Journal of Mathematical Psychology 66 (2015) 70–82

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Individual differences in the algebraic structure of preferences

Clintin P. Davis-Stober [a,*], Nicholas Brown [a], Daniel R. Cavagnaro [b]

[a] *University of Missouri, USA*

[b] *Mihaylo College of Business and Economics, California State University Fullerton, P.O. Box 6848, Fullerton, CA 92834-6848, USA*

HIGHLIGHTS

- We examine the algebraic structure of individual preference.
- We evaluate a large class of weak-order and lexicographic semiorder based theories.
- We present a new study as well as a re-analysis of an existing data set.
- We find that a majority of subjects' preferences are consistent with weak orders.
- We find that the remaining subjects are well-described by lexicographic semiorders.

ARTICLE INFO

Article history:
Received 14 December 2013
Received in revised form
29 October 2014
Available online 2 March 2015

Keywords:
Preferential choice
Choice modeling
Lexicographic semiorders
Weak orders
Convex polytopes
Random preference
Order-constrained inference

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 across multiple data sets. We report the results of a new experiment, which 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 semi-ordered preferences.

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. Theorists commonly impose idealized conditions on these preferences in order to generate tractable models. 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 many of the most prominent models of decision making, including Expected Utility Theory (von Neumann & Morgenstern, 1947) and Cumulative Prospect Theory (Tversky & Kahneman, 1992). Yet, 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), Fishburn (1991), and Regenwetter, Dana, and Davis-Stober (2011) for comprehensive reviews of the arguments both for and against transitivity.

Tversky's (1969) experiment, and many of those that followed (e.g. Birnbaum, 2010; Birnbaum & Gutierrez, 2007; Montgomery, 1977; Ranyard, 1977, 1982), were designed to elicit intransitive choice patterns arising from a particular ordered collection of semi-ordered preferences, called a lexicographic semiorder. The idea of representing preferences as semiorders was introduced by Luce (1956) (although see Armstrong, 1939; Georgescu-Roegen, 1936, for earlier, related work) 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 without sugar, the other with one microgram of sugar. Since

* Correspondence to: University of Missouri, Department of Psychological Sciences, Columbia, MO, 65211, USA.
E-mail address: stoberc@missouri.edu (C.P. Davis-Stober).

http://dx.doi.org/10.1016/j.jmp.2014.12.003
0022-2496/

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 US 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 differ by a sufficient margin, i.e., are "distinguishable." 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 lexical-based heuristics, such as lexicographic semiorders, 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 & Brighton, 2009), yet recent tests of lexicographic models such as Take the Best (Gigerenzer & Goldstein, 1996) and the Priority Heuristic[1] (Brandstätter, Gigerenzer, & Hertwig, 2006), have found mixed to little empirical support (e.g. Birnbaum, 2008; Birnbaum & Gutierrez, 2007; Glöckner & Betsch, 2008; Lee & Cummins, 2004).

Why has the literature been unable to reach a consensus on the algebraic structure of preferences despite decades of research? We argue that four, long-standing, theoretical and methodological conventions have hindered progress:

1. Due to the mathematical complexity of higher-order choice structures, studies have used a binary forced choice framework that does not explicitly include indifference, even though indifference is a defining aspect of the lexicographic semiorder structure (see Regenwetter & 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 to test the algebraic structures, e.g., trembling hand (Harless & Camerer, 1994), true-and-error (Birnbaum, 2011), and random utility (Regenwetter, Dana, Davis-Stober, & Guo, 2011). This confounds the test of the algebraic structure with a test of the stochastic framework. See Hey (2005) and Loomes, Moffatt, and Sugden (2002) for discussion.
3. Due to the limitations of statistical analyses based solely on classical goodness-of-fit tests and/or the complications of order-constrained inference, many previous analyses have not appropriately penalized models for complexity (e.g. Tversky, 1969); — see Davis-Stober & Brown, 2011, for discussion. This can bias the results in favor of 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 (e.g., Regenwetter, Dana, Davis-Stober, & Guo, 2011; Regenwetter & Davis-Stober, 2012). As a result, rather than offering an alternative explanation to "rationalize" violating data, statistically significant violations are attributed to either irrational behavior or Type I error.

[1] Depending on the choice stimuli being considered, the Priority Heuristic is either a lexicographic semiorder or a lexicographic interval order.

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 re-analysis 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 preference, thereby providing a richer mapping between true preference and choice behavior. Under an 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, while also being testable and extremely parsimonious, as we will show.

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 class of LSMMs to that of a viable competitor model. However, to overcome the limitations of methodological convention (2) 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 (2012). Like the LSMM, the WOMM is also defined for ternary choice, and also allows DMs to move from one preference state to 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. Heyer & Niederée, 1992; Loomes & Sugden, 1995). 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. In other words, the models allow for different individuals to have very different distributions over their preference states, while still capturing the core property of interest (weak orders or lexicographic semiorders). Yet, while both the WOMM and the 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. In order to compute a Bayes factor, for each model, we place a prior distribution over the set of all possible probability distributions over the set of viable preference relations for that model. As we describe in detail in later sections, following a data-driven approach, we assume a uniform prior over these distributions for each of the WOMM and LSMM models.

1.2. Re-analysis 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 an LSMM, could provide a better, more parsimonious description of the data. Therefore, to overcome the limitations of methodological conventions (3) and (4) from the list above, we use Bayesian model selection to compare the WOMM with our LSMMs in the data from Regenwetter and Davis-Stober (2012). In particular, we compute the Bayes factor for each model using the order-constrained inference method of Klugkist and Hoijtink (2007). The Bayes factor (Kass & 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). There are alternative model selection measures that properly account for complexity, such as normalized maximum likelihood (Myung, Navarro, & Pitt, 2006) which is based on the minimum description length principle (Grünwald, 2007). See Davis-Stober and Brown (2011) for an application of normalized maximum likelihood to distribution-free random preference models.

In our re-analysis, 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

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 and Hoffrage (1999, 2008) showed that DMs are more likely to apply lexicographic strategies when the time to make a decision is greatly limited. Similarly, many studies have demonstrated that the display format of stimuli 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) (e.g. Brandstätter, 2011; Kleinmuntz & Schkade, 1993). 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 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 (3) 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 Estes (1956) and Luce (2000).

Our findings from the experiment confirm our conclusions from the re-analysis in that a majority of participants are best described by the WOMM, with a substantial minority of participants best described by the LSMM. Individuals rarely switched classifications across our experimental conditions. These results further suggest that the individual differences we found in the re-analysis 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 the 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 re-analysis of the data from Regenwetter and Davis-Stober (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 (numerical utility representation)

Let $\mathcal{A}$ be a finite 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. 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 the 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 the WOMM is that no additional structure is placed on the probability distribution over the set of strict weak orders, i.e., the WOMM allows any probability distribution over strict weak orders. Let $\succ$ be a strict weak order on $\mathcal{A}$. Let $\mathcal{WO}_{\mathcal{A}}$ denote the set of all strict weak orders on $\mathcal{A}$. A system of ternary choice probabilities satisfies the WOMM if there exists a probability distribution on $\mathcal{WO}_{\mathcal{A}}$,

$$Prob: \mathcal{WO}_{\mathcal{A}} \to [0, 1], \quad \succ \mapsto \mathcal{P}_{\succ},$$

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

$$\mathcal{P}_{a,b} = \sum_{\succ \in \mathcal{WO}_{\mathcal{A}}, a \succ b} \mathcal{P}_{\succ}.$$

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), the 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).

In principle, the 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 a numerical algorithm, the POlyhedron Representation Transformation Algorithm (PORTA) (Christof & Löbel, 1997), 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 the WOMM for n larger than 5.

2.2. Interpretation of the WOMM

Since the WOMM allows an arbitrary probability distribution over all strict weak orders, a violation of the 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 the WOMM is defined at the level of ternary choice probabilities and does not require the specification of any particular utility function. In this way, the 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 the WOMM is a very restrictive model. For $n = 5$, the number of allowable preference states under the WOMM is equivalent to the number of weak orders over n alternatives, which is equal to 541; this is 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 the WOMM for $n = 5$, and, using Monte Carlo methods, found this proportion to be slightly less than 0.0005.

[2] From a geometric perspective, the 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.

2.3. Lexicographic semiorder mixture models (LSMM1 and LSMM2)

Consider the case of a DM that does not satisfy the 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 is defined on ternary choice probabilities and is compatible with *intransitive* preference. We begin with some preliminary definitions.

Let S be a binary relation on $\mathcal{A}$. S is a *strict semiorder* if, and only if, there exist a real-valued function g, defined on $\mathcal{A}$, and a non-negative constant q such that, $\forall a, b \in \mathcal{A}$,

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

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

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

It is routine to show that T defines a weak ordering on $\mathcal{A}$.

In general, a lexicographic semiorder is defined as a binary relation on $\mathcal{A}$ characterized by an ordered collection of semiorders (Pirlot & Vincke, 1997). In this article, we consider a special case of lexicographic semiorders, i.e., our definition of a lexicographic semiorder maintains a similar structure to previous definitions but introduces two important constraints. A *simple lexicographic semiorder* (Davis-Stober, 2010, 2012) is a relation P on $\mathcal{A}$ such that there exist two semiorders S_1 and S_2 on $\mathcal{A}$, that satisfy, $\forall a, b \in \mathcal{A}$:

(i) the trace of S_1 is the reverse of the trace of S_2.
(ii) $aPb \iff aS_1b$, or $[aS_2b$ and $\neg[aS_1b$ 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 $a, b \in \mathcal{A}$, if $(a, b) \in T_1$, then $(b, a) \in T_2$, and reciprocally.

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 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 level of quality 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 \$60 but camera B is of higher quality, and would prefer C to B, since the price difference is again less than \$60 but

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.

Camera	Price	Quality
A	\$200	Low
B	\$250	Medium
C	\$300	High

camera C is of higher quality, yet would also prefer A to C, since A is more than \$60 cheaper than C.

To create viable competitor models for the WOMM, we will consider a particular class of simple lexicographic semiorder. A relation, R_1, defined on $\mathcal{A}$, is *compatible* with another relation, R_2, also defined on $\mathcal{A}$, if $\forall a, b \in \mathcal{A}$, $(a, b) \in R_1 \Rightarrow (a, b) \in R_2$. Let $\mathcal{SLS}_{\mathcal{A}}$ denote the set of all simple lexicographic semiorders such that the trace of S_1 is compatible with a fixed linear ordering on $\mathcal{A}$. In general, this fixed linear ordering on $\mathcal{A}$ 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, $\mathcal{SLS}_{\mathcal{A}}$ 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 $\mathcal{SLS}_{\mathcal{A}}$, 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 $\mathcal{SLS}_{\mathcal{A}}$,

$$Prob: \mathcal{SLS}_{\mathcal{A}} \rightarrow [0, 1], \quad P \mapsto \mathcal{P}_P,$$

such that $\forall a, b \in \mathcal{A}, a \neq b$,

$$\mathcal{P}_{a,b} = \sum_{P \in \mathcal{SLS}_{\mathcal{A}}, (a,b) \in P} \mathcal{P}_P.$$

Said another way, the 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 $\mathcal{SLS}_{\mathcal{A}}$ such that a is strictly preferred to b. Similar to the interpretation of the 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 the WOMM, we place no restrictions on the probability distribution over the elements of $\mathcal{SLS}_{\mathcal{A}}$.

In our empirical analysis, $\mathcal{A}$ 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 relabeling the corresponding ternary choice probabilities according the fixed ordering on $\mathcal{A}$, see Davis-Stober (2012) for a complete discussion.

In contrast to the 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 $\mathcal{A}$ are ordered in a strictly increasing fashion according to a fixed linear ordering of $\mathcal{A}$. A collection of ternary choice probabilities satisfy the LSMM if, and only if, the following seven families of inequalities are satisfied:

$$\mathcal{P}_{i,j} - \mathcal{P}_{i,j+1} \leq 0, \quad \forall (i,j) \in \{1, 2, \ldots, n\} : i < j < n, \tag{2}$$

$$\mathcal{P}_{i+1,j} - \mathcal{P}_{i,j} \leq 0, \quad \forall (i,j) \in \{1, 2, \ldots, n\} : i+1 < j, \tag{3}$$

$$\mathcal{P}_{j,i} + \mathcal{P}_{i,j} - \mathcal{P}_{j+1,i} - \mathcal{P}_{i,j+1} \leq 0, \quad \forall (i,j) \in \{1, 2, \ldots, n\} : i < j < n, \tag{4}$$

$$\mathcal{P}_{i+1,j} + \mathcal{P}_{j,i+1} - \mathcal{P}_{i,j} - \mathcal{P}_{j,i} \leq 0, \quad \forall (i,j) \in \{1, 2, \ldots, n\} : i+1 < j, \tag{5}$$

$$-\mathcal{P}_{i,i+1} \leq 0, \quad \forall i \in \{1, 2, \ldots, n-1\}, \tag{6}$$

$$-\mathcal{P}_{i,j} \leq 0, \quad \forall (i,j) \in \{1, 2, \ldots, n\} : i > j, \tag{7}$$

$$\mathcal{P}_{1,n} + \mathcal{P}_{n,1} \leq 1. \tag{8}$$

The total number of non-redundant linear inequalities that characterize an 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 et al., 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 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 the LSMMs and the 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 $\mathcal{SLS}_{\mathcal{A}}$ is a simple function of the Catalan numbers,

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

where $C_n = \frac{1}{n+1}\binom{2n}{n}$.

This result allows us to directly compare the WOMM and the LSMM to one another in terms of how many preference states are allowed under the two models. Remarkably, the WOMM and the LSMM have comparable numbers of preference states for many values of n, see Table 2. In all of the experiments we consider in this paper, $n = 5$, thus there are 594 possible preference states under an LSMM as compared to 541 for the WOMM. Thus, one model is not a simple "straw-man" model as compared to the other. It is interesting that the two classes of models have such similar numbers of preference states given that they are derived from very different preference structures.

Table 2
Number of allowable preference states for the WOMM and the LSMM as a function of the number of choice alternatives, $|\mathcal{A}| = n$.

n	WOMM	LSMM
3	13	14
4	75	84
5	541	594
6	4,683	4,719
7	47,293	40,898

While the number of preference states are comparable to the WOMM, 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 the LSMM to be roughly 0.00002 (for $n = 5$). This is an order of magnitude more restrictive than the WOMM.

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 & 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,

$$L(\boldsymbol{\mathcal{P}}|\mathbf{N}) = \prod_{\forall(a,b)\in\mathcal{A}} \frac{N!}{N_{a,b}!N_{b,a}!(N-N_{a,b}-N_{b,a})!}\mathcal{P}_{a,b}^{N_{a,b}}\mathcal{P}_{b,a}^{N_{b,a}} \times (1-\mathcal{P}_{a,b}-\mathcal{P}_{b,a})^{N-N_{a,b}-N_{b,a}}, \quad (9)$$

where $\boldsymbol{\mathcal{P}} = P_{a,b}, \forall a, b \in \mathcal{A}, a \neq b, \mathbf{N} = N_{a,b}, \forall a, b \in \mathcal{A}$, $a \neq b$, 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, $\mathcal{P}_{a,b}$, 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). The Bayes factor (BF) is a measure of the strength of evidence for a particular model over another. It is calculated as the ratio of the marginal likelihoods for two models. The resulting BF can then be used in model comparison, allowing us to state evidence for which, if any, model is likely to have produced the data. Larger (resp. smaller) BFs indicate evidence in favor of the model in the numerator (resp. denominator) of the BF. For an introduction to the use of BFs and Bayesian methodology see Jackman (2009).

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 the WOMM, LSMM1, and LSMM2 are nested in M_1, which affords relatively simple numerical calculation of the Bayes factor via M_1. Let M_t be the mixture model specification being evaluated, either the 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,

$$BF_{t1} = \frac{p(\mathbf{N}|M_t)}{p(\mathbf{N}|M_1)} = \frac{\int L(\mathbf{N}|\boldsymbol{\mathcal{P}})\pi(\boldsymbol{\mathcal{P}}|M_t)d\boldsymbol{\mathcal{P}}}{\int L(\mathbf{N}|\mathcal{P})\pi(\boldsymbol{\mathcal{P}}|M_1)d\boldsymbol{\mathcal{P}}}, \quad (10)$$

where $L(\mathbf{N}|\boldsymbol{\mathcal{P}})$ is the density corresponding to the statistical model in (9) and $\pi(\boldsymbol{\mathcal{P}}|M_t)$ is the prior distribution of $\boldsymbol{\mathcal{P}}$ under model M_t, which is defined to be uniform on the support of M_t. Similarly, $\pi(\boldsymbol{\mathcal{P}}|M_1)$ is the prior distribution of $\boldsymbol{\mathcal{P}}$ under model M_1, which is defined to be uniform over the set of all ternary choice probabilities. 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, the WOMM and the LSMMs are extraordinarily parsimonious by this measure.

As described more fully in Klugkist and Hoijtink (2007), Eq. (10) can be further simplified. The Bayes factor under this inequality-constrained framework can be described as the ratio of two proportions: the proportion of the encompassing prior in agreement with the constraints of M_t and the proportion of the encompassing posterior distribution in agreement with the constraints of M_t. This simplification gives,

$$BF_{t1} = \frac{c_t}{d_t}, \quad (11)$$

where $\frac{1}{c_t}$ is the proportion of the encompassing prior in agreement with the constraints of M_t and $\frac{1}{d_t}$ is the proportion of the encompassing posterior distribution in agreement with the constraints of M_t. The proportion, $\frac{1}{c_t}$, for both the LSMM and the WOMM is already known via uniform Monte Carlo sampling, – see Davis-Stober (2012) and Regenwetter and Davis-Stober (2012). We calculated the $\frac{1}{d_t}$ terms using standard Monte Carlo sampling methods, i.e., random draws from the conjugate beta posterior formed by a uniform prior over $\boldsymbol{\mathcal{P}}$ in Eq. (9). We refer interested readers to Klugkist and Hoijtink (2007) for additional details on this statistical method.

For each pair $a, b \in \mathcal{A}$, we assume a Dirichlet prior over $\boldsymbol{\mathcal{P}}_{a,b} = \{\mathcal{P}_{a,b}, \mathcal{P}_{b,a}, 1-\mathcal{P}_{a,b}-\mathcal{P}_{b,a}\}$, which is conjugate to the multinomial likelihood function in Eq. (9). Setting the concentration parameters of the Dirichlet prior equal to 1, so that the prior is uniform over systems of ternary choice probabilities, the posterior distribution given a set of ternary choice data $\mathbf{N}_{a,b} = \{N_{a,b}, N_{b,a}\}$ is

$$\boldsymbol{\mathcal{P}}_{a,b}|\mathbf{N}_{a,b} \sim Dir(1+N_{a,b}, 1+N_{b,a}, 1+N-N_{a,b}-N_{b,a}).$$

We estimated d_t for each model and data set by generating 5,000,0000 random draws from this posterior distribution, each time checking whether the sampled point satisfies the model constraints (e.g., Eqs. (2)–(8)).

Table 3
Regenwetter and Davis-Stober (2012) experimental Stimuli, where $\{a, b, c, d, e\}$ constitute the set of choice alternatives.

a	b	c	d	e
Gamble Set I				
(\$26.32, $\frac{7}{24}$)	(\$25.00, $\frac{8}{24}$)	(\$23.68, $\frac{9}{24}$)	(\$22.36, $\frac{10}{24}$)	(\$21.04, $\frac{11}{24}$)
Gamble Set II				
(\$31.43, 0.28)	(\$27.50, 0.32)	(\$24.44, 0.36)	(\$22.00, 0.40)	(\$20.00, 0.44)

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 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 ($p < 0.05$), out of 90 total data sets (30 participants times 3 stimulus sets). They concluded that the violations of the 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 neither compare the WOMM to a competitor model, nor adequately penalize the WOMM for complexity.

We conducted a re-analysis of this data, using the Klugkist and Hoijtink (2007) Bayes factor methodology, to compare the WOMM to 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 3. Stimulus Set 1 was an updated version of the classic Tversky (1969) stimulus 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 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 4 and 5 present the results of our re-analysis. The second column of Table 4 displays the best fitting model, WOMM, LSMM1, or LSMM2, according to the Bayes factor analysis for stimulus Set 1. All Bayes factors are reported on a natural log scale. 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 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 the 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 "$\surd$" indicates that the marginal choice proportions were perfectly consistent with the WOMM. We repeat this information for Set 2 in Table 5.

As seen in Tables 4 and 5, our results largely agree with the findings of Regenwetter and Davis-Stober (2012) with some exceptions. We found that the WOMM was the best model for 48 out of the 60 data sets (30 participants over two stimulus sets) with natural log scale Bayes factors much larger than 4.61 (corresponding to a Bayes factor of over 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 the 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 stimulus Set 2, for whom the best model according to the Bayes factor was the encompassing model (the Bayes factor for the WOMM, the best performing mixture model, against the encompassing was 0.05). 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 $\alpha = 0.05$), 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 the WOMM was the best model for the majority of DMs, LSMM1 provided a superior description of behavior for several DMs, including those that did not fit the 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 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 stimuli, can we manipulate experimental conditions in such a way that DMs will adjust their underlying preference structures, in the sense of Brandstätter (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.

Table 4
This table displays the Bayes Factor (BF) analysis for stimulus Set 1 of the Regenwetter and Davis-Stober experiment. All BFs are reported on a natural log scale. 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. The sixth column reproduces the classical goodness-of-fit analysis of Regenwetter and Davis-Stober (2012). A "√" indicates that the marginal choice proportions satisfied the WOMM — since these data do not violate the model they cannot be used as evidence against it. If the marginal choice proportions violated the WOMM, then the G^2 value is listed along with its respective p-value in parentheses.

Participant	Set 1	BF	Next best (set 1)	BF	Original analysis
1	WOMM	9.49[c]	LSMM1	6.73[c]	√
2	WOMM	7.11[c]	LSMM1	5.37[c]	√
3	WOMM	9.49[c]	LSMM1	10.20[c]	√
4	WOMM	8.39[c]	LSMM1	0.22	√
5	WOMM	8.52[c]	LSMM1	3.94[b]	√
6	WOMM	8.78[c]	LSMM1	6.47[c]	√
7	WOMM	9.73[c]	LSMM1	6.05[c]	√
8	**LSMM1**	6.09[c]	WOMM	1.72[a]	√
9	WOMM	6.37[c]	LSMM1	8.62[c]	0.52 (0.82)
10	WOMM	7.65[c]	LSMM1	7.15[c]	2.82 (0.47)
11	WOMM	6.84[c]	LSMM1	5.21[c]	√
12	WOMM	9.21[c]	LSMM1	6.25[c]	√
13	WOMM	5.85[c]	LSMM1	4.65[c]	√
14	**LSMM1**	12.07[c]	WOMM	9.41[c]	8.01 (**0.02**)
15	**LSMM1**	12.07[c]	WOMM	9.40[c]	√
16	WOMM	9.66[c]	LSMM1	8.19[c]	√
17	WOMM	8.93[c]	LSMM1	4.72[c]	√
18	WOMM	9.73[c]	LSMM1	9.68[c]	√
19	WOMM	9.74[c]	LSMM1	9.68[c]	√
20	WOMM	9.46[c]	NA	NA	√
21	WOMM	9.54[c]	LSMM1	5.21[c]	√
22	**LSMM1**	6.27[c]	WOMM	4.48[b]	5.01 (0.10)
23	WOMM	8.64[c]	LSMM1	6.91[c]	0.16 (0.95)
24	WOMM	8.61[c]	LSMM1	6.36[c]	√
25	**LSMM1**	10.37[c]	WOMM	1.73[a]	√
26	WOMM	5.75[c]	LSMM1	4.54[b]	1.93 (0.48)
27	WOMM	9.06[c]	LSMM2	11.31[c]	√
28	WOMM	9.41[c]	LSMM1	6.76[c]	√
29	**LSMM1**	1.91[a]	WOMM	0.70	6.04 (0.26)
30	**LSMM1**	7.80[c]	WOMM	1.71[a]	1.03 (0.54)

[a] Indicates a BF greater than 3 (1.10 on natural log scale).
[b] Indicates a BF greater than 10 (2.30 on natural log scale).
[c] Indicates a BF greater than 100 (4.61 on natural log scale).

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, totaling 960 trials.

Participants indicated preference for each ternary choice trial via key presses on a keyboard. Presentation of the trials was 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.

5.1.1. Time pressure manipulation

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 s. The average response time for the timed condition across participants was 1.681 s. 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 levels. 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 an LSMM model for the time pressure condition.

5.1.2. Display format manipulation

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,

Table 5
This table displays the Bayes Factor (BF) analysis for stimulus Set 2 of the Regenwetter and Davis-Stober experiment. All BFs are reported on a natural log scale. 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. The sixth column reproduces the classical goodness-of-fit analysis of Regenwetter and Davis-Stober (2012). A "√" indicates that the marginal choice proportions satisfied the WOMM — since these data do not violate the model they cannot be used as evidence against it. If the marginal choice proportions violated the WOMM, then the G^2 value is listed along with its respective p-value in parentheses. Underlined models indicate the best-fitting model was not favored over the encompassing model.

Participant	Set 2	BF	Next best (set 2)	BF	Original analysis
1	WOMM	9.71[c]	LSMM1	5.73[c]	√
2	WOMM	9.71[c]	NA	NA	√
3	WOMM	8.02[c]	LSMM2	11.37[c]	1.11 (0.77)
4	WOMM	9.72[c]	LSMM1	5.17[c]	√
5	WOMM	7.77[c]	LSMM2	11.11[c]	0.17 (0.89)
6	WOMM	9.28[c]	LSMM1	6.24[c]	√
7	WOMM	9.68[c]	NA	NA	√
8	WOMM	7.67[c]	LSMM1	3.69[b]	√
9	**LSMM1**	5.01[c]	WOMM	4.07[b]	5.95 (0.18)
10	WOMM	9.46[c]	LSMM1	6.15[c]	√
11	WOMM	7.26[c]	LSMM1	3.75[c]	√
12	WOMM	8.88[c]	LSMM1	6.60[c]	√
13	WOMM	5.85[c]	LSMM1	4.42[c]	√
14	**LSMM1**	8.33[c]	WOMM	2.09[a]	1.22 (0.38)
15	**LSMM1**	8.33[c]	WOMM	2.20[c]	0.80 (0.40)
16	WOMM	9.71[c]	LSMM1	9.30[c]	√
17	WOMM	8.95[c]	NA	NA	√
18	WOMM	9.07[c]	NA	NA	√
19	WOMM	9.22[c]	LSMM1	7.58[c]	√
20	WOMM	9.26[c]	LSMM2	12.61[c]	√
21	WOMM	9.25[c]	LSMM1	4.52[b]	√
22	<u>WOMM</u>	−3.00	NA	NA	14.90 (<**0.01**)
23	WOMM	8.80[c]	LSMM1	4.97[c]	0.04 (0.89)
24	WOMM	8.90[c]	LSMM1	5.55[c]	√
25	**LSMM1**	11.25[c]	WOMM	2.57[b]	√
26	WOMM	5.96[b]	LSMM1	8.62[c]	1.30 (0.75)
27	WOMM	8.93[c]	LSMM2	10.67[c]	√
28	WOMM	9.59[c]	LSMM1	6.71[c]	√
29	**LSMM1**	3.26[b]	WOMM	8.15[c]	14.81 (**0.01**)
30	WOMM	7.67[c]	LSMM1	3.66[b]	0.04 (0.88)

[a] Indicates a BF greater than 3 (1.10 on natural log scale).
[b] Indicates a BF greater than 10 (2.30 on natural log scale).
[c] Indicates a BF greater than 100 (4.61 on natural log scale).

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 stimulus Sets 1 and 2. In the "Pie" condition, we displayed the gambles according to the common "pie" format used in Regenwetter, Dana, Davis-Stober, and Guo (2011), Regenwetter and Davis-Stober (2012), Tversky (1969), among others (Birnbaum & LaCroix, 2008; Montgomery, 1977), see the left-hand side of Fig. 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 Fig. 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 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.

5.1.3. *Stimulus manipulation*

For this study, we used two stimulus sets of five gambles each. These gambles are displayed in Table 6. 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 stimulus Set 1, the expected value of the binary gambles increases as the probability of a gain increases. In our stimulus 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 5. To the extent that different stimuli could alter a DM's underlying preference structure, this suggests an increase in the use of LSMM2.

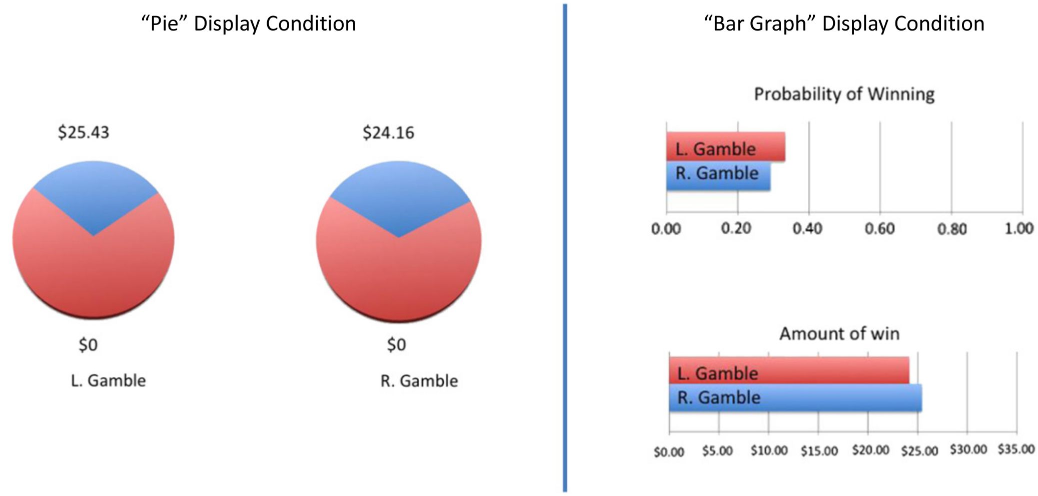

Fig. 1. This figure shows the same gamble pair under both the "Pie" and "Bar Graph" display conditions.

Table 6
Experimental stimuli, time pressure/perceptual display manipulation study.

a	b	c	d	e
Gamble Set I				
(\$25.43, $\frac{7}{24}$)	(\$24.16, $\frac{8}{24}$)	(\$22.89, $\frac{9}{24}$)	(\$21.62, $\frac{10}{24}$)	(\$20.35, $\frac{11}{24}$)
Gamble Set II				
(\$31.99, $\frac{7}{24}$)	(\$27.03, $\frac{8}{24}$)	(\$22.89, $\frac{9}{24}$)	(\$19.32, $\frac{10}{24}$)	(\$16.19, $\frac{11}{24}$)

5.2. Results

Across all combinations of conditions, there were a total of 480 data sets to analyze (60 participants, two stimulus sets, two display conditions, two timing conditions, $60 \times 2 \times 2 \times 2 = 480$). For each of the 480 data sets, we computed Bayes factors for each substantive model (WOMM, LSMM1, and LSMM2) against the encompassing model, using the methodology of Klugkist and Hoijtink (2007). To simplify presentation, we only report a model as best if it scored the largest Bayes factor among the three substantive models *and* scored a Bayes factor of at least $10^{0.5} \approx 3.16$ against the encompassing model (at least 1.15 on the natural log scale), as this is the recommended threshold for interpreting the Bayes factor as providing "substantial" evidence (Jeffreys, 1998). However, the analyses generally yielded very large Bayes factors for at least one substantive model, meaning that there was a clearly superior model for most data sets. The average natural log scale Bayes factor for the WOMM when it was the best model was 7.88, and the average natural log scale Bayes factor for LSMM1 and LSMM2 when one was classified as the best model was 8.51. All calculated Bayes factors for all subjects across all conditions are reported in the accompanying online appendix.

As shown in Table 7, the WOMM was the best model (highest Bayes factor) for 396 out of the 480 data sets (approximately 82%). LSMM1 was best for 45 out of 480 (approximately 9%), and LSMM2 was best for 13 out of 480 (approximately 3%). At the participant level, 18 out of 60 participants (30%) were best fit by one of the lexicographic models for at least one condition, closely matching the proportion that was found in our re-analysis of the Regenwetter and Davis-Stober (2012) data: 8 out of 30 participants (27%) best fit by an LSMM for at least one condition.

5.2.1. Effect of time pressure manipulation

The time pressure manipulation did not affect the proportions of data sets that were best described by each model. As shown in Table 8, the number of data sets that were best described by each substantive model are roughly equal in the timed and un-timed conditions. Out of 240 participant/stimulus/display combinations, the number of times that the WOMM, LSMM1 and LSMM2 were best were 200, 23, and 5, respectively, in the timed condition, and 196, 22, and 8, respectively, in the un-timed condition. This result suggests that the underlying preference structure is robust (at the aggregate level) with respect to the time pressure manipulation. To see if it is also robust at the individual level, we investigated whether a DM originally classified as the WOMM could be induced to 'switch' to either LSMM1 or LSMM2, or vice-versa, under a different time pressure condition. In fact, there were few switches within participants among substantive models under the timing manipulation. Combining counts across stimulus sets and display conditions, there were only 24 data sets (out of 240) that exhibited a switch from one model to another. Surprisingly, only 9 were in the hypothesized direction of being classified as WOMM under the un-timed condition to either LSMM1 or LSMM2 under the timed condition. The remaining 15 switches were in the opposite direction, although this difference in rate was not statistically significant. The absolute number of switches was very small relative to the total number of data sets, and there were no significant interactions between time pressure and the other conditions.

Table 7
Number of data sets (out of 480 participant/timing/display/stimulus combinations) for which each model provided the best explanation (highest Bayes factor on natural log scale among substantive models and at least 1.15 against the encompassing model).

	Overall results
WOMM	396
LSMM1	45
LSMM2	13

Table 8
Number of data sets in each timing condition (out of 240 for participant/display/stimulus combinations) for which each model provided the best explanation (highest Bayes factor on natural log scale among substantive models and at least 1.15 against the encompassing model).

	Timing condition	
	Timed	Un-timed
WOMM	200	196
LSMM1	23	22
LSMM2	5	8

5.2.2. Effect of display format manipulation

Combining counts across stimulus sets and time pressure conditions, the total classification counts in each display condition

Table 9
Number of data sets in each display condition (out of 240 participant/timing/stimulus combinations) for which each model provided the best explanation (highest Bayes factor on natural log scale among substantive models and at least 1.15 against the encompassing model).

	Display condition	
	Pie	Bar
WOMM	201	195
LSMM1	26	19
LSMM2	1	12

Table 10
Number of data sets in each stimulus condition (out of 240 participant/timing/display combinations) for which each model provided the best explanation (highest Bayes factor on natural log scale among substantive models and at least 1.15 against the encompassing model).

	Stimulus condition	
	Set 1	Set 2
WOMM	203	202
LSMM1	31	18
LSMM2	5	9

(out of 240 participant/stimulus/timing combinations) were: 201 and 195 for the WOMM under "Pie" and "Bar Graph", respectively; 26 and 19 for LSMM1 under "Pie" and "Bar Graph", respectively; and 1 and 12 for LSMM2 under "Pie" and "Bar Graph", respectively. This difference in counts for LSMM2, while too small to allow a statistical test, is in the hypothesized direction. See Table 9.

We investigated whether a DM originally classified under the WOMM could be induced to 'switch' to either LSMM1 or LSMM2. Again combining counts across stimulus sets and timing conditions, a total of 52 sets (out of 240) exhibited a 'switch' from one substantive model to another. Twenty-nine of these switches were in the hypothesized direction (i.e., from the WOMM under the "Pie" condition to either LSMM1 or LSMM2 under the "Bar Graph" condition) with the remaining 23 data sets switching in the opposite direction (i.e., from either LSMM1 or LSMM2 under the "Pie" condition to the WOMM under the "Bar Graph" condition). This difference in switching rate was not significantly greater than chance. There were no significant interactions between display format and the other conditions.

5.2.3. *Effect of stimulus manipulation*

As shown in Table 10, combining counts across timing and display conditions, the number of data sets for which the WOMM was best under the two stimulus sets were: 203 data sets (out of 240) under stimulus Set 1 and 202 (out of 240) under stimulus Set 2. LSMM1 was the best model by our Bayes factor methodology for 31 data sets (out of 240) under stimulus Set 1 and 18 data sets (out of 240) under stimulus Set 2. LSMM2 was the best 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 counts of the three models across stimulus Set 1 were comparable to those obtained in our re-analysis of the Regenwetter and Davis-Stober (2012) study. In support of our hypothesis, the classification counts for LSMM2 were larger for stimulus Set 2 than stimulus Set 1; however, these counts were too small to carry out a statistical test.

6. Discussion

In this article, we presented a comprehensive analysis of two of the most prominent preference structures in the decision literature: weak orders and lexicographic semiorders. We considered mixture model formulations of these algebraic structures. This approach is powerful, as it requires few assumptions at the level of the DM, e.g., it does not require a specific functional form to model DM choice. Further, by formulating a mixture model over all possible preference states consistent with a specified algebraic structure, we are able to draw strong conclusions from data analysis. Should such a mixture model be rejected, then there does not exist a probability distribution over preferences consistent with the algebraic structure of interest that would well-describe that DM's data. Conversely, as we demonstrated, both classes of mixture models that we consider place extremely strong restrictions on the choice data that could satisfy them. Thus, should a DM satisfy such a mixture model, this too is informative as the theories are so restrictive.

Our approach moves beyond simply testing goodness-of-fit for such mixture theories. Using Bayesian order-constrained techniques, we put both classes of mixture theories in direct competition with one another for each set of individual choice data. This allows us to determine whether preferences consistent with weak orders or lexicographic semiorders, or neither, best account for a DM's choices.

Our statistical methodology, coupled with the parsimony of the mixture model theories and richness of the ternary choice paradigm, afforded excellent model distinguishability in our analyses. The Bayes factors between the weak order (i.e., numerical utility) 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 empirically distinguishable from one another.

Our re-analysis of the data from Regenwetter and Davis-Stober (2012) showed that nearly all DMs in this study are well-described by either the WOMM or an LSMM. The lone DM who violated both models did so in only one stimulus condition. This result is notable in and of itself because of the fact that the models are so parsimonious (as noted before, DMs choosing randomly would violate either model more than 99.9% of the time). Overall, roughly 80% are best described as having weak-ordered preferences (i.e., a numerical utility representation). The 20% who are best described as having lexicographic semi-ordered preferences includes all of the participants who were found to violate transitivity in the original, classical analysis, plus some who were found to be consistent with the WOMM in the original analysis. Thus, even when a DM does not violate transitivity from a classical statistical perspective, the LSMM can still give a more parsimonious, generalizable explanation of their preferences than the WOMM. This result highlights the fact that the LSMM is not solely a model of intransitivity; both transitive and intransitive preferences can be consistent with it.

We were able to replicate these results in a larger, follow-up study, which also included various manipulations to test the robustness of preference structures within-participants. In general, we found that preference structure within-participants was very stable. In other words, the vast majority of participants did not 'switch' in their best-fitting mixture model across different experimental conditions. We found that direct time pressure did not induce more participants to make choices consistent with lexicographic semiorders, despite lexicographic semiorders being less cognitively demanding from a process perspective. The time pressure manipulation also failed to induce individual participants to switch from one model to another. A possible explanation of the time pressure result is that the choice alternatives we used had only two attributes, in contrast to the more complex stimuli of Rieskamp and Hoffrage (1999).

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, 2014), 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, in agreement with the conclusions of Hey (2005) and Loomes et al. (2002).

Our analyses operate at two levels. The mixture models, analyzed via Bayes Factors, are defined on individual, within-subject data, while the experimental manipulation analyses are between-subject in nature, e.g., differences with respect to time pressure and display format. Future work could include the development of a multi-level Bayesian model to unify these within and between-subject data analyses. This type of modeling carries several conceptual benefits over classical methods, see Shiffrin, Lee, Kim, and Wagenmakers (2008) for an overview and introduction to the topic.

There is evidence in the literature that individual differences may also extend to stochastic specification (Cavagnaro & Davis-Stober, 2014). 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 the 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 is possible, but would need to be defined and extended to the ternary choice framework. Future work could explore evaluating alternative stochastic specifications such as "error" specifications where an individual has a single preference state but, with a given probability, makes choices inconsistent with this preference state (e.g. Harless & Camerer, 1994). Other possible specifications include "distance-based" and combinations of mixture and error, — see Regenwetter et al. (2014) for a more complete discussion of alternative stochastic specifications.

Acknowledgments

This work was supported by a University of Missouri Research Board Grant ("Empirical Tests of Decision Models Based on Lexicographic Semiorders", Davis-Stober, PI). Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the University of Missouri or the California State University at Fullerton. We would like to thank Katelin Coonce, Jacob Nicholson, and Amanda Cleaver for their help in data collection. We would also like to thank Michael Lee for feedback on the manuscript as well as an anonymous reviewer. Any errors or inconsistencies are the sole responsibility of the authors.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmp.2014.12.003.

References

Armstrong, W. E. (1939). The determinateness of the utility function. *Economics Journal*, *49*, 453–467.

Ben-Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica*, *47*, 89–104.

Bernardo, J. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences*, *4*, 111–122.

Birnbaum, M. H. (2008). Evaluation of the priority heuristic as a descriptive model of risky decision making: comment on Brandstätter, Gigerenzer, and Hertwig. *Psychological Review*, *115*, 253–262.

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

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: comments on Regenwetter et al. (2011). *Psychological Review*, *118*, 675–683.

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.

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

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

Böckenholt, U., & Kroeger, K. (1993). The effect of time pressure in multiattribute binary choice tasks. In O. Svenson, & A. J. Maule (Eds.), *Time pressure and stress in human judgment and decision making* (pp. 195–214). New York: Plenum Press.

Brandstätter, E. (2011). On the stability of choice processes. *Frontiers in Psychology*, *2*, http://dx.doi.org/10.3389/fpsyg.2011.00295.

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

Cavagnaro, D. R., & Davis-Stober, C. P. (2014). Transitive in our preferences but transitive in different ways: an analysis of choice variability. *Decision*, *1*, 102–122.

Christof, T., & Löbel, A. (1997). Porta: polyhedron representation transformation algorithm [Software]. Retrieved from http://typo.zib.de/opt-long_projects/Software/Porta/.

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

Davis-Stober, C. P. (2010). A bijection between a set of lexicographic semiorders and pairs of non-crossing Dyck paths. *Journal of Mathematical Psychology*, *54*, 471–474.

Davis-Stober, C. P. (2012). A lexicographic semiorder polytope and probabilistic representations of choice. *Journal of Mathematical Psychology*, *56*, 86–94.

Davis-Stober, C. P., & Brown, N. (2011). A shift in strategy or error? strategy classification over multipole stochastic specifications. *Judgment and Decision Making*, *6*, 800–813.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140.

Fiorini, S. (2001). *Polyhedral combinatorics of order polytopes*. (Doctoral dissertation), Brussels, Belgium: Université Libre de Bruxelles.

Fiorini, S., & Fishburn, P. C. (2004). Weak order polytopes. *Discrete Mathematics*, *275*, 111–127.

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

Georgescu-Roegen, N. (1936). The pure theory of consumer's behavior. *Quarterly Journal of Economics*, *50*, 545–593.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*, 650–669.

Glöckner, A., & 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*, 75–95.

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

Grünwald, P. D. (2007). *The minimum description length principle*. MIT Press.

Harless, D. W. (1992). Actions versus prospects: the effect of problem representation on regret. *American Economic Review*, *82*, 634–649.

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

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

Heyer, D., & Niederée, R. (1992). Generalizing the concept of binary choice systems induced by rankings: one way of probabilizing deterministic measurement structures. *Mathematical Social Sciences*, *23*, 31–44.

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

Jackman, S. (2009). *Bayesian analysis for the social sciences*. Hoboken, NJ: Wiley.

Jeffreys, H. (1998). *Theory of probability*. Oxford University Press.

Johnson, E. J., Payne, J. W., & Bettman, J. R. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes*, *42*, 1–21.

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

Kleinmuntz, D. N., & Schkade, D. A. (1993). Information displays and decision processes. *Psychological Science*, *4*, 221–227.

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

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, *11*, 343–352.

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. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 178–191.

Luce, R. D. (2000). *Utility of gains and losses: measurement-theoretical and experimental approaches*. Mahwah, N. J.: Erlbaum.

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

Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review*, *101*, 505–518.
Montgomery, H. (1977). A study of intransitive preferences using a think aloud procedure. In H. Jungerman, & G. de Zeeuw (Eds.), *Decision-making and change in human affairs* (pp. 347–362). Dordrecht, the Netherlands: Reidel.
Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190–204.
Myung, J. I., Karabatsos, G., & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, *49*, 205–225.
Myung, J. I., Navarro, D. J., & Pitt, M. A. (2006). Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology*, *50*, 167–179.
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.
Pirlot, M., & Vincke, P. (1997). *Semiorders: properties, representations, applications*. Dordrecht: Kluwer.
Ranyard, R. H. (1977). Risky decisions which violate transitivity and double cancellation. *Acta Psychologica*, *41*, 449–459.
Ranyard, R. H. (1982). Binary choice patterns and reasons given for simple risky choice. *Acta Psychologica*, *52*, 125–135.
Regenwetter, M. (1996). Random utility representations of finite m-ary relations. *Journal of Mathematical Psychology*, *40*, 219–234.
Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preference. *Psychological Review*, *118*, 42–56.
Regenwetter, M., Dana, J., Davis-Stober, C. P., & Guo, Y. (2011). Parsimonious testing of transitive or intransitive preferences: reply to Birnbaum (2011). *Psychological Review*, *118*, 684–688.
Regenwetter, M., & Davis-Stober, C. P. (2012). Choice variability versus structural inconsistency of preferences. *Psychological Review*, *119*, 408–416.
Regenwetter, M., Davis-Stober, C. P., Lim, S.-H., Guo, Y., Popova, A., Zwilling, C., et al. (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.
Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics and how can we tell? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make use smart* (pp. 141–167). New York: Oxford University Press.
Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: how opportunity costs affect strategy selection. *Acta Psychologica*, *127*, 258–276.
Russo, J. E. (1977). The value of unit price information. *Journal of Marketing*, *14*, 193–201.
Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.
Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: inequality, order, and shape restrictions*. New York: John Wiley & Sons.
Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty*, *6*, 235–254.
Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48.
Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, *79*, 281–299.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.
von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior*. Princeton University Press.
Ziegler, G. M. (1995). *Lectures on polytopes*. Berlin, Germany: Springer-Verlag.