Evaluating Decision Makers’ Preferences via Lexicographic Semiorders

Clintin P. Davis-Stober¹, Nicholas Brown¹, and Daniel R. Cavagnaro²

¹Department of Psychological Sciences, University of Missouri
²Mihaylo College of Business and Economics, California State University Fullerton

Abstract

Two divergent theories regarding the algebraic structure of preferences are the strict weak-order (i.e., utility) representation, and the lexicographic semiorder representation. We carry out a novel comparison of these theories by formulating them as mixture models of ternary choice that are general yet parsimonious. We apply Bayesian model selection to see which representation (if any) best explains each decision maker’s choices in an existing dataset, as well as in our own new experiment, which also tests the robustness of each representation with respect to manipulations of stimuli, display format, and time pressure. We find that a majority of participants are best described by strict weak-ordered preferences with a substantial minority best described by lexicographically semiordered preferences. More than 96% of choice datasets could be rationalized by at least one of the mixture models, suggesting that previously reported violations of one representation or the other could be explained by individual differences in the underlying preference structure. Among our experimental manipulations, we find only a weak main effect of display format, suggesting that, after accounting for individual differences, the algebraic structure of preferences is a fairly robust property of human decision making.

Key Words: Preferential Choice, Choice Modeling, Lexicographic Semiorders, Weak Orders, Convex Polytopes, Random Preference.

1 Introduction

When buyers choose from among different bundles of goods and services, their choices are assumed to be based on underlying states-of-mind called preferences. In order to generate tractable models of these preferences, theorists commonly impose idealized conditions on them. One such condition is that preferences can be represented by a unidimensional, numerical utility function $U(\cdot)$ such that alternative $A$ is preferred to alternative $B$ if and only if $U(A) > U(B)$. The existence of such a utility representation is assumed by most,
but not all, models of decision making, including Expected Utility Theory (von Neumann & Morgenstern, 1947) and Cumulative Prospect Theory (Tversky & Kahneman, 1992). Despite its centrality in modeling preferential choice, the decision-making literature is internally divided on the question of whether a numerical utility function can well describe the actual choices of individual decision makers. In particular, numerous studies, beginning with Tversky’s (1969) ‘Intransitivity of Preference,’ have questioned whether the preferences of human decision makers satisfy transitivity, which is a necessary condition for the existence of a utility representation as described above - see Mellers and Biagini (1994), Regenwetter, Dana, and Davis-Stober (2011), and Fishburn (1991) for comprehensive reviews of the arguments both for and against transitivity.

Tversky’s (1969) experiment, and many of those that followed (e.g., Birnbaum & Gutierrez, 2007; Ranyard, 1977; 1982; Montgomery, 1977), were designed to elicit intransitive preferences arising from a lexicographic semiorder structure, which is defined as an ordered collection of semiorders. The idea of representing preferences as semiorders was initiated by Luce (1956) and extended to lexicographic semiorders by Tversky (1969, 1972). The core feature of decision models based on semiorders is that “small” differences in attribute values are ignored by the decision maker. A canonical example is that of a decision maker choosing between two cups of coffee: one cup without sugar, the other cup with one microgram of sugar. Since this amount of sugar is below what a human tongue can detect, the decision maker would be indifferent between the two cups of coffee. Similarly, a decision maker would likely be indifferent between two similar goods whose difference in price is a single U.S. penny. The idea of a lexicographic semiorder representation of preference is that, when comparing any two choice alternatives, attribute values are compared sequentially via semiorders until a set of attribute values are reached on which the choice alternatives are distinguishable by a sufficient margin. At that point the process stops and the alternative that is superior based on that attribute is preferred.

Not all lexicographic semiorders are compatible with a utility representation, and not all utility representations are compatible with a lexicographic semiorder, hence these two representations constitute divergent theories of the algebraic structure of preferences. Representing preferences as lexicographic semiorders is intuitively appealing for its apparent simplicity and realism, and provides a model of bounded rational choice that can be characterized by direct axioms on choice behavior (Manzini & Mariotti, 2012). However, as with the utility representation of preferences, the literature remains divided on whether a lexicographic semiorder representation can accurately describe real human choice data. Proponents of the lexicographic semiorder representation tout the ecological rationality of fast and frugal heuristics from which semi-ordered preferences can arise (Gigerenzer and Brighton, 2009), yet recent empirical tests of the most prominent of such models, the Priority Heuristic\(^1\) (Brandstätter, Gigerenzer, & Hertwig, 2006), have found little support for it in the data (e.g., Birnbaum and Gutierrez 2007, Birnbaum 2008, Glöckner and Betsch, 2008).

Why has the literature been unable to reach a consensus on the algebraic structure

---

\(^1\) Depending on the choice stimuli being considered, the Priority Heuristic is either a lexicographic semiorder or a lexicographic interval order.
of preferences, despite decades of research? We argue that four, long-standing, methodo-
logical conventions have hindered progress:

1. Due to the mathematical complexity of higher-order choice structures, studies have
used a binary choice framework that does not explicitly include indifference, even
though indifference is a defining aspect of the lexicographic semiorder structure (see
Regenwetter and Davis-Stober, 2012, for an in-depth discussion of this limitation).

2. Due to a lack of consensus on how to appropriately specify algebraic structures
as stochastic models, different studies have used different stochastic specifications
(e.g., trembling hand, true-and-error, random utility) to test the algebraic struc-
tures. This confounds the test of the algebraic structure with a test of the stochas-
tic framework.

3. Due to the limitations of statistical analyses based solely on goodness-of-fit, pre-
vious analyses have not appropriately penalized models for complexity, potentially
biasing the results in favor the models that are more flexible but not necessarily
more generalizable.

4. Due to the limitations of statistical analyses that are not well-suited for non-nested
model comparison, many previous studies have not directly compared competing
theories to one another. As a result, rather than offering an alternative explanation
to “rationalize” the data, statistically significant violations are attributed to either
irrational behavior or type-1 error.

In this article, we aim to bring some clarity to the issue of preference representation
by addressing all of the limitations listed above, with a reanalysis of existing data as well
as our own new experiment. Using a Bayesian model selection methodology to directly
evaluate competing mixture specifications of ‘ternary choice,’ we find that a majority of
decision makers (about 80%) are best described by a numerical utility representation,
while a substantial minority (about 20%) are best described by a lexicographic semiorder
representation. Very few participants seem to violate both representations, even though
decision makers choosing randomly would do so more than 99.9% of the time due to
the extremely strong restrictions placed on choice data by our parsimonious mixture
framework. Further details on our approach, and how it addresses the limitations listed
above, are given next.

1.1 Mixture models of ternary choice

To address the limitations of binary choice, we test a new class of lexicographic semiorder
mixture models (LSMM) for ‘ternary choice’ data (Davis-Stober, 2010, 2012). The
ternary choice framework extends the standard binary choice framework by allowing
participants to report indifference, instead of forcing them to always report a strict pref-
erence, thereby providing a richer mapping between true preference and choice behavior.
Under a LSMM, at every experimental time point, decision makers (DMs) are required
to make choices consistent with a lexicographic semiorder over the choice alternatives,
with the particular lexicographic semiorder used by the DM allowed to shift over the course of an experiment. In other words, on each choice, the DM is assumed to draw his or her preference from a mixture of lexicographic semiorders. Moving to ternary choice is critical for this model to be testable, because a mixture model on binary choice data would be unable to distinguish between the case of a DM truly being indifferent between two alternatives and a DM having a mixture of opposite, strict preferences. This class of model provides a very general instantiation of the lexicographic semiorder representation. Nevertheless, as we will show, it is testable and extremely parsimonious.

To test whether a numerical utility representation or a lexicographic semiorder representation provides a better description of individual preferences, we will compare the fit of our LSMMs to that of a viable competitor model. However, to overcome the second methodological problem from the list above, the competitor model should use the same stochastic specification. Such a competitor is provided by the weak order mixture model (WOMM) of Regenwetter and Davis-Stober’s (2012). Like the LSMM, the WOMM is also defined for ternary choice, and also allows decision makers to move from one preference another. However, the WOMM requires that DMs always make choices consistent with a strict weak order (i.e., a ranking with ties) over the choice alternatives, rather than a lexicographic semiorder. In the ternary choice framework, a strict weak order representation of preferences is equivalent to a numerical utility representation, so the WOMM provides a general instantiation of the numerical utility representation that includes other utility-based models like Expected Utility Theory and Cumulative Prospect Theory as special cases.

Both the WOMM and LSMM models can be described as a type of distribution-free “random preference model” (e.g., Loomes & Sugden, 1995; Heyer & Niederée, 1992). This perspective allows for a general test of the mathematical structures of interest. By operating at the level of preference relations, we are not assuming any particular functional form that gives rise to them, i.e., we are not engaging in model-fitting. Similarly, by allowing an arbitrary distribution over the preference relations of interest, we are not limiting our results or analyses to any particular choice of mixture distribution or estimation method. Yet, while both WOMM and LSMM are general representations of their corresponding structures, both place extremely strong restrictions on the set of observable data that would be consistent with them.

1.2 Reanalysis of existing data using Bayesian model selection

Regenwetter and Davis-Stober (2012) found strong empirical support for their weak order mixture model by evaluating its goodness-of-fit using data from a study designed to induce intransitive choice. Although a few participants significantly violated the model, they concluded that the transitivity axiom, and hence the existence of a utility representation, was well supported. However, simply evaluating the goodness of fit of the WOMM does not rule out the possibility that a different model, such as a LSMM, could provide a better, more parsimonious description of the data. Therefore, to address the third and fourth methodological limitations from the list above, we use Bayesian model selection to compare the WOMM with our LSMMs in the data from Regenwetter and
In particular, we compute the Bayes factor for each model using the order-constrained inference method of Klugkist & Hoijtink (2007). The Bayes factor (Kass and Raftery, 1995), as a model selection measure, properly accounts for model complexity so as to avoid preferring overly flexible models (over-fitting), unlike measures that assess only goodness-of-fit such as the proportion of correctly predicted choices. In this way, the Bayes factor selects the model that is the most generalizable (Myung, 2000).

In our reanalysis, we find the lexicographic semiorder representation to be superior for approximately 20% of participants in Regenwetter and Davis-Stober’s (2012) experiment, including all but one of the participants who were classified as ‘intransitive’ in the original analysis. Thus, we conclude that the lexicographic semiorder representation does indeed provide an empirically valid, boundedly rational explanation of the choices of participants who violated the more restrictive (but still bounded) rationality constraints of the utility representation. We find the utility representation (WOMM) to be superior for the remaining 80% of participants, which implies that the transitivity axiom is still well supported for many, but not all, participants in this study. The within-participant results were largely consistent across experimental conditions, suggesting that the individual differences are stable.

1.3 A new empirical test of individual differences, stimulus effects, time-pressure effects, and task-format effects

Prior research has suggested that the conditions of the choice task can affect decision makers’ choice strategies, and possibly influence the algebraic structures of their preferences. For instance, Rieskamp & Hoffrage (1999, 2008) showed that DMs are more likely to apply lexicographic strategies when the time to make a decision is greatly limited. Similarly, Brandstätter (2011) showed that the display format of gambles can influence the order in which DMs process information, encouraging comparisons within alternatives (as in a utility calculation) or within attributes (as in a lexicographic heuristic). Therefore, to investigate the stability of preference structures across experimental conditions, and to further test for individual differences, we report the results of a partial replication of the Regenwetter and Davis-Stober (2012) study with the following additional experimental manipulations. We introduce: 1) a time pressure condition, where the participants are given a very limited amount of time to make their choices, and 2) a display format condition, where the choice alternatives are displayed in such a way so as to encourage within-attribute comparisons, as in a lexicographic semiorder structure. In addition to the classic Tversky (1969) gambles, we also introduce new gamble stimuli with larger variances in payout values. Our study follows a within-participants design, with all participants completing all experimental conditions. This type of design prevents aggregation artifacts caused by averaging disparate preference states across individuals, see Luce (2000).

Our findings from the experiment confirm our conclusions from the re-analysis in that a majority of participants are best described by WOMM, with a substantial minority of participants best described by LSMM. We find a weak, display-format main effect, and no main effect for the time pressure manipulation. These results further suggest that the
individual differences we found in the reanalysis are robust with respect to manipulations of both time-pressure and display format.

The rest of the paper is organized as follows. In Section 2, we precisely formalize the WOMM and LSMMs, which instantiate the hypotheses of numerical utility and lexicographic semiorder representations, respectively. Section 3 then describes the statistical methodology that will be used to compare these models. Section 4 gives the results of our reanalysis of the data from Regenwetter et al. (2012), using the models described in Section 2 and the methodology described in Section 3. Section 5 gives the method and results of our new experiment, and Section 6 offers overall conclusions and a general discussion of our results.

2 Model specification

2.1 Weak Order Mixture Model

Let $\mathcal{A}$ be a set of $n$ choice alternatives with $n \geq 3$. The mixture models we consider are defined on ternary choice. In a ternary choice experiment, a DM is presented with two choice alternatives and may express preference for one alternative or the other, or express indifference between the two. Thus, we model DM behavior using ternary choice probabilities, in contrast to binary choice probabilities used in specifications such as WST. Let $\mathcal{P}_{a,b}$ denote the probability of choosing alternative $a$ over $b$. The collection $(\mathcal{P}_{a,b})_{a,b \in \mathcal{A}, a \neq b}$ is called a system of ternary choice probabilities if, and only if,

$$0 \leq \mathcal{P}_{a,b} \leq 1, \quad \forall a, b \in \mathcal{A}, a \neq b,$$

$$\mathcal{P}_{a,b} + \mathcal{P}_{b,a} \leq 1, \quad \forall a, b \in \mathcal{A}, a \neq b.$$

In words, for any pair of alternatives, $a, b \in \mathcal{A}$, $\mathcal{P}_{a,b}$ is the probability of the DM choosing $a$ over $b$, $\mathcal{P}_{b,a}$ is the probability of choosing $b$ over $a$, and $1 - \mathcal{P}_{a,b} - \mathcal{P}_{b,a}$ is the probability of the DM expressing indifference between $a$ and $b$.

The strict weak order mixture model (WOMM) of Regenwetter and Davis-Stober (2012) is a model of probabilistic choice defined over strict weak orders, i.e., asymmetric and negatively transitive binary relations. Under WOMM, each choice made by a DM must be consistent with a strict weak order, but over repeated paired comparisons that strict weak order may fluctuate. This fluctuation in strict weak order may represent intrinsic uncertainty among choice alternatives or simply a “change of mind.” A key feature of WOMM is that no additional structure is placed on the probability distribution over the set of strict weak orders, i.e., WOMM allows any probability distribution over strict weak orders. A system of ternary choice probabilities satisfies WOMM if there exists a probability distribution on $\omega_\mathcal{A}$,

$$\text{Prob} : \omega_\mathcal{A} \rightarrow [0, 1]$$

$$\triangleright \mapsto \mathcal{P}_{\triangleright}.$$
that assigns probability $P_{\succ}$ to any strict weak order $\succ$, such that $\forall a, b \in A, a \neq b$,

$$P_{a,b} = \sum_{\succ \in WO_A, a \succ b} P_{\succ},$$

where $WO_A$ is the set of all strict weak orders on $A$. In other words, the probability that the DM chooses $a$, when offered $a$ versus $b$, is the total probability of all those strict weak orders $\succ$ in which $a$ is strictly preferred to $b$. In the terminology of Loomes and Sugden (1995), WOMM is a random preference model whose core theory is strict weak orders. This model can be equivalently characterized as a “distribution-free random utility model” (Heyer & Niederée, 1992; Regenwetter, 1996; Regenwetter & Marley, 2001).

WOMM can be equivalently characterized as a set of linear inequality constraints on a DM’s system of ternary choice probabilities$^2$. In general, a complete enumeration of these linear inequality constraints is not known for arbitrary $n$. Regenwetter and Davis-Stober (2012) used numerical methods to obtain the necessary system of linear inequality constraints for $n = 5$, obtaining 75,834 non-redundant, linear inequalities. To our knowledge, there is not a complete linear inequality description of WOMM for $n$ larger than 5.

### 2.2 Interpretation of WOMM

Since WOMM allows an arbitrary probability distribution over all strict weak orders, a violation of WOMM, in turn, implies a violation of any utility theory that places a unidimensional scale over a set of choice alternatives, including expected utility theory and prospect theory. Note that WOMM is defined at the level of ternary choice probabilities and does not require the specification of any particular utility function. In this way, WOMM provides a direct test of any model consistent with strict weak orders and does not confound this test with the goodness-of-fit of any particular choice of utility function. Should a DM’s set of ternary choice probabilities satisfy the model, then this too is informative as WOMM is a very restrictive model. For $n = 5$, the number of allowable preference states under WOMM is equivalent to the number of weak orders over $n$ alternatives, which is equal to 541, a fraction of the total number of possible ternary preferences over 5 alternatives, which is equal to 59,049. Regenwetter and Davis-Stober (2012) also examined the proportion of ternary choice probabilities that conform to WOMM for $n = 5$, and, using Monte Carlo methods, found this proportion to be slightly less than .0005.

### 2.3 Lexicographic Semiorder Mixture Models (LSMM1 and LSMM2)

Consider the case of a DM that does not satisfy WOMM. How can we model their preferences in a way that accommodates probabilistic choice? In this section, we present a class of mixture model derived from lexicographic semiorders. This class of mixture model $^2$

---

$^2$From a geometric perspective, WOMM is equivalent to the weak order polytope on $n$-many choice alternatives. The facet structure of this polytope has been studied by Fiorini (2001) and Fiorini and Fishburn (2004). See Ziegler (1995) for an introduction to convex polytopes. 
is defined on ternary choice probabilities and is compatible with intransitive preference. We begin with some preliminary definitions.

A binary relation, $S$, is a strict semiorder if, and only if, there exists a real-valued function $g$, defined on $A$, and a non-negative constant $q$ such that, $\forall a, b \in A$,

$$aSb \iff g(a) > g(b) + q.$$  \hspace{1cm} (1)

Using predecessor and successor sets we can associate any strict semiorder with a weak ordering of the elements in $A$. Define the trace of a semiorder, $S$, as the relation $T$ such that

$$aTb \iff [bSc, \forall c \in A] \text{ and } [dSa \Rightarrow dSb, \forall d \in A].$$

It is routine to show that $T$ defines a weak ordering on $A$.

In general, a lexicographic semiorder is defined as a binary relation on $A$ characterized by an ordered collection of semiorders (Pirlot & Vincke, 1997). In this article, we consider a special case of lexicographic semiorders. A simple lexicographic semiorder (Davis-Stober, 2012) is a relation $P$ on $A$ such that there exist two semiorders $S_1$ and $S_2$ on $A$, that satisfy, $\forall a, b \in A$:

(i) the trace of $S_1$ is the reverse of the trace of $S_2$.

(ii) $aPb \iff aS_1b$, or $[aS_2b \text{ and } \neg[aS_1b \text{ or } bS_1a]]$.

Let $T_1$ be the trace of $S_1$ and let $T_2$ be the trace of $S_2$. $T_1$ is the reverse of $T_2$ if, for any $a, b \in A$, if $(a, b) \in T_1$ then $(b, a) \in T_2$.

For modeling purposes, we associate the semiorders $S_1$ and $S_2$ with the attributes of the choice alternatives under consideration. As described in Davis-Stober (2010; 2012), simple lexicographic semiorders have been used across many studies to model decision environments where choice alternatives have two attributes that “trade-off” with one another. This is often the case for the gamble stimuli used in risky choice experiments. For mixed gambles with one gain and one loss, the probability of winning will tend to increase as the payoff amount decreases, and vice versa. Hence, the trace of a semiorder associated with the “probability of a gain” attribute will be the reverse of the trace of the semiorder associated with the “payoff amount” attribute.

To see how preferences based on a lexicographic semiorder can be intransitive, consider the following example of a decision maker (DM) who is interested in purchasing a new camera. Suppose that cameras have two relevant attributes, price and quality, and that this DM is considering the three cameras whose prices and qualities are shown in Table 1. Camera A costs $200 and is of ‘low’ quality, camera B costs $250 and is of ‘medium’ quality, and camera C costs $300 and is of ‘high’ quality. Suppose further that this DM is looking for a bargain, so that price takes precedence over quality. Then the price attribute defines $S_1$ and the quality attribute defines $S_2$ in the simple lexicographic semiorder definition, and in comparing any two cameras, the DM would first look at the prices, and then only compare them by quality if the two cameras were sufficiently close in price. This consumer could have intransitive preferences, depending upon the semiorders $S_1$ and $S_2$. Suppose this DM considers differences in price of less than $60
to be irrelevant to his decision. Then this DM would prefer $B$ to $A$, since the price difference is less than $50$ but camera $B$ is of higher quality, and would prefer $C$ to $B$, since the price difference is again less than $50$ but camera $C$ is of higher quality, yet would also prefer $A$ to $C$, since $A$ is more than $60$ cheaper than $C$.

Table 1: Illustration of an intransitive cycle of preference implied by a lexicographic semiorder. $A$ is preferred to $C$ by virtue of the price difference exceeding $60$, but $B$ is preferred to $A$ and $C$ is preferred to $B$ by virtue of higher quality since the price differences are less than $60$ in those pairs.

<table>
<thead>
<tr>
<th>Camera</th>
<th>Price</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$200$</td>
<td>Low</td>
</tr>
<tr>
<td>$B$</td>
<td>$250$</td>
<td>Medium</td>
</tr>
<tr>
<td>$C$</td>
<td>$300$</td>
<td>High</td>
</tr>
</tbody>
</table>

To create viable competitor models for WOMM, we will consider a particular class of simple lexicographic semiorders. Let $\mathcal{SLS}_A$ denote the set of all simple lexicographic semiorders such that the trace of $S_1$ is compatible with a fixed linear ordering on $A$. In general, this fixed linear ordering on $A$ is arbitrary, but we will consider the “natural” linear orderings defined by the attributes of the choice alternatives. For example, in risky choice we could linearly order a set of gambles strictly by the probability of a gain. Under this linear ordering, $\mathcal{SLS}_A$ would become the set of all simple lexicographic semiorders in which the probability of a gain ($S_1$) is considered before the payoff attribute ($S_2$). Conversely, we could consider the set of all simple lexicographic semiorders in which the payoff attribute is considered before the probability of a gain.

By considering a mixture model over the set $\mathcal{SLS}_A$ with an appropriately chosen fixed linear ordering, we define a mixture model over all simple lexicographic semiorders in which a specific attribute is always considered first in the lexicographic ordering. Said differently, each lexicographic semiorder mixture model we consider does not allow a DM to switch the order that he or she examines the attributes. More formally, a collection of ternary choice probabilities satisfies a lexicographic semiorder mixture model (LSMM) if, and only if, there exists a probability distribution on $\mathcal{SLS}_A$,

$$\text{Prob} : \mathcal{SLS}_A \rightarrow [0,1]$$

$$P \mapsto P_{\mathcal{SLS}_A},$$

such that $\forall a, b \in A, a \neq b$,

$$P_{a,b} = \sum_{P \in \mathcal{SLS}_A : (a,b) \in P} P_{\mathcal{SLS}_A}.$$  

Said another way, LSMM states that the probability of a DM choosing choice alternative $a$ over choice alternative $b$ is the total probability of all simple lexicographic semiorders in $\mathcal{SLS}_A$ such that $a$ is strictly preferred to $b$. Similar to the interpretation of WOMM, the
particular simple lexicographic semiorder used by the DM is allowed to fluctuate across comparisons, for example due to a “change of mind,” fatigue, or intrinsic uncertainty among choice alternatives. As in WOMM, we place no restrictions on the probability distribution over the elements of $\mathcal{SLS}_A$.

In our empirical analysis, $A$ is comprised of binary gambles and therefore we consider two LSMMs: the mixture over all simple lexicographic semiorders in which the probability of a gain is considered before payoff values (LSMM1), and the mixture over all simple lexicographic semiorders in which the payoff attribute is considered before the probability of a gain (LSMM2). LSMM1 and LSMM2 have exactly the same number of preference states and are described by the same class of linear inequality constraints, i.e., one can obtain LSMM1 from LSMM2 (or any LSMM) by simply re-labeling the corresponding ternary choice probabilities according the fixed ordering on $A$. See Davis-Stober (2012) for a complete discussion.

In contrast to WOMM, a complete, minimal set of non-redundant linear inequalities that completely characterize this class of model is known for arbitrary $n$ (Davis-Stober, 2012). Assume that the elements of $A$ are ordered in a strictly increasing fashion according to a fixed linear ordering of $A$. A collection of ternary choice probabilities satisfy LSMM if, and only if, the following seven families of inequalities are satisfied:

1. $P_{i,j} - P_{i,j+1} \leq 0$, $\forall (i,j) \in \{1,2,\ldots,n\} : i < j < n$, (2)
2. $P_{i+1,j} - P_{i,j} \leq 0$, $\forall (i,j) \in \{1,2,\ldots,n\} : i + 1 < j$, (3)
3. $P_{j,i} + P_{i,j} - P_{j+1,i} - P_{i,j+1} \leq 0$, $\forall (i,j) \in \{1,2,\ldots,n\} : i < j < n$, (4)
4. $P_{i+1,j} + P_{j,i+1} - P_{i,j} - P_{j,i} \leq 0$, $\forall (i,j) \in \{1,2,\ldots,n\} : i + 1 < j$, (5)
5. $-P_{i,i+1} \leq 0$, $\forall i \in \{1,2,\ldots,n-1\}$, (6)
6. $-P_{i,j} \leq 0$, $\forall (i,j) \in \{1,2,\ldots,n\} : i > j$, (7)
7. $P_{1,n} + P_{n,1} \leq 1$. (8)

The total number of non-redundant linear inequalities that characterize a LSMM is equal to $\frac{5n^2-11n+8}{2}$ (Davis-Stober, 2012). For the experiments we consider in our analysis, $n = 5$, yielding 39 non-redundant linear inequalities.

2.4 Interpretation of LSMM1 and LSMM2

This class of model generalizes previous approaches to evaluating the descriptive accuracy of lexicographic semiorder models. Birnbaum and LaCroix (2008) investigated the descriptive accuracy of a particular type of lexicographic semiorder mixture model. Their mixture model allowed six preference states and was constructed as a test of the priority heuristic theory (Brandstätter, Gigerenzer, & Hertwig, 2006), which, depending on the choice stimuli, is either a lexicographic semiorder or lexicographic interval order over four attributes. In their study, which utilized a two-alternative, forced-choice design, they found that this mixture model was not well supported by the data, - see also

\footnote{A relation, $R_1$, defined on $A$, is compatible with another relation, $R_2$, also defined on $A$, if $\forall a,b \in A$, $(a,b) \in R_1 \Rightarrow (a,b) \in R_2.$}
Birnbaum and Gutierrez (2007). Similarly, Regenwetter, Dana, Davis-Stober, and Guo (2011) also found little empirical support for a lexicographic semiorder mixture model defined on binary choice probabilities. Our extension to ternary choice is important, as indifference is a critical component of the semiorder structure. For example, the lexicographic semiorder mixture model used by Regenwetter et al. is unable to distinguish between the case of a DM truly being indifferent between two alternatives and a DM having a mixture of opposite strict preferences.

While LSMMs and WOMM are derived under very different assumptions, they are comparable in terms of the number of allowable preference states. Davis-Stober (2010) proved that the number of elements in $\mathcal{SLS}_A$ is a simple function of the Catalan numbers, 

$$|\mathcal{SLS}_A| = C_n C_{n+2} - C_{n+1}^2,$$

where $C_n = \frac{1}{n+1} \binom{2n}{n}$. In all of the experiments we consider in this paper, $n = 5$, thus there are 594 possible preference states under a LSMM. While the number of preference states are comparable to WOMM, 594 as compared to 541, the LSMMs place even tighter constraints on the set of permissible ternary choice probabilities. Davis-Stober (2012), using Monte Carlo methods, estimated the proportion of permissible ternary choice probabilities for LSMM to be roughly .00002 (for $n = 5$).

3 Statistical Methodology

The choice data we consider is comprised of DM responses to a series of repeated gamble pair presentations. The DM indicates preference/indifference on each presentation trial. By including repetitions of gamble pairs, we are better able to distinguish choice variability from the underlying structural preference of the individual (e.g., Tversky, 1969, see also Regenwetter and Davis-Stober, 2012, for a discussion). Assuming a balanced design, let $N$ be the number of times each individual gamble pair is presented to the DM. Assuming independence between trials, a DM’s choice responses follow a trinomial distribution with the following likelihood,

$$L(P|N) = \prod_{\forall(a,b) \in \mathcal{A}} P_{a,b}^{N_{a,b}} P_{b,a}^{N_{b,a}} (1 - P_{a,b} - P_{b,a})^{N - N_{a,b} - N_{b,a}},$$

where $N = N_{a,b}, \forall a, b \in \mathcal{A}, a \neq b$, and $N_{a,b}$ (respectively $N_{b,a}$) is the number of times choice alternative $a$ was selected over $b$ (respectively $b$ over $a$). The ternary choice probabilities, $P_{a,b}$, are estimated using their corresponding marginal choice proportions, $\frac{N_{a,b}}{N}$. While we state independence in the likelihood definition, it is important to note that the Bayesian methodology we adopt to carry out model comparison only requires the weaker assumption of exchangeability (e.g., Bernardo, 1996).

3.1 Order-constrained Bayesian inference

Each mixture model is defined as a system of linear inequalities on ternary choice probabilities. As such, standard statistical techniques for evaluating these models are not
appropriate due to a violation of necessary likelihood assumptions, i.e., ‘boundary conditions’ (e.g., Silvapulle & Sen, 2005). To provide a proper statistical analysis of these models, we must employ order-constrained statistical inference techniques. This necessity has been discussed at length in the decision literature (Davis-Stober, 2009; Iverson & Falmagne, 1985; Myung, Karabatsos, & Iverson, 2005; Tversky, 1969).

To this end, we employ the order-constrained Bayes factor methodology of Klugkist and Hoijtink (2007). Let $M_1$ be defined as the ‘encompassing model’ formed by placing no a priori restrictions on the ternary choice probabilities. In other words, $M_1$ is a type of reference or “null” model in which DMs are allowed to have any set of ternary choice probabilities. This statistical method takes advantage of the fact that WOMM, LSMM1, and LSMM2 are nested in $M_1$, thus, we need only compute the Bayes factor of each model relative to $M_1$. Let $M_t$ be the mixture model specification being evaluated, either WOMM, LSMM1, or LSMM2. Then the Bayes factor for $M_1$ and $M_t$, denoted $BF_{t1}$, is defined as the ratio of the two marginal likelihoods,

$$BF_{t1} = \frac{p(N|M_t)}{p(N|M_1)} = \frac{\int L(N|\mathcal{P})\pi(\mathcal{P}|M_t)d\mathcal{P}}{\int L(N|\mathcal{P})\pi(\mathcal{P}|M_1)d\mathcal{P}},$$

where $\pi(\mathcal{P}|M_t)$ is the prior distribution of $\mathcal{P}$ under model $M_t$, which is defined to be uniform on the support of $M_t$. The $BF_{t1}$ is defined with respect to the encompassing model and evaluates the strength of evidence, in terms of the likelihood of generating the observed data, of the theoretic model against the encompassing model. For example, $BF_{t1}=10$ means that the theoretic model is 10 times more likely to have generated the data. The Bayes factor between any two mixture models can be constructed by taking the ratio of their respective $BF_{t1}$ values. This Bayes factor methodology provides a measure of empirical evidence for each model while appropriately penalizing for complexity, defined as the volume of the parameter space that each substantive model occupies relative to the encompassing model. As discussed in the previous sections, WOMM and the LSMMs are extraordinarily parsimonious by this measure. We refer interested readers to Klugkist and Hoijtink (2007) for additional details on this statistical method.

4 Re-analysis of Regenwetter and Davis-Stober (2012)

In this section, we re-analyze the Regenwetter and Davis-Stober (2012) data. This experiment followed a very similar design to that of Tversky (1969). Participants were repeatedly presented pairs of binary gambles and asked to select the gamble they preferred. Their experiment deviated from that of Tversky as they allowed the participant to indicate indifference between the presented gambles, i.e., ternary choice as opposed to binary forced-choice. In their experiment, they tested a total of 30 participants across three stimulus sets following a within-participants design, i.e., all participants completed all trials for all stimulus sets. Each stimulus set consisted of five distinct gambles, providing ten distinct gamble pairs per stimulus set. Each gamble pair, in each stimulus set, was presented to each participant a total of 45 times. This sample size provided excellent statistical power for the classical order-constrained method used in the original
Table 2: Regenwetter and Davis-Stober (2012) Experimental Simuli

<table>
<thead>
<tr>
<th>Gamble Set I</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>($26.32, \frac{7}{21}$)</td>
<td>b</td>
<td>($25.00, \frac{8}{21}$)</td>
<td>c</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gamble Set II</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>($31.43, .28$)</td>
<td>b</td>
<td>($27.50, .32$)</td>
<td>c</td>
</tr>
</tbody>
</table>

analysis (Davis-Stober, 2009). We refer readers to Regenwetter and Davis-Stober (2012) for complete details of the experiment.

Regenwetter and Davis-Stober (2012) found a total of four data sets that rejected the weak order mixture model ($p < .05$), out of 90 total data sets (30 participants times 3 stimulus sets). They concluded that the violations of WOMM were well within the range of Type I error and that transitivity was well supported in this sample. It is important to reiterate that the classical methodology they used is not well suited for direct model comparison. Therefore, this analysis could not compare WOMM to a competitor model, nor adequately penalize WOMM for complexity.

We conducted a re-analysis of this data, using the Klugkist and Hoijtink (2007) Bayes factor methodology, to compare WOMM to competitor models compatible with intransitive preference, LSMM1 and LSMM2. We re-analyzed data from two of the three stimulus sets. Sets 1 and 2 were binary gambles over monetary values - see Table 2. Stimulus Set 1 was an updated version of the classic Tversky (1969) gamble set, with all monetary values updated to 2006 dollars. Stimulus Set 2 was similar in structure, with the modification that all gamble pairs had equal expected value, roughly $8.80. These stimuli provide a natural two-attribute structure for simple lexicographic semiorders, assuming individuals prefer winning more money to less. Regenwetter and Davis-Stober’s (2012) stimulus Set 3 were binary gambles involving non-monetary outcomes, such as lunch items, coffee gift cards, and movie rentals - all with roughly the same monetary value. While the WOMM model is well-defined for these experimental stimuli, there is not a unique way to model these stimuli using lexicographic semiorders as they do not have any attributes in common other than probability of winning. Thus, we omitted this set from analysis.

4.1 Results

Tables 3-4 present the results of our re-analysis. The second column of Table 3 displays the best fitting model, WOMM, LSMM1, or LSMM2, according to the Bayes factor analysis for stimulus Set 1. The Bayes factor for the best fitting model compared to the encompassing model is presented in column 3. To provide a measure of distinguishability between these models, in columns 4 and 5 we present the second-best model for that data.
Table 3: This table displays the Bayes Factor (BF) analysis for gamble Set 1 of the Regenwetter and Davis-Stober experiment. The first column displays the participant number. The second and third columns display the best performing model for that participant and the BF for that model compared to the encompassing model. The fourth column displays the second-best performing model. The fifth column is the BF between the best performing model (column 2) and the second-best performing model (column 4) with larger numbers favoring the model in column 2. “†” indicates a BF greater than 3, “∗” indicates a BF greater than 10, and “∗∗” indicates a BF greater than 100. The sixth column reproduces the original classical, WOMM goodness-of-fit analysis of Regenwetter and Davis-Stober (2012). A “√” indicates that the marginal choice proportions satisfied the WOMM; if the marginal choice proportions violated the WOMM then the $G^2$ value is listed along with its respective p-value in parentheses.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Set 1 BF</th>
<th>Next Best (Set 1) BF</th>
<th>Original Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WOMM 13228.27**</td>
<td>LSMM1 837.82**</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>WOMM 1219.91**</td>
<td>LSMM1 214**</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>WOMM 13166.72**</td>
<td>LSMM1 26804.74**</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>WOMM 4417.42**</td>
<td>LSMM1 1.25</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>WOMM 5006.71**</td>
<td>LSMM1 51.65*</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>WOMM 6486.52**</td>
<td>LSMM1 644.16**</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>WOMM 16851.72**</td>
<td>LSMM1 425.04**</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>LSMM1 441.24**</td>
<td>WOMM 5.59†</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>WOMM 582.55**</td>
<td>LSMM1 5534.46**</td>
<td>0.52 (.82)</td>
</tr>
<tr>
<td>10</td>
<td>WOMM 2105.66**</td>
<td>LSMM1 1276.89**</td>
<td>2.82 (.47)</td>
</tr>
<tr>
<td>11</td>
<td>WOMM 936.04**</td>
<td>LSMM1 182.73**</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>WOMM 9988.93**</td>
<td>LSMM1 515.75**</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>WOMM 349.13**</td>
<td>LSMM1 104.74**</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>LSMM1 174361.50**</td>
<td>WOMM 12302.77**</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>LSMM1 174196.9**</td>
<td>WOMM 12057.07**</td>
<td>✓</td>
</tr>
<tr>
<td>16</td>
<td>WOMM 15737.00**</td>
<td>LSMM1 3617.12**</td>
<td>✓</td>
</tr>
<tr>
<td>17</td>
<td>WOMM 7566.21**</td>
<td>LSMM1 112.73**</td>
<td>✓</td>
</tr>
<tr>
<td>18</td>
<td>WOMM 16817.18**</td>
<td>LSMM1 15976.96**</td>
<td>✓</td>
</tr>
<tr>
<td>19</td>
<td>WOMM 17023.17**</td>
<td>LSMM1 15976.96**</td>
<td>✓</td>
</tr>
<tr>
<td>20</td>
<td>WOMM 12821.84**</td>
<td>WOMM 183.81**</td>
<td>✓</td>
</tr>
<tr>
<td>21</td>
<td>WOMM 13879**</td>
<td>LSMM1 88.20*</td>
<td>5.01 (.10)</td>
</tr>
<tr>
<td>22</td>
<td>LSMM1 527.21**</td>
<td>WOMM 998.42**</td>
<td>✓</td>
</tr>
<tr>
<td>23</td>
<td>WOMM 5493.51**</td>
<td>LSMM1 575.63**</td>
<td>✓</td>
</tr>
<tr>
<td>24</td>
<td>WOMM 31801.25**</td>
<td>WOMM 5.65†</td>
<td>✓</td>
</tr>
<tr>
<td>25</td>
<td>WOMM 315.53**</td>
<td>LSMM1 93.68*</td>
<td>1.93 (.48)</td>
</tr>
<tr>
<td>26</td>
<td>WOMM 8580.10**</td>
<td>LSMM2 81514.21**</td>
<td>✓</td>
</tr>
<tr>
<td>27</td>
<td>WOMM 12232.93**</td>
<td>LSMM1 863.00**</td>
<td>✓</td>
</tr>
<tr>
<td>28</td>
<td>LSMM1 6.74†</td>
<td>WOMM 2.02</td>
<td>6.04 (.26)</td>
</tr>
<tr>
<td>29</td>
<td>LSMM1 2446.60**</td>
<td>WOMM 5.54†</td>
<td>1.03 (.54)</td>
</tr>
</tbody>
</table>

14
Table 4: This table displays the Bayes Factor (BF) analysis for gamble Set 2 of the Regenwetter and Davis-Stober experiment. The first column displays the participant number. The second and third columns display the best performing model for that participant and the BF for that model compared to the encompassing model. The fourth column displays the second-best performing model. The fifth column is the BF between the best performing model (column 2) and the second-best performing model (column 4) with larger numbers favoring the model in column 2. “†” indicates a BF greater than 3, “∗” indicates a BF greater than 10, and “∗∗” indicates a BF greater than 100. The sixth column reproduces the original classical, WOMM goodness-of-fit analysis of Regenwetter and Davis-Stober (2012). A “√” indicates that the marginal choice proportions satisfied the WOMM; if the marginal choice proportions violated the WOMM then the $G^2$ value is listed along with its respective p-value in parentheses.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Set 2</th>
<th>BF</th>
<th>Next Best (Set 2)</th>
<th>BF</th>
<th>Original Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WOMM</td>
<td>16419.45**</td>
<td>LSMM1</td>
<td>306.87**</td>
<td>√</td>
</tr>
<tr>
<td>2</td>
<td>WOMM</td>
<td>16418.45**</td>
<td>NA</td>
<td>NA</td>
<td>√</td>
</tr>
<tr>
<td>3</td>
<td>WOMM</td>
<td>3044.74**</td>
<td>LSMM2</td>
<td>86778.51**</td>
<td>1.11 (.77)</td>
</tr>
<tr>
<td>4</td>
<td>WOMM</td>
<td>16679.89**</td>
<td>LSMM1</td>
<td>176.14**</td>
<td>√</td>
</tr>
<tr>
<td>5</td>
<td>WOMM</td>
<td>2356.72**</td>
<td>LSMM2</td>
<td>67169.23**</td>
<td>.17 (.89)</td>
</tr>
<tr>
<td>6</td>
<td>WOMM</td>
<td>10765.23**</td>
<td>LSMM1</td>
<td>514.80**</td>
<td>√</td>
</tr>
<tr>
<td>7</td>
<td>WOMM</td>
<td>16018.44**</td>
<td>NA</td>
<td>NA</td>
<td>√</td>
</tr>
<tr>
<td>8</td>
<td>WOMM</td>
<td>2150.36**</td>
<td>LSMM1</td>
<td>40.16*</td>
<td>√</td>
</tr>
<tr>
<td>9</td>
<td>LSMM1</td>
<td>149.22**</td>
<td>WOMM</td>
<td>58.47*</td>
<td>5.95 (.18)</td>
</tr>
<tr>
<td>10</td>
<td>WOMM</td>
<td>12775.04**</td>
<td>LSMM1</td>
<td>471.03**</td>
<td>√</td>
</tr>
<tr>
<td>11</td>
<td>WOMM</td>
<td>1427.35**</td>
<td>LSMM1</td>
<td>42.55**</td>
<td>√</td>
</tr>
<tr>
<td>12</td>
<td>WOMM</td>
<td>7190.93**</td>
<td>LSMM1</td>
<td>737.23**</td>
<td>√</td>
</tr>
<tr>
<td>13</td>
<td>WOMM</td>
<td>348.64**</td>
<td>LSMM1</td>
<td>83.50**</td>
<td>√</td>
</tr>
<tr>
<td>14</td>
<td>LSMM1</td>
<td>4160.78**</td>
<td>WOMM</td>
<td>8.09†</td>
<td>1.22 (.38)</td>
</tr>
<tr>
<td>15</td>
<td>LSMM1</td>
<td>4144.75**</td>
<td>WOMM</td>
<td>8.06**</td>
<td>.80 (.40)</td>
</tr>
<tr>
<td>16</td>
<td>WOMM</td>
<td>16535.58**</td>
<td>LSMM1</td>
<td>10960.06**</td>
<td>√</td>
</tr>
<tr>
<td>17</td>
<td>WOMM</td>
<td>7726.53**</td>
<td>NA</td>
<td>NA</td>
<td>√</td>
</tr>
<tr>
<td>18</td>
<td>WOMM</td>
<td>8712.31**</td>
<td>NA</td>
<td>NA</td>
<td>√</td>
</tr>
<tr>
<td>19</td>
<td>WOMM</td>
<td>10151.24**</td>
<td>LSMM1</td>
<td>1954.88**</td>
<td>√</td>
</tr>
<tr>
<td>20</td>
<td>WOMM</td>
<td>10556.35**</td>
<td>LSMM2</td>
<td>300868.10**</td>
<td>√</td>
</tr>
<tr>
<td>21</td>
<td>WOMM</td>
<td>10402.16**</td>
<td>LSMM1</td>
<td>91.90*</td>
<td>√</td>
</tr>
<tr>
<td>22</td>
<td>WOMM</td>
<td>.05</td>
<td>NA</td>
<td>NA</td>
<td>14.90 (&lt; .01)</td>
</tr>
<tr>
<td>23</td>
<td>WOMM</td>
<td>6659.05**</td>
<td>LSMM1</td>
<td>144.55**</td>
<td>.04 (.89)</td>
</tr>
<tr>
<td>24</td>
<td>WOMM</td>
<td>7331.23**</td>
<td>LSMM1</td>
<td>257.64**</td>
<td>√</td>
</tr>
<tr>
<td>25</td>
<td>LSMM1</td>
<td>76868.89**</td>
<td>WOMM</td>
<td>13.09*</td>
<td>√</td>
</tr>
<tr>
<td>26</td>
<td>WOMM</td>
<td>388.38*</td>
<td>LSMM1</td>
<td>5534.60**</td>
<td>1.30 (.75)</td>
</tr>
<tr>
<td>27</td>
<td>WOMM</td>
<td>7542.09**</td>
<td>LSMM2</td>
<td>42991.63**</td>
<td>√</td>
</tr>
<tr>
<td>28</td>
<td>WOMM</td>
<td>14591.03**</td>
<td>LSMM1</td>
<td>823.49**</td>
<td>√</td>
</tr>
<tr>
<td>29</td>
<td>LSMM1</td>
<td>26.17*</td>
<td>WOMM</td>
<td>3450.40**</td>
<td>14.81 (.01)</td>
</tr>
<tr>
<td>30</td>
<td>WOMM</td>
<td>2151.15**</td>
<td>LSMM1</td>
<td>38.83*</td>
<td>.04 (.88)</td>
</tr>
</tbody>
</table>
set with accompanying Bayes factors between the best and second-best models. Column 6 displays the results of the original Regenwetter and Davis-Stober (2012) goodness-of-fit analysis of WOMM. This column lists the $G^2$ values they obtained, the larger the value the worse the fit to the WOMM; p-values are given in parentheses and a “√” indicates that the marginal choice proportions were consistent with WOMM. We repeat this information for Set 2 in Table 4.

As seen in Tables 3-4, our results largely agree with the findings of Regenwetter and Davis-Stober with some exceptions. We found that WOMM was the best model for 48 out of the 60 data sets (30 participants over two stimulus sets) with Bayes factors much larger than 100, indicating decisive evidence (Jeffreys, 1998). In other words, a utility representation was supported for a majority of data sets. We found that two of the three participants whose choices significantly violated WOMM from the original analysis were well explained by LSMM1 (Participant 14, Set 1, and Participant 29, Set 2), with the sole exception being Participant 22 under gamble Set 2, for whom the most preferred model according to the Bayes factor was the encompassing model (the Bayes factor for WOMM, the best performing mixture model, against the encompassing was .05). In general, many of the participants who were best fit by LSMM1 were participants whose choice proportions had violated the weak order mixture model in at least one condition, but not necessarily significantly so (at $\alpha = .05$), in the original analysis. The lexicographic semiorder mixture model formed by considering payoff values before the probability of a gain (LSMM2) was decisively worse than at least one competitor model for all data sets. Similar to the conclusions of Tversky (1969), participants employing a lexicographic semiorder strategy tended to examine probability of a gain before payoff values.

5 Experiment

In our re-analysis of Regenwetter and Davis-Stober (2012), we found that while WOMM was the most preferred model for the majority of DMs, LSMM1 provided a superior description of behavior for several DMs, including those that did not fit WOMM in the original analysis. To investigate the generality and robustness of this result, we carried out a new study in which DMs were repeatedly presented pairs of gambles and asked to select their preferred gamble within a ternary choice framework. Our new experiment followed a fully crossed 2 by 2 by 2 design examining three experimental manipulations: ‘time pressure,’ ‘stimulus display format,’ and two stimulus sets. All three experimental manipulations were designed to distinguish empirically among decision theories consistent with weak orders and those based upon lexicographic semiorders.

The purpose of these experimental manipulations is to investigate to the extent to which these mathematical representations, utility-based (WOMM) or lexicographic-based (LSMMs), are “hard-wired” properties of decision making. In other words, for the same set of simuli, can we manipulate experimental conditions in such a way that DMs will adjust their underlying preference structures, in the sense of Bradstätter (2011), or, are preferences robust to such manipulations, suggesting a more ‘automatic’ process, in the
sense of Glöckner and Herbold (2011)?

5.1 Method

We recruited sixty-five students from the University of Missouri to participate; only sixty completed all sessions. We omitted from the analysis the five participants who did not complete all sessions. The experiment was divided into two sessions, with each session lasting approximately one hour. The two sessions were separated by at least one full day. Each session consisted of 480 repeated presentations of the gamble pairs from stimulus set 1 and stimulus set 2 (20 distinct gamble pairs, 12 repetitions of each gamble pair, over two combinations of time pressure and display format manipulations, thus \(20 \times 12 \times 2 = 480\) trials). Each participant completed two sessions, totalling 960 trials.

Participants indicated preference for each ternary choice trial via key presses on a keyboard. Presentation of the trials were fully randomized and counter-balanced (each gamble from each set appeared on the left-hand, respectively right-hand side, an equal number of times). Gamble pairs from the two stimulus sets were randomly inter-mixed to serve as distractors for one another. Participants were randomly assigned to the order the conditions were presented in. As an incentive, all gambles chosen by the participant were recorded on the computer and at the end of each session, two such gambles were randomly selected (one from each of the two conditions) and the participant was allowed to play these gambles for real money. In addition, each participant was paid $10 for participation per session.

5.2 Time Pressure Condition

Prior work has demonstrated that the amount of time a DM has to make a decision affects both information search (Ben-Zur & Breznitz, 1981; Böckenholt & Kroeger, 1993; Payne, Bettman, & Johnson, 1988) and strategy selection (Rieskamp & Hoffrage, 1999; 2008). In particular, Rieskamp and Hoffrage (1999) found that by directly limiting the amount of time to make a decision, participants were more likely to use non-compensatory strategies that are lexicographic in structure. Rieskamp and Hoffrage (2008) later generalized this finding to time pressure involving opportunity costs.

To investigate the effect of time pressure on transitive preference, we added a direct time pressure condition. Under this condition, participants were given only four seconds to indicate their preference among the gambles for each presentation trial. If they did not respond within four seconds, the computer screen flashed a message indicating that they ran out of time and their “chosen” gamble for the purposes of playing for actual payment at the end of the session was randomly determined - for data analysis purposes, these trials were dropped. The average response time for the un-timed condition across participants was 2.010 seconds. The average response time for the timed condition across participants was 1.681 seconds. The average rate in which participants failed to respond within the allotted time across participants was only 1.10%.

If preference structure is robust to experimental manipulations, we would expect similar results across the two time pressure conditions, at both the group and individual
level. On the other hand, should preference structure be dependent on this type of manipulation, prior empirical work suggests DMs would be more likely to utilize a LSMM model for the time pressure condition.

### 5.3 Display Format Condition

There are many studies demonstrating that strategy selection can be influenced by how information is organized and presented to the DM (e.g., Kleinmuntz & Schkade, 1993; Russo, 1977). In a well-cited study, Johnson, Payne and Bettman (1988) demonstrated that the frequency of preference reversals could be influenced by the format in which probabilities were displayed, with formats requiring greater cognitive effort to process eliciting more frequent preference reversals. See also the work on “juxtaposition effects” within the context of regret theory (Harless, 1992; Starmer & Sugden, 1993).

The lexicographic semiorder structure is built upon ‘attribute-wise’ examination of the choice stimuli. In other words, under this structure a DM sequentially compares the gambles one attribute at a time - e.g., the DM examines the probability of a gain and, if the difference between these values is not large, then examines payoffs. It is reasonable to ask if the display format of the gambles could influence choice in this context. Similar to the organizational properties described by Kleinmuntz and Schkade (1993), we developed an experimental condition in which the gamble properties were more readily compared “attribute-wise”. To this end, we constructed two display formats for gamble Sets 1 and 2. In the ‘Pie’ condition, we displayed the gambles according to the common ‘pie’ format used in Tversky (1969), Regenwetter et al., (2011), Regenwetter and Davis-Stober (2012), among others (Montgomery, 1977; Birnbaum, & LaCroix, 2008), see the left-hand side of Figure 1. Under the “Bar Graph” condition, we displayed the gambles so as to encourage attribute-wise comparison. The gambles under this condition are grouped according to attribute, with the probability of a gain for each gamble displayed as a bar chart in the top half of the screen, and the payoff amount information for each gamble displayed in the bottom half of the screen, - see the right-hand side of Figure 1.

Similar to the case of time pressure manipulation, if preference structure is a stable property of decision making at the individual level we would expect similar classification rates across the two display format conditions with few individual-level classification switches. Should preference structure not be robust to this manipulation, we would expect the “Bar Graph” condition to result in a greater likelihood for DMs to utilize either LSMM1 or LSMM2 as the attribute information for the gambles is conveniently grouped for attribute-wise comparison. Additionally, we would expect an increase in the usage of LSMM2 under the “Bar Graph” condition, given that probabilities and payoffs are displayed in an identical fashion, as opposed to the “Pie” format, in which probabilities are displayed more prominently than payoff values, potentially leading DMs to favor examining probabilities before payoffs as in LSMM1.
5.4 Stimulus Manipulation

For this study, we used two stimulus sets of five gambles each. These gambles are displayed in Table 5. Similar to Regenwetter and Davis-Stober (2012), our stimulus set 1 is based on the Tversky (1969) gambles with monetary values updated to 2009 dollar values via the Consumer Price Index. Our stimulus set 2 is a new stimulus set with probabilities of a gain identical to those in Set 1 but with payoffs modified to have larger variances. In Tversky (1969), and gamble Set 1, the expected value of the binary gambles increases as the probability of a gain increases. In our gamble Set 2, the opposite is true; as probability of a gain increases, the expected value of the gambles decreases. In this
way, a DM choosing by expected value would choose gambles with larger payouts - see Table 4. To the extent that different stimuli could alter a DM’s underlying preference structure, this suggests an increase in the use of LSMM2.

5.5 Individual Differences

Collapsing across all stimulus sets, conditions and participants, there were a total of 480 data sets to analyze (60 participants, two stimulus sets, two display conditions, two timed conditions, 60*2*2*2 = 480). For each of the 480 data sets, we simultaneously evaluated the WOMM, LSMM1, LSMM2, and encompassing model using the Bayes factor methodology of Klugkist and Hoijtink (2007). To simplify presentation, we only report a model as the preferred model if it scored the largest Bayes factor among all substantive models and scored a Bayes factor of at least $10^5 = 3.16$ (Jeffreys, 1998) against the encompassing model. Overall, WOMM was the preferred model for 393 out of all 480 data sets (approximately 82%). LSMM1 was the preferred model for 38 out of 480 analyses (approximately 8%), and LSMM2 was the preferred model for 13 out of 480 analyses (approximately 3%). At the participant level, 17 out of 60 participants (.28%) were best fit by one of the lexicographic semiorder mixture models for at least one condition. This is very similar to our re-analysis of the Regenwetter and Davis-Stober (2012) data, finding that 8 out of 30 participants (.27%) were best fit by an LSMM for at least one condition. Our analysis yielded relatively large Bayes factors. The average Bayes factor for WOMM when it was the preferred model was 2645.23; the average Bayes factor for LSMM1 and LSMM2 when classified as the preferred model was 4957.19.

Table 6: The table above/below presents the overall classification rates for WOMM, LSMM1 and LSMM2 across all conditions. The table shows how often (out of 480 data sets) each model was the preferred model.

<table>
<thead>
<tr>
<th>Overall Results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMM</td>
<td>393</td>
</tr>
<tr>
<td>LSMM1</td>
<td>38</td>
</tr>
<tr>
<td>LSMM2</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 7: The table above/below presents the marginal classification rates for WOMM, LSMM1 and LSMM2 across display and stimulus conditions. The table shows how often (out of 240 for each timing condition) each model was the preferred model.

<table>
<thead>
<tr>
<th>Time Condition</th>
<th>Timed</th>
<th>Un-timed</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMM</td>
<td>199</td>
<td>194</td>
</tr>
<tr>
<td>LSMM1</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>LSMM2</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

5.6 Time Pressure Results

Averaging across stimulus sets and display format conditions, the marginal classification rates under the time pressure condition were: WOMM preferred for 199 data sets (out of 240) under the timed condition, 194 data sets (out of 240) under the un-timed condition. LSMM1 was preferred for 15 data sets (out of 240) under the timed condition, 23 data sets (out of 240) under the un-timed condition. Finally, LSMM2 was the preferred model for 5 data sets (out of 240) under the timed condition, 8 data sets (out of 240) under the un-timed condition.

There were few switches among substantive models under the timing manipulation. Collapsing across stimulus sets and display conditions, there were only 23 data sets (out of 240) that exhibited a switch from one model to another. Surprisingly, only 7 were in the hypothesized direction of being classified WOMM under the un-timed condition to either LSMM1 or LSMM2 under the timed condition. The remaining 16 switches were in the opposite direction, although this difference in rate was not statistically significant. As in the previous condition, the absolute number of switches was very small relative to the total number of data sets. There were no significant interaction effects across conditions.

5.7 Display Format Results

Averaging across stimulus sets and time pressure conditions, the marginal classification rates for WOMM under the display manipulation were: 198 data sets (out of 240) under the “Pie” condition and 195 data sets (out of 240) under the “Bar Graph” condition. LSMM1 was the preferred model for 19 data sets (out of 240) under the “Pie” condition and for 19 data sets (out of 240) under the “Bar Graph” condition. LSMM2 was the preferred model for 1 data set (out of 240) under the “Pie” condition and for 12 data sets (out of 240) under the “Bar Graph” condition. This difference in classification rate for the LSMM2, while too small to allow a statistical test, is in the predicted direction.

We investigated whether a DM originally classified under WOMM could be induced to ‘switch’ to either LSMM1 or LSMM2. Collapsing across stimulus sets and timing conditions, a total of 48 data sets (out of 240) exhibited a ‘switch’ from one substantive model to another. Thirty of these switches were in the hypothesized direction, from
WOMM under the “Pie” condition to either LSMM1 or LSMM2 under the “Bar Graph” condition, with the remaining 18 data sets switching in the opposite direction, from either LSMM1 or LSMM2 under the “Pie” condition to WOMM under the “Bar Graph” condition. This difference in switching rate was significantly greater than chance, yielding a $p$-value equal to 0.037, but we caution that this effect was very small and the majority of participants did not “switch” in their best-fitting substantive model across conditions. There were no significant interaction effects across conditions.

Table 8: The table above/below presents the marginal classification rates for WOMM, LSMM1 and LSMM2 across timing and stimulus conditions. The table shows how often (out of 240 for each display condition) each model was the preferred model.

<table>
<thead>
<tr>
<th>Display Condition</th>
<th>Pie</th>
<th>Bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMM</td>
<td>198</td>
<td>195</td>
</tr>
<tr>
<td>LSMM1</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>LSMM2</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 9: The table above/below presents the marginal classification rates for WOMM, LSMM1 and LSMM2 across display and timing conditions. The table shows how often (out of 240 for each stimulus set) each model was the preferred model.

<table>
<thead>
<tr>
<th>Stimulus Condition</th>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMM</td>
<td>200</td>
<td>202</td>
</tr>
<tr>
<td>LSMM1</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>LSMM2</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

5.8 Stimulus Results

Collapsing across timing and display format conditions, the marginal classification rates for WOMM under the two stimulus sets were: 200 data sets (out of 240) under stimulus set 1 and 202 (out of 240) under stimulus set 2. LSMM1 was the preferred model by our Bayes factor methodology for 24 data sets (out of 240) under stimulus set 1 and 18 data sets (out of 240) under stimulus set 2. LSMM2 was the preferred model for 5 data sets (out of 240) under stimulus set 1 and 9 data sets (out of 240) under stimulus set 2. The classification rates of the three models across stimulus set one were comparable to those obtained in our re-analysis of the Regenwetter and Davis-Stober (2012) study. In support of our hypothesis, the classification rate for LSMM2 was larger for stimulus set 2 than stimulus set 1; however, these rates are too small to for carrying out a statistical test.
6 Discussion

We presented a new class of mixture model based upon lexicographic semiorders. We demonstrated that such models can explain the choice behavior of DMs whose preferences are intransitive. Using Bayesian order-constrained statistical methodology, we carried out a comprehensive test of algebraic preference structure by simultaneously evaluating the weak order mixture model of Regenwetter and Davis-Stober (2012) and a new class of mixture model based on lexicographic semiorders. By analyzing prior data, as well as a new study, we confirmed the general conclusions of Regenwetter and Davis-Stober (2012) and Regenwetter et al. (2011), finding that a majority of participants made choices consistent with strict weak orders. Our analysis adds several additional conclusions, namely that a minority of individuals appear to make choices consistent with lexicographic semiorders.

This paper adds to a growing literature demonstrating the robustness of algebraic structure of preference among DMs (Birnbaum & Gutierrez, 2007; Birnbaum & Schmidt, 2008; Cavagnaro & Davis-Stober, 2013), especially that of utility representations. Contrary to previous studies, however, we consistently find support for the lexicographic semiorder structure among a substantial minority of individuals. This result indicates that utility representations are likely descriptive for many, but not all individuals - cautioning against a “one-size-fits all” approach to decision modeling, analogous to the conclusions by Loomes, Moffatt, and Sugden (2002) and Hey (2005).

Our statistical methodology coupled with the parsimony of the mixture model theories and richness of the ternary choice paradigm afforded excellent model distinguishability. The Bayes factors between the weak order and lexicographic semiorder mixture models were quite large for nearly all of the participants in both our re-analysis and new study. This indicates that the weak order and lexicographic semiorder mixture models are distinguishable from one another, in the sense that one is orders of magnitude more likely than the other from our Bayesian model selection perspective.

There is evidence in the literature that individual differences may also extend to stochastic specification (Cavagnaro & Davis-Stober, 2013). In our analyses, we considered only one stochastic specification to provide a direct test of algebraic preference structure. However, even with the extreme parsimony of both WOMM and LSMM models, at least one of these was found to be an adequate fit for the vast majority of data sets. Introducing alternative stochastic specifications are possible, but would need to be defined and extended to the ternary choice framework. Future work could explore evaluating alternative stochastic specifications. Additionally testing for differences in stochastic and algebraic structure simultaneously open up combinatorial difficulties.

We found weak evidence that display format can influence whether participants’ choices are consistent with either model. We stress that this effect was relatively small in size, and that a majority of participants did not ‘switch’ in their preferred model across experimental conditions. We found that direct time pressure did not induce more participants to make choices consistent with lexicographic semiorders, nor did this manipulation induce individual participants to switch from one model to another. A possible explana-
tion of this result is that the choice alternatives we used were relatively simple with only two attributes, in contrast to the stimuli of Rieskamp and Hoffrage (1999).
7 References


Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints:


