respondents, whereas the GBF aggregates *likelihoods* across respondents. The PBF is the ratio of the marginal likelihoods of two models given the pooled data from all respondents, whereas the GBF is the product of respondent-level Bayes factors. Thus, the model with the highest PBF is the one that best accounts for the pooled data, while the model with the highest GBF is the one that *jointly* best accounts for each respondent's data.[20]

Table 4 ranks each model based on the GBF and PBF, respectively, in each stimulus set (the log_{10} transformed GBF and PBF values are reported in Table S2). For pooled data, it only makes sense to evaluate models which, if there is more than one core deterministic preference, can inherently accommodate heterogeneity of preferences. Formally, these are models whose parameter spaces form convex sets, that is, we must omit the noisy-$\mathcal{PI}$ model and the noisy-$\mathcal{LO}$ model (such as weak stochastic transitivity).

The noisy-$\mathcal{PI}$-mix model was by far the most successful, according to both the GBF and PBF, in almost all stimulus sets. The exceptions were Set 2, in which noisy-$\mathcal{LO}$ was best according to GBF, and R&R, in which noisy-$\mathcal{LO}$ and noisy-$\mathcal{P}$ were best according to the GBF and PBF, respectively. What is most notable about this result, besides the near-unanimity across stimulus sets, is that noisy-$\mathcal{PI}$-mix assumes probabilistic preferences, whereas a vast majority of respondents were best described as having deterministic preferences. These results are not contradictory, as they may seem at first, because probabilistic preferences at the group level need not imply that every decision maker in the group has uncertain preferences. Rather, probabilistic preferences at the group level implies that the sample comprises a heterogeneous mix of up to three types of decision makers: those with deterministic preferences who respond in a noisy fashion, those who have uncertain preferences and respond in a deterministic fashion, and those who have uncertain preferences and respond in a noisy fashion. The group-level analyses cannot identify the nature of the heterogeneity more precisely because they do not distinguish between variability within respondents (such as, preference uncertainty) and variability between respondents (such as, individual differences in core preferences).

Despite the limitations of the group-level analyses, they are essential for obtaining results that generalize beyond each particular decision maker. The current GBF results suggest that the model that will generalize best to data from a randomly selected respondent is noisy-$\mathcal{PI}$-mix. Although this model implies probabilistic pref-

[20] This interpretation of the GBF rests on two assumptions: that every respondent has the same model (i.e., the same set of restrictions on choice probabilities, but not necessarily the same choice probabilities) and that the model evidences are independent. The latter assumption is tenable for GBFs as long as respondents are sampled independently from the population.

Table 4
Experiment 1: Ranking of Each Model From Best to Worst, in Terms of the Joint (Group Bayes Factor [GBF]) and Pooled (Pooled Bayes Factor [PBF]) Analyses, in Each Stimulus Set (Column), Combined Across Locations

		Joint (GBF)						Pooled (PBF)					
Model	τ	Set 1	Set 2	Set 3	Set 4	Set 5	R&R	Set 1	Set 2	Set 3	Set 4	Set 5	R&R
Noisy-$\mathcal{P}$	.10	(19)	(18)	(18)	(17)	(16)	(18)	(11)	(9)	(11)	**2**	**1**	(10)
Noisy-$\mathcal{P}$	.25	(17)	(17)	(17)	(16)	9	(16)	(10)	(8)	**2**	**3**	**3**	(9)
Noisy-$\mathcal{P}$	.50	(16)	(15)	(15)	10	10	(13)	**2**	**2**	4	5	5	**1**
Noisy-$\mathcal{I}$	.10	(21)	(21)	(21)	(21)	(21)	(21)	(11)	(9)	(12)	(12)	(12)	(10)
Noisy-$\mathcal{I}$	.25	(20)	(20)	(20)	(20)	(20)	(20)	(11)	(9)	(12)	(12)	(12)	(10)
Noisy-$\mathcal{I}$	.50	(18)	(19)	(19)	(19)	(19)	(19)	(11)	(9)	(12)	(12)	(12)	(10)
Noisy-$\mathcal{PI}$	.10	(15)	(16)	(14)	**2**	**2**	(17)	—	—	—	—	—	—
Noisy-$\mathcal{PI}$	.25	(12)	(13)	**3**	**3**	**3**	(15)	—	—	—	—	—	—
Noisy-$\mathcal{PI}$	.50	**3**	**3**	4	7	7	(9)	—	—	—	—	—	—
Noisy-$\mathcal{LO}$	.10	(14)	(12)	(13)	4	5	5	—	—	—	—	—	—
Noisy-$\mathcal{LO}$	.25	5	**1**	5	6	6	**1**	—	—	—	—	—	—
Noisy-$\mathcal{LO}$	.50	6	**2**	7	9	11	**2**	—	—	—	—	—	—
Random-$\mathcal{LO}$		9	(11)	(11)	(14)	(15)	4	4	(7)	6	7	7	4
Random-$\mathcal{LOT}$		(13)	(14)	(16)	(18)	(18)	(12)	4	(9)	(12)	(12)	(12)	**3**
Noisy-$\mathcal{PI}$-mix	.10	**1**	(10)	**1**	**1**	**1**	(14)	(11)	(9)	**1**	**1**	**2**	(10)
Noisy-$\mathcal{PI}$-mix	.25	**2**	(6)	**2**	5	4	(11)	**1**	(9)	**3**	4	4	(10)
Noisy-$\mathcal{PI}$-mix	.50	4	(5)	6	8	8	(10)	**3**	**1**	5	6	6	**2**
Noisy-$\mathcal{LO}$-mix	.10	7	(9)	8	11	13	**3**	6	**3**	7	8	8	5
Noisy-$\mathcal{LO}$-mix	.25	8	(7)	9	12	12	6	7	4	8	9	9	6
Noisy-$\mathcal{LO}$-mix	.50	(11)	(8)	(12)	(15)	(17)	(8)	8	5	9	10	10	7
Saturated		10	4	10	13	14	7	9	6	10	11	11	8

Note. Rankings in parentheses are worse than the saturated model in the same stimulus set. Ties are given identical ranks. For ease of reading, the three best models (**1**, **2**, and **3**) are marked in boldfaced font.

erences $\succ_A$ and $\succ_t$, we can see from the respondent-level results, in Table 3, that it is unlikely for a randomly selected respondent to be best described by such a model (most are best described by models with deterministic preference $\succ_A$). However, because there are individual differences, the randomly selected respondent may be best described as having deterministic preference $\succ_A$, or deterministic preference $\succ_t$, both of which are part of $\succ_A \bigvee \succ_t$. Thus, noisy-$\mathcal{PI}$-mix is selected by the GBF because it is deemed to provide the most parsimonious account that is consistent with the behavior of most respondents.

It also stands out that noisy-$\mathcal{PI}$-mix does well in only four of the six stimulus sets, whereas noisy-$\mathcal{LO}$ does well in Set 2 and R&R. In fact, the only models that beat the unconstrained model across all six stimulus sets are noisy-$\mathcal{LO}$ with error rates of 0.25 and 0.5. This suggests that generalizing across multiple stimulus sets requires more preference patterns than just $\succ_A$ and $\succ_t$. This result highlights the importance of the choice of stimulus sets when testing models of intertemporal choice. If one is only concerned with modeling choices on a narrow set of stimuli, such as those in Sets 3–5, then a small set of preference patterns may suffice. However, generalizing to a broader set of stimuli may require additional preference patterns, perhaps even intransitive patterns. Identifying the minimal set of preference patterns that generalizes to any arbitrary stimulus sets is beyond the scope of this article. Later, in the Roadmap section, we provide additional guidance on investigating this issue.

The PBF results suggest a slightly different interpretation than the GBF results. Because the PBF is based on pooled data, the model selected by the PBF is the one that is deemed to generalize best to future pooled data. That is, it may not be representative of any particular respondent, but it parsimoniously captures the aggregate choice proportions. This distinction be-

tween the PBF and the GBF helps to explain why the noisy-$\mathcal{P}$ model fares well according to the PBF but not the GBF. The noisy-$\mathcal{P}$ model fares well according to the PBF because, in the pooled data, any influence from the minority of respondents whose choices are not consistent with noisy-$\mathcal{P}$ (i.e., those in Table 3 whose best core theory was not $\succ_A$) is washed out by the vast majority of respondents whose choices are best described by noisy-$\mathcal{P}$. On the other hand, the GBF is not based on pooled data, but rather aims to simultaneously describe each respondent's choice proportions. Thus, the noisy-$\mathcal{P}$ model does not fare well according to the GBF, because the noisy-$\mathcal{P}$ model provides such an extremely poor account of the choice data from those respondents who were best described by other models (e.g., those in Table 3 whose best core theory was $\succ_t$).

Results of Experiment 2

Experiment 2 aimed to diagnose systematic changes in respondent behavior caused by the number of questions. For instance, the large number of choices in Experiment 1 might have led decision makers to switch their decision-making strategy from a compensatory strategy to a simple heuristic of attending only to either reward or time. Thus, in Experiment 2, each respondent saw and made a choice on each option pair only once, not 20 times as in Experiment 1. The drawback is that these data do not permit fine-grained individual level analyses. We interpret the models as describing between-subjects heterogeneity and we focus on pooled analyses. Like in the pooled analysis of Experiment 1, it only makes sense to evaluate convex models (that inherently accommodate heterogeneity of preferences wherever multiple core preferences are allowed).

Table 5 gives the model rankings in each stimulus set, according to the PBF, for Experiment 2 (the log transformed Bayes factor values are available in Table S3). Notably, the rankings in this table nearly match those of Experiment 1 in the right panel of Table 4. In particular, the best model in each stimulus set in Experiment 2, according to the PBF, is either noisy-$\mathcal{P}$ or noisy-$\mathcal{PI}$-mix. These models fare well at nearly all τ

Table 5
Experiment 2: Ranking of Each Model From Best (Highest Pooled Bayes Factor) to Worst (Lowest Pooled Bayes Factor) in Each Stimulus Set

Model	τ	Set 1	Set 2	Set 3	Set 4	Set 5	R&R
Noisy-$\mathcal{P}$	.10	(13)	(12)	(12)	**1**	**1**	(12)
Noisy-$\mathcal{P}$	.25	(11)	4	**3**	**3**	**3**	(10)
Noisy-$\mathcal{P}$	.50	6	**1**	4	5	5	**1**
Noisy-$\mathcal{I}$	.10	(15)	(15)	(15)	(15)	(15)	(15)
Noisy-$\mathcal{I}$	.25	(14)	(14)	(14)	(14)	(14)	(14)
Noisy-$\mathcal{I}$	.50	(12)	(13)	(13)	(13)	(13)	(13)
Noisy-$\mathcal{PI}$	.10	—	—	—	—	—	—
Noisy-$\mathcal{PI}$	.25	—	—	—	—	—	—
Noisy-$\mathcal{PI}$	.50	—	—	—	—	—	—
Noisy-$\mathcal{LO}$	.10	—	—	—	—	—	—
Noisy-$\mathcal{LO}$	.25	—	—	—	—	—	—
Noisy-$\mathcal{LO}$	.50	—	—	—	—	—	—
Random-$\mathcal{LO}$		5	7	6	8	(8)	5
Random-$\mathcal{LOT}$		4	(11)	8	(12)	(12)	(8)
Noisy-$\mathcal{PI}$-mix	.10	**2**	5	**1**	**2**	**2**	(11)
Noisy-$\mathcal{PI}$-mix	.25	**1**	**2**	**2**	4	4	(9)
Noisy-$\mathcal{PI}$-mix	.50	**3**	**3**	5	6	6	**2**
Noisy-$\mathcal{LO}$-mix	.10	7	6	7	7	(11)	**3**
Noisy-$\mathcal{LO}$-mix	.25	8	8	9	(10)	(10)	4
Noisy-$\mathcal{LO}$-mix	.50	9	9	10	(11)	(9)	(7)
Saturated		10	10	11	9	7	6

Note. Rankings in parentheses are worse than the saturated model in the same stimulus set. Ties are given identical ranks. For ease of reading, the three best models (**1**, **2**, and **3**) are marked in boldfaced font.

levels. None of the other models fares particularly well in any stimulus set or with any τ level, with the exception of noisy-$\mathcal{P}$-mix in the R&R stimulus set.

To put these results into perspective, recall from Experiment 1 that we found heterogeneity between subjects was best characterized by a mixture of two types of respondents: those attending only to time and those attending only to reward amount (noisy-$\mathcal{P}$ and noisy-$\mathcal{PI}$-mix were the best explanations of the pooled data). If this pattern were merely a consequence of the large number of choices made by each respondent in Experiment 1 then we would expect to see a different pattern in Experiment 2. Because model selection favors the same core in both experiments, we see no reason to suspect a dramatic change. Note that this evidence is only suggestive and not a formal implication, because the aggregate choice proportions do not uniquely identify the mixture components. This is an inherent weakness of analyzing pooled data and the key reason why one can only draw conclusions about individual behavior if one gathers sufficient data from the individual. For instance, choice proportions that are consistent with noisy-$\mathcal{PI}$-mix are also consistent with mixtures of other core theories besides just $\succ_A$ and $\succ_t$. It is possible for noisy-$\mathcal{PI}$-mix to be the best model according to the GBF even when the data are generated by some mixture of compensatory strategies. This problem is particularly vexing for models like noisy-$\mathcal{PI}$-mix, because vectors of choice proportions that are near one half on every dimension can be generated by nearly limitless combinations of deterministic components. However, in Experiment 2 we actually found that noisy-$\mathcal{P}$ was the best model in four out of six stimulus sets, with $\tau = 0.1$ in one case. The geometry of the parameter space makes it implausible that aggregate data could favor noisy-$\mathcal{P}$ with $\tau = 0.1$ unless the vast majority of individual respondents actually chose according to that model.

Roadmap

This article has been about the interplay between heterogeneity and parsimony in modeling intertemporal preferences. To highlight how this issue affects model selection, we have focused specifically on transitive intertemporal preference. Furthermore, instead of considering the menagerie of specific, parametric, transitive theories, we have considered a handful of more general, parameter-free models that are characterized by subsets of viable linear ordered preferences. In particular, we have considered the 'extreme' cases where either just one or two, or all linear orders were considered viable. However, for a given set of stimuli, a parametric theory of the form $u(x) = v(A) \odot \Psi(t)$ typically falls between these two extremes by predicting potentially many, but not all, linear orders as permissible preferences. Other types of theories furthermore predict preferences other than linear orders, such as intransitive preferences. Next, we briefly discuss a roadmap for studying competing theories in a way that formally accounts for heterogeneity. Future analyses of discounting models and intransitive models alike can emulate our approach of modeling either the core preferences, or the responses, or both, as probabilistic processes. Future work can also leverage order-constrained inference methods for statistical inferences and model selection to tackle the complex trade-off between parsimony and heterogeneity. Without much loss of generality and for ease of exposition for rest of this section, we concentrate on the scenario in which two or more theories of the form $u(x) = v(A) \odot \Psi(t)$ compete against each other.

Stimulus Design

Our Stimulus Sets 1–5 are 'standard' intransitivity stimuli in which two attributes trade-off against each other in equal steps as we move through the list of stimuli (similar to the lotteries of Tversky, 1969, in risky choice). Stimulus Set R&R was based on a prior article on intransitivity of intertemporal preference. If, instead of transitivity, one were rather interested in specific theories of the form $u(x) = v(A) \odot \Psi(t)$, then stimulus design could leverage the specifics of those theories to create choice options that are diagnostic among the theories under consideration. To distinguish these theories, one should use stimuli for which different theories predict minimally overlapping sets of preference patterns. In addition, if the primary goal is to test competing theories (i.e., to either validate or falsify each theory in its own right), one should design the stimuli in such a way that

each theory under consideration would also permit as few distinct preference patterns as possible so as to create maximally parsimonious predictions. On the other hand, if the goal is to estimate and identify parameters, say, discount rates, with maximal precision, then one should design stimuli that are maximally diagnostic in that regard, namely, stimuli that lead to many different preference patterns as one varies the discount rate of each theory. In so doing, one ensures that each preference pattern is consistent with only a small range of parameter values of the core theory, say, a narrow range of discount rates. In addition, stimulus design also depends on the type of heterogeneity one wants to either accommodate or critically test.

Heterogeneity

The type of heterogeneity one wants to account for has strong implications for the type of probabilistic model and level of data aggregation that are suitable. For example, if each individual decision maker satisfies a logit model, but there are individual differences in the parameters of this logit model, then the population generally does not satisfy a logit model because the average of logit probabilities need not be logit probabilities. More generally, if each individual has a core deterministic preference or utility function and only responses are probabilistic, it usually does not make sense to model the population with a single deterministic core preference or utility function, unless it makes sense to treat preferences or utilities as unanimous.

If one were to compare, say, exponential and hyperbolic discounting, it would be advisable to consider multiple different specifications. The first step would be to identify, for the given stimulus set, the set of linear orders that are consistent with exponential and hyperbolic discounting by varying their free parameters. Then, one could consider probabilistic models of the following types.

1. Like our noisy-$\mathcal{P}$, noisy-$\mathcal{I}$, noisy-$\mathcal{PI}$, and noisy-$\mathcal{LO}$ models, it would make sense to consider models with deterministic core preferences that are defined by precisely those linear orders that are consistent with the discounting model at hand, and responses are modeled probabilistically. In addition to the distribution-free error specifications we used, many models of the form $u(x) = v(A) \odot \Psi(t)$, including discounting models, interface naturally with Fechnerian specifications, such as logit and probit models. It is important to reiterate that many of these specifications can be hard to interpret as models of individual behavior if applied exclusively to data pooled across individuals, unless one is willing to assume that those individuals are unanimous in their underlying preferences or utilities.

2. Like our random-$\mathcal{LO}$ and random-$\mathcal{LOT}$ models, it would make sense to consider random preference models that permit a probability distribution over precisely those preference states that are permitted by a given core theory. Because these models feature convex parameters spaces, they can model both within and between person heterogeneity. Interesting parametric special cases to consider, say for exponential and hyperbolic discounting, are random preference models constructed via a parametric distribution over the permissible discount rates in each core theory.

3. Last but not least, like our noisy-$\mathcal{PI}$-mix and noisy-$\mathcal{LO}$-mix models, it is worth considering hybrid models that permit heterogeneity in the preference states consistent with each given core theory, as well as probabilistic error in responses.

Model Selection Criteria

In our analysis, we have emphasized the interplay of heterogeneity and parsimony. In addition to multiple different criteria for goodness-of-fit, we have leveraged the Bayes factor as a model selection tool that is well-suited to quantify parsimony of probabilistic models and to select among models that, like ours, are neither disjoint nor nested. The same methods are useful also for model competitions more generally, including among models based on a core representation of the form $u(x) = v(A) \odot \Psi(t)$. For parametrized theories like that, there are many additional tools available for model selection. For example, some probabilistic models, especially Fechnerian models, naturally plug

into adaptive design optimization methods (Cavagnaro, Gonzalez, Myung, & Pitt, 2013) at the individual level. Furthermore, when using models to estimate core parameters, such as an individual's discount rate, it is natural to test the validity of parameter estimates through prediction to new data sets on different stimuli (e.g., the generalization criterion of Busemeyer & Wang, 2000).

Sketch of a Model Selection Study

We briefly sketch how our roadmap would help design a study aimed at diagnostic design that facilitates replication studies while balancing heterogeneity with parsimony. Table 6 sketches an example of a model competition between exponential and hyperbolic discounting. Imagine that a lab plans a study consisting of a three-stage competition between these two core theories. In Stage I, the lab proposes a set of stimuli that balances two types of diagnosticity. (a) By permitting only few different preference patterns under either theory, it places empirical pressure on both theories. (b) By predicting rather different collections of preference patterns from the two theories, it helps distinguish exponential from hyperbolic discounting. The lab includes several different nonparametric probabilistic models that broadly model probabilistic preferences, or probabilistic responses, or both. The lab also plans frequentist and Bayesian analyses on several different levels, including individual level and group level analyses. In Stage II, the study focuses on the "best performing" core theory from Stage I to attempt to estimate and identify discount rates.

Table 6
Sketch of an Example Model Competition Between Exponential and Hyperbolic Discounting

Theory, models, stimuli	Considerations or Recommendations
Stage I: Theory testing and screening	
Algebraic core	Exponential versus hyperbolic discounting
Stimuli	Permit few preference patterns overall
	Preference patterns diagnostic between these theories
Deterministic preferences and probabilistic responses	Supermajority specification with three different error bounds
Probabilistic preferences and deterministic responses	Random preference over permissible preference states
Probabilistic preferences and probabilistic responses	Hybrid model (convex hull of the previous two)
Stage II: Identifying discount rates for best theory from Stage I	
Algebraic core	Best theory from Stage I
Stimuli	Permit many preference patterns
Fixed discount rate and probabilistic responses	Supermajority specification with three different error bounds
	Logit, probit, Luce, and other Fechnerian models
Probabilistic discount rate and deterministic responses	Parametric random preference over permissible preference states induced by a normal distribution over discount rates
Probabilistic discount rate and probabilistic responses	Hybrid models
Stage III: Generalizability to new stimuli	
Algebraic core	Same as Stage II
Stimuli	Permit few patterns based on parameter estimates of Stage II
Model of heterogeneity	Best from Stage II
Data	Statistical method
Types of analyses in each stage	
Within subject	Frequentist p, Bayes p, Bayes factor
Pooled	Frequentist p, Bayes p, Bayes factor
Other	Group Bayes factor, hierarchical Bayes models

Note. A study like this can be preregistered. It specifies how theories compete, what sources of heterogeneity are permissible, and how they are modeled.

The stimuli for this stage are designed to be maximally diagnostic for that core theory by permitting a broad array of preference patterns as a function of the discount rate. The probabilistic specifications now also include a variety of parametric special cases of the specifications in Stage II. Parametric Fechnerian and random preference models lend additional structure that can help identify discount rates more precisely than the earlier nonparametric models. Depending on the source of heterogeneity, the goal is to obtain either a 'best' single discount rate from each individual or a parametric distribution of each individual's discount rates, or to estimate a population level distribution over discount rates through a variety of probabilistic models and statistical procedures. A major component of Stage II is to evaluate whether and how the 'best' discount rate (point estimate or estimated distribution) varies with the assumed source and the model of heterogeneity. Finally, Stage III is a generalizability study that critically tests the 'best' core theory, 'best' probabilistic specification, and 'best' parameters from Stages I and II on additional stimuli. These stimuli are dependent on the results of Stages I and II and are designed to place maximal pressure on the hypothesized theory, model of heterogeneity, and parameters from Stages I and II. The quantitative performance in all three stages can be evaluated with similar methods.

Conclusions

Heterogeneity causes great challenges in measuring and predicting individual preferences and choices. A common way to think of heterogeneity is that different decision makers might differ in their parameter values (such as their discount rates) within a shared theoretical account (such as exponential discounting) or that a given decision maker might differ in her parameter values for different types of stimuli. Another common way of tackling heterogeneity is to relax restrictions on the functional form of a given theory without changing the probabilistic specification or the response mechanism. Rather than spelling out a refined theory of choice behavior, such approaches pursue increasingly complicated theories of hypothetical constructs. The common practice of inferring parameter values (e.g., discount rates) of a 'prototypical' decision maker from pooled binary choice data of heterogeneous decision makers is rarely grounded in an explicit and compelling model of heterogeneity.

A common way to think of parsimony of a theory is to count the number of parameters in the deterministic core of a theory (and to earmark one or more additional parameters for noise or for heterogeneity of parameter values). Counting parameters is only a coarse heuristic in characterizing how flexible or inflexible a theory is in accounting for potential empirical data. As a case in point, on our Set 5, hyperbolic discounting with one free parameter in the algebraic core permits just one preference state, namely $\succ_A$, regardless of the discount parameter. On the other hand, for exponential discounting, which also has one free parameter in the algebraic core, we have found 11 different linear orders, depending on the discount rate. Hence, if we are interested in testing theories empirically, we must keep a close eye on the interplay between the functional form, the probabilistic specification, as well as the stimuli we use in a given study, to account for parsimony in a suitable fashion when analyzing our data. A more rigorous account of model complexity, rather than counting parameters of an algebraic functional form, is to spell out the sources of heterogeneity mathematically and to quantify the flexibility with which the resulting probabilistic model accommodates possible data as a function of the stimuli used.

Here, we aimed to abstract away from distributional assumptions and parametric accounts of heterogeneity and parsimony in intertemporal choice. We focused instead on general characterizations of two crucially important sources of heterogeneity in choices on a given stimulus: the latent intertemporal preferences and the response process. In particular, we considered that the latent preferences may be probabilistic or the responses (based on a given preference) may be probabilistic, or both processes may be probabilistic. While these types of processes have a long history of scientific study, they have been largely neglected in intertemporal choice research. Even though our models differ strongly in their parsimony, every one can be characterized by 10 order-constrained binomial parameters. We have taken a Bayesian approach to quantifying model complexity.

We found that the core preferences $\succ_A$ and $\succ_t$ appeared to drive the performance of the

winning models in most cases, suggesting that models draw most of their strength from being able to predict simple patterns of behavior, such as always preferring the highest reward or always preferring the shortest time. However, developing a robust model of intertemporal choice requires attention to a number of issues besides just the core preferences permitted by the underlying theory. Our various levels and types of analyses have shown that both model performance and model selection are sensitive also to the chosen stimulus set, the assumed response process, and whether we analyze data within each individual, jointly across many individuals (GBF), or pooled from many individuals (PBF). We did not find evidence for systematic differences between the U.S and the German study. Also, even though respondents in Experiment 1 each had to handle 20 times as many questions as respondents in Experiment 2, we did not find evidence for systematic differences between the two experiments.

References

Arfer, K., & Luhmann, C. (2015). The predictive accuracy of intertemporal-choice models. *British Journal of Mathematical and Statistical Psychology, 68,* 326–341.

Becker, G., DeGroot, M., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science, 8,* 41–55.

Bernardo, J. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences, 4,* 111–122.

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

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

Birnbaum, M., & Navarrete, J. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17,* 49–78.

Blavatskyy, P. (2011). A model of probabilistic choice satisfying first-order stochastic dominance. *Management Science, 57,* 542–548.

Blavatskyy, P., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25,* 963–986.

Block, H., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford, CA: Stanford University Press.

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171–189.

Carbone, E., & Hey, J. (2000). Which error story is best? *Journal of Risk and Uncertainty, 20,* 161–176.

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

Cavagnaro, D. R., Gonzalez, R., Myung, J. I., & Pitt, M. A. (2013). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science, 59,* 358–375.

Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* [Essay on the application of the probabilistic analysis of majority vote decisions]. Paris, France: Imprimerie Royale.

Dai, J. (2014). *Using test of intransitivity to compare competing static and dynamic models of intertemporal choice* (Doctoral thesis). Indiana University, Bloomington, Indiana.

Dai, J., & Busemeyer, J. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General, 143,* 1489–1514.

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

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

Doyle, J. (2013). Survey of time preference, delay discounting models. *Judgment and Decision Making, 8,* 116–135.

Doyle, J., & Chen, C. (2012). *The wages of waiting and simple models of delay discounting*. SSRN scholarly paper, Social Science Research Network, Rochester, NY.

Ebert, J., & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Management Science, 53,* 1423–1438.

Ericson, K., White, J., Laibson, D., & Cohen, J. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting. *Psychological Science, 26,* 826–833.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, XL, 351*–401.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733–760.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130,* 769–792.

Guo, Y., & Regenwetter, M. (2014). Quantitative tests of the perceived relative argument model: Commentary on Loomes (2010). *Psychological Review, 121,* 696–705.

Harless, D., & Camerer, C. (1994). The predictive value of generalized expected utility theories. *Econometrica, 62,* 1251–1289.

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

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

Iverson, G. (1990). *Probabilistic measurement theory* (Tech. Rep. No. MBS 90–23). Irvine, CA: University of California, Irvine.

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

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

Killeen, P. (2009). An additive-utility model of delay discounting. *Psychological Review, 116,* 602–619.

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

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 112,* 443–477.

Leland, J. (2002). Similarity judgments and anomalies in intertemporal choice. *Economic Inquiry, 40,* 574–581.

Liu, C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology, 52,* 362–375.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics, 107,* 573–597.

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

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

Luce, R. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

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

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

Luce, R., & Narens, L. (1994). Fifteen problems in the representational theory of measurement. In P. Humphreys (Ed.), *Patrick Suppes: Scientific philosopher, Vol. 2: Philosophy of Physics, theory structure, measurement theory, philosophy of language, and logic* (pp. 219–245). Dordrecht, the Netherlands: Kluwer.

Luce, R., & Suppes, P. (1965). Preference, utility and subjective probability. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology: Volume III* (pp. 249–410). New York, NY: Wiley.

Manski, C., & McFadden, D. (Eds.). (1981). *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press.

Manzini, P., & Mariotti, M. (2006). A vague theory of choice over time. *Advances in Theoretical Economics, 6,* 1–27.

Marschak, J. (1960). Binary-choice constraints and random utility indicators. In K. Arrow, S. Karlin, & P. Suppes (Eds.), *Proceedings of the first Stanford Symposium on Mathematical Methods in the Social Sciences, 1959* (pp. 312–329). Stanford, CA: Stanford University Press.

Mazur, J. (1984). Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 426–436.

Mazur, J. (1987). An adjusting procedure for studying delayed reinforcement. In M. Commons, J. Mazur, J. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: The effect of delay and of intervening events on reinforcement value,* Vol. 5 (pp. 55–73). Hillsdale, NJ: Erlbaum.

McCausland, W., & Marley, A. (2014). Bayesian inference and model comparison for random choice structures. *Journal of Mathematical Psychology, 62,* 33–46.

McClure, S., Ericson, K., Laibson, D., Loewenstein, G., & Cohen, J. (2007). Time discounting for primary rewards. *The Journal of Neuroscience, 27,* 5796–5804.

McFadden, D. (2001). Economic choices. *American Economic Review, 91,* 351–378.

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

Pitt, M., & Myung, I. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6,* 421–425.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty, 23,* 5–32.

Read, D. (2004). Intertemporal choice. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 424–443). Malden, MA: Blackwell.

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

Regenwetter, M., Davis-Stober, C., Lim, S., Guo, Y., Popova, A., Zwilling, C., . . . Messner, W. (2014). QTest: Quantitative testing of theories of binary choice. *Decision, 1,* 2–34.

Regenwetter, M., & Marley, A. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology, 45,* 864–912.

Roelofsma, P., & Read, D. (2000). Intransitive intertemporal choice. *Journal of Behavioral Decision Making, 13,* 161–177.

Rubinstein, A. (2003). "Economics and psychology"? The case of hyperbolic discounting. *International Economic Review, 44,* 1207–1216.

Samuelson, P. (1937). A note on measurement of utility. *Review of Economic Studies, 4,* 155–161.

Scholten, M., & Read, D. (2006). Discounting by intervals: A generalized model of intertemporal choice. *Management Science, 52,* 1424–1436.

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review, 117,* 925–944.

Stephan, K., Weiskopf, N., Drysdale, P., Robinson, P., & Friston, K. (2007). Comparing hemodynamic models with DCM. *Neuroimage, 38,* 387–401.

Stevens, J. R. (2016). Intertemporal similarity: Discounting as a last resort. *Journal of Behavioral Decision Making, 29,* 12–24.

Stott, H. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32,* 101–130.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34,* 273–286.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., . . . Wilkens-Diehr, N. (2014). XSEDE: Accelerating scientific discovery. *Computing in Science & Engineering, 16,* 62–74.

Tsai, R.-C., & Böckenholt, U. (2006). Modelling intransitive preferences: A random-effects approach. *Journal of Mathematical Psychology, 50,* 1–14.

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

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J. Cox & G. Harrison (Eds.), *Risk aversion in experiments, Volume 12: Research in experimental economics* (pp. 197–292). Bingley, United Kingdom: Emerald.

Yellott, J. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology, 15,* 109–144.

Received August 14, 2015
Revision received August 29, 2016
Accepted September 1, 2016 ■

Appendix E: Supplementary File

In the 'SupplementalFiles' folder, the folder 'Chapter5' includes all the predicted patterns and facet-defining inequalities and equalities for 49 versions of CPT for Experiments 2009 and 2012; the file 'Chapter6SupplementMaterials.pdf' includes all the predicted patterns and facet-defining inequalities and equalities of all the decision models.