THEORETICAL NOTE

Behavioral Variability of Choices Versus Structural Inconsistency of Preferences

Michel Regenwetter
University of Illinois at Urbana–Champaign

Clintin P. Davis-Stober
University of Missouri

Theories of rational choice often make the structural consistency assumption that every decision maker's binary strict preference among choice alternatives forms a *strict weak order*. Likewise, the very concept of a *utility function* over lotteries in normative, prescriptive, and descriptive theory is mathematically equivalent to strict weak order preferences over those lotteries, while intransitive heuristic models violate such weak orders. Using new quantitative interdisciplinary methodologies, we dissociate the variability of choices from the structural inconsistency of preferences. We show that laboratory choice behavior among stimuli of a classical "intransitivity" paradigm is, in fact, consistent with variable strict weak order preferences. We find that decision makers act in accordance with a restrictive mathematical model that, for the behavioral sciences, is extraordinarily parsimonious. Our findings suggest that the best place to invest future behavioral decision research is not in the development of new *intransitive* decision models but rather in the specification of parsimonious models consistent with strict weak order(s), as well as heuristics and other process models that explain why preferences appear to be weakly ordered.

Keywords: rationality, strict weak order, utility of uncertain prospects, random utility

Supplemental materials: http://dx.doi.org/10.1037/a0027372.supp

Bob and Joe meet for three meals. Restaurants A, B, and C are compatible with Joe's diet. On Monday, C is closed, and Bob prefers A to B. On Tuesday, A is closed, and Bob would rather eat at B than eat at C. On Friday, Bob is indifferent between A and C, but he asks not to go to B. Joe expected Bob to like A the best and C the least. Is Bob inconsistent?

On Monday, Bob craves a sandwich, for which A outshines B. On Tuesday, he wants a soup, with B being an easy choice. On Friday, Bob craves a salad, which B does not offer, while A and C have salad bars. Even though Bob first chooses A over B, then B over C, then is indifferent between A and C, and prefers both to B, this cannot be interpreted to mean that Bob makes incoherent decisions. Rather, Bob compares the restaurants by quality of his desired entree, and the resulting rankings (with possible ties) of the restaurants vary as he seeks out different entrees. We need to dissociate behavioral variability from structural inconsistency.

The literature on individual decision making is internally conflicted over whether preferences can be modeled with unidimensional numerical utility representations and, in particular, whether variable choice behavior can be reconciled with transitive preferences. Even a leading thinker in the field, Amos Tversky, was the author of both the most influential intransitivity study (Tversky, 1969) and, later, the most influential (transitive) numerical representation, (cumulative) prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Regenwetter, Dana, and Davis-Stober (2011) questioned the literature on intransitivity of preferences and reported that the variable binary choice data they considered were generally consistent with variable (transitive) linear order preferences. In this article, we develop a much more direct, general, and yet also powerful test of unidimensional utility representations and associated (transitive) "weak order" preferences. Even though our experimental stimuli have a historical track record of allegedly revealing inconsistent preferences, our findings support the highly restrictive class of numerical utility theories.

Consider the two data sets at the top of Figure 1. These show the number of times, out of 45 repetitions, that decision makers chose a row gamble over a column gamble. For instance, Participant 9 (Gamble Set III of our experiment) chose gamble *a* over gamble *b* on 28 trials, chose *b* over *a* on eight trials, and expressed indif-

REAL DATA:

Participant # 9 (Set III)						Participant # 22 (Set II)					
Observed # of times row chosen over column						Observed # of times row chosen over column					
	a	b	c	d	e		a	b	c	d	e
a	-	28	35	41	43	a	-	43	42	34	29
b	8	-	21	29	40	b	1	-	44	30	20
c	10	18	-	35	41	c	1	0	-	30	23
d	4	11	5	-	39	d	1	8	5	-	17
e	2	4	3	4	-	e	5	12	9	6	-

The choice proportions satisfy all 75,834 FDIs.	The model is violated significantly ($p \leq 0.001$).

FACET-DEFINING INEQUALITY EXAMPLE:
(Real Choice Proportions of Participant # 9, Set III)

FDI in (8) holds: $\begin{cases} 3\left(\frac{28}{45} - \frac{8}{45} + \frac{11}{45} - \frac{29}{45} + \frac{41}{45} + \frac{3}{45}\right) + \frac{35}{45} + \frac{10}{45} + \frac{21}{45} - \frac{18}{45} - \frac{39}{45} + \frac{4}{45} \\ +3\left(\frac{40}{45}\right) - \frac{4}{45} - 3\left(\frac{43}{45}\right) + \frac{2}{45} - 3\left(\frac{4}{45}\right) - \frac{41}{45} - 3\left(\frac{5}{45}\right) - \frac{35}{45} = \frac{37}{45} \leq \frac{180}{45} = 4. \end{cases}$

HOW DOES THE MODEL GENERATE DATA?
(Hypothetical Example)

Random Utility					Probability	Mixture (Random Preference) Strict Weak Order $\succ$
U(a)	U(b)	U(c)	U(d)	U(e)		
100	50	10	5	0	0.15 $\left.\begin{matrix} \\ \end{matrix}\right\} \leftrightarrow 0.4$	$a \succ b \succ c \succ d \succ e$
20	10	5	0	−5	0.25	
50	50	10	5	0	$0.2 \leftrightarrow 0.2$	$a \sim b \succ c \succ d \succ e$
-5	0	5	100	20	$0.3 \leftrightarrow 0.3$	$d \succ e \succ c \succ b \succ a$
105	5	15	30	5	$0.1 \leftrightarrow 0.1$	$a \succ d \succ c \succ b \sim e$

By Equations (5) and (6):

Probability P_{ij} that row i is chosen over column j							Expected # of times row is chosen over column				
P_{ij}	a	b	c	d	e			a	b	c	d	e
a	-	0.5	0.7	0.7	0.7	$\Rightarrow$	a	-	22.5	31.5	31.5	31.5
b	0.3	-	0.6	0.6	0.6		b	13.5	-	27	27	27
c	0.3	0.4	-	0.6	0.7		c	13.5	18	-	27	31.5
d	0.3	0.4	0.4	-	1.0		d	13.5	18	18	-	45
e	0.3	0.3	0.3	0	-		e	13.5	13.5	13.5	0	-

Figure 1. Illustration with real and hypothetical ternary paired comparison data. The top gives two real ternary paired comparison data sets. The model fits one set perfectly and is violated significantly by the other. We illustrate one facet-defining inequality (FDI) for Participant 9 (Set III). The center panel illustrates a hypothetical random utility representation and the corresponding mixture. The joint distribution of the random utilities is on the left, the weak order probabilities are on the right. Below are the implied binary choice probabilities and the corresponding expected frequencies for ternary choice (45 repetitions).

ference among *a* and *b* on nine trials. The primary goal in this article is to disentangle such variable overt choices from inconsistent latent preferences. We show that a restrictive quantitative model of variable, but structurally consistent, decision making explains a large set of empirical data. According to this model, individual preferences form *strict weak orders,* that is, asymmetric and negatively transitive binary relations, as defined in Formulas 2 and 3 below. We consider the class of all possible decision theories that share the structural property that decision makers, knowingly or unknowingly, rank-order choice options, allowing the possibility that they are indifferent between some options. Should this

structural property be violated over those lotteries, then *all* theories of (unidimensional) utility would be violated, including expected utility theory and its many offspring, such as (cumulative) prospect theory. Should preferences be strict weak orders, however, then an even larger class of binary relations would be rejected, including *all* theories with *intransitive* preferences, such as intransitive heuristics and regret theory (Loomes & Sugden, 1982). Our approach is unique in that we found a way to test the pivotal concept of weak order preferences directly, hence to test the very notion of a utility function directly without limiting ourselves to particular functional forms (e.g., "expected utility" with "power" utility for money) or

even to cash stimuli. Returning to Figure 1, we later show that Participant 9 (Set III) is not just approximately, but perfectly, consistent with variable weak order preferences, whereas Participant 22 (Set II), whose data look superficially similar, in fact, violates the model. Throughout the article, we use the term *preference* to refer to a hypothetical construct that might not be directly observable and the term *choice* to refer to the observed behavior of participants when asked to select an option.

Motivation and Background

Why is the concept of rank ordering choice alternatives such a central and pivotal construct in psychological and economic theories of decision making? Many laboratory and real life decisions share a common structure: They typically involve various outcomes that are contingent on uncertain states of the world. Decision making is often complicated by intricate trade-offs among different attributes, outcomes, and uncertainty or risk. Psychological experiments in the laboratory often use cash lotteries as choice alternatives. In addition to such cash lotteries, we offered our participants noncash options similar to the following prospects A and C.

$Prospect\ A$: $\begin{cases} \text{Receive 40 movie rentals} & \text{with a probability of .26} \\ \text{Receive or lose nothing} & \text{otherwise.} \end{cases}$

$Prospect\ C$: $\begin{cases} \text{Receive 7 books} & \text{with a probability of .22} \\ \text{Receive or lose nothing} & \text{otherwise.} \end{cases}$

A major class of decision theories hypothesizes a numerical value $U(i)$ associated with each prospect i. These numerical theories state that a decision maker prefers prospect i to prospect j if she or he values i more highly than j, that is, if $U(i) > U(j)$. Writing $>$ to denote strict pairwise preference, such theories predict, for example,

$$A > C \Leftrightarrow U(A) > U(C). \qquad (1)$$

There is a substantial literature in the decision sciences, primarily for monetary gambles, that discusses the functional form of the numerical function $U(\cdot)$. Famous examples of such theories are expected utility theory (Savage, 1954; von Neumann & Morgenstern, 1947) or (cumulative) prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). All of these theories share one very restrictive property: They model decision behavior through the hypothetical numerical construct of a utility value associated with each uncertain prospect. These theories assume that decision makers act in accordance with Formula 1. Even though the utilities in Formula 1 are numerical, the prospects themselves need not be. For example, Prospects A and C are nonmonetary and, in our experiment, their probabilities are displayed geometrically using pie charts depicting wheels of chance. Any theory in which pairwise preferences are derived from hypothetical numerical values, as illustrated in Formula 1, implies that binary preferences $>$ form strict weak orders. It has been shown that, likewise, strict weak order preferences can always be represented numerically, whether the decision makers mentally compute numerical utilities or not (see, e.g., F. S. Roberts, 1979).

This answers our question: All theories that rely on the hypothetical construct of a utility function as in Formula 1 imply that preference is a strict weak order, and all theories that model

preference using a strict weak order can be recast mathematically as Formula 1 via a function $U(\cdot)$.

Hence, by considering the class of all theories in which preferences form strict weak orders, we study the pivotal property that distinguishes preferences that are structurally consistent with utility functions from those that are inconsistent with utility. To study the structural property of strict weak orders, we must disentangle varying choice (as in Figure 1) from structurally inconsistent preference. We start by considering basic concepts.

We are interested in the structural constraint that binary strict preference must form *a strict weak order*. This means that (a) strict preference $>$ is an *asymmetric binary relation:* If a person strictly prefers i to j then she or he does not strictly prefer j to i:

$$\text{For all } i \text{ and } j: [i > j] \Rightarrow [j \not> i], \qquad (2)$$

and that (b) strict preference $>$ is *negatively transitive:* If a person does not strictly prefer i to j and does not strictly prefer j to k then she or he does not strictly prefer i to k:

$$\text{For all } i, j \text{ and } k: [i \not> j] \wedge [j \not> k] \Rightarrow [i \not> k]. \qquad (3)$$

As a consequence of these two axioms, strict weak orders also satisfy *transitivity:*[1] If a participant strictly prefers i to j and strictly prefers j to k then she or he strictly prefers i to k:

$$\text{For all } i, j, \text{ and } k: [i > j] \wedge [j > k] \Rightarrow [i > k]. \qquad (4)$$

Transitivity is often viewed as a rationality axiom. It is satisfied by the vast majority of normative, prescriptive, and descriptive theories of decision making in the literature.

Asymmetry strikes us as a property that we may safely assume to hold inherently for strict preference. Hence, we do not test it empirically. We consider stimulus sets of five choice alternatives in our experiment. For five alternatives, there are $3^{\binom{5}{2}} = 59,049$ different, possible asymmetric binary relations. Of these, only 541, that is fewer than 1%, are strict weak orders (Fiorini, 2001). Hence, fewer than 1% of all conceivable strict preference relations on five choice options can be represented with a numerical function $U(\cdot)$. The remaining 99% feature violations of negative transitivity and/or violations of transitivity.

It is already clear, even from a purely algebraic point of view, that strict weak orders form quite a restrictive class of preference relations. We later show, once we move to our probabilistic specification of strict weak orders, that our quantitative model is characterized by more than 75,000 nonredundant simultaneous mathematical constraints on choice probabilities. Our model also allows us to separate the wheat from the chaff by distinguishing variability in observable choices from inconsistency of underlying preferences.

The Mixture Model of Strict Weak Order Preferences

We propose a *strict weak order* model for *ternary paired comparisons.* A key distinguishing feature of our empirical paradigm is that decision makers compare pairs of objects (e.g., gambles) and either

---

[1] Note, however, that the converse does not hold. For example, for five choice alternatives, there are 541 strict weak orders (Fiorini, 2001), whereas there are 154,303 different transitive relations (Pfeiffer, 2004).

choose one of the objects as their "preferred" option or indicate that they are "indifferent" among the offered pair of objects. Each decision maker repeats each such ternary paired comparison multiple times over the course of an experiment. All repetitions are carefully separated by decoys in order to avoid memory effects and to reduce statistical interdependencies among repeated observations for each binary choice. For five choice alternatives, this paradigm generates 20 degrees of freedom in the data: each pair $i, j \in \mathcal{C}$ yields some number of i choices, some number of j choices, and some number of indifference statements, hence yielding 2 degrees of freedom for each of the 10 gamble pairs. In comparison, a two-alternative forced choice paradigm for the same stimuli has only 10 degrees of freedom in the data because it does not permit indifference responses; hence, each of the 10 gamble pairs only has 1 degree of freedom associated with it. In other words, even though we only marginally increase the amount of information required from the respondent on each trial (Miller, 1956), we double the degrees of freedom and thereby dramatically increase the total amount of information extracted from the respondent over the entire experiment as we combine multiple stimuli.[2]

We are interested in the structural property of strict weak order preference. We also need to separate variable overt choice behavior from structurally inconsistent latent preferences. We achieve these goals by evaluating the goodness-of-fit of a probabilistic model. According to our model, whenever facing a pair of choice alternatives, a decision maker evaluates these choice alternatives in a fashion that induces a strict weak order on the set of choice alternatives. Over repeated ternary paired comparisons, the decision maker may fluctuate in the strict weak order that he or she uses in the various decisions. This is either because he or she varies in his or her preferences over time or because he or she experiences uncertainty about his or her own preferences and, when asked to decide, ends up fluctuating in those forced choices. We introduce a probability space where every sample point must be a strict weak order. We deliberately avoid peripheral mathematical assumptions because all we do is take the concept of weak order preferences and place it into a probability space that has no additional structure except for requiring every sample point to be a strict weak order.

Equivalently, we are interested in the structural property of preferences that are consistent with numerical utilities. In the face of variable choice behavior, we need to distinguish whether variable choices are consistent with variable numerical utilities, or not. According to our model, when considering a pair of choice alternatives, a decision maker chooses in accordance with a numerical function $U(\cdot)$ as specified in Formula 1. Over repeated ternary paired comparisons, the value of $U(c)$ of a choice option c may vary. Accordingly, we replace the function $U(\cdot)$ with a family of random variables, $(\mathbf{U}_c)_{c \in \mathcal{C}}$, whose joint distribution is not restricted in any way. We allow decision makers to waver in the numerical utility values that they associate with the choice options. Again, we deliberately avoid peripheral mathematical assumptions because all we do is take the concept of a numerical function $U(\cdot)$ and turn the numerical value $U(c)$ of each prospect c into a random variable that has no additional constraints because we allow the random variables to have any distribution whatsoever.

At first blush, our weak order model appears extremely flexible: For five choice options, it permits 541 different preference states, that is, it has 540 free parameters. Its formulation in terms of

arbitrary random variables, $(\mathbf{U}_c)_{c \in \mathcal{C}}$, has no limit on free parameters since we explicitly forego constraints on the joint distribution. The empirical data only have 20 degrees of freedom. Do we engage in overfitting? Appearances are misleading. First, we have already seen that strict weak orders are rather restrictive compared with asymmetric relations. Even more important is the statistical perspective. Using Monte Carlo methods, we have computed the approximate relative volume of the parameter space within the outcome space (both are of the same dimension). It turns out that the model is, in fact, highly restrictive: For five choice alternatives, the permissible ternary paired comparison probabilities of the strict weak order model occupy only one two-thousandth of the empirical outcome space. It is crucial to distinguish statistical identifiability from statistical testability. While the 540 free parameters are not uniquely identifiable when there are only 20 degrees of freedom in the data, the model is testable because it makes extraordinarily restrictive predictions. The strict weak order model, stated formally in Equation 5, can test the structural property of strict weak order preferences in the presence of variable choice behavior.

Denote by $\mathcal{WO}_\mathcal{C}$ the collection of all strict weak orders on a given finite set $\mathcal{C}$ of choice alternatives. A collection $(P_{ij})_{i,j \in \mathcal{C}}$ is called a *system of ternary paired comparison probabilities*; if $\forall i, j \in \mathcal{C}$, with $i \neq j$, we have $0 \leq P_{ij} \leq 1$ and $P_{ij} + P_{ji} \leq 1$. This system satisfies the *strict weak order model* if there exists a probability distribution on $\mathcal{WO}_\mathcal{C}$,

$$P: \mathcal{WO}_\mathcal{C} \rightarrow [0, 1]$$

$$> \mapsto P_>,$$

that assigns probability $P_>$ to any strict weak order $>$, such that $\forall i, j \in \mathcal{C}, i \neq j$,

$$P_{ij} = \sum_{\substack{> \in \mathcal{WO}_\mathcal{C} \\ i > j}} P_>. \tag{5}$$

In words, this is as follows: The probability that the respondent chooses i, when offered i versus j in a ternary paired comparison trial, is the (marginal) total probability of all those strict weak orders $>$ in which i is strictly preferred to j.

In the terminology of Loomes and Sugden (1995), Equation 5 is a random preference model whose core theory is the strict weak order. It can be rewritten as a "distribution-free random utility model" in which each sample point is a strict weak order (Heyer & Niederée, 1992; Niederée & Heyer, 1997; Regenwetter, 1996; Regenwetter & Marley, 2001).

A system of ternary paired comparison probabilities satisfies a *distribution-free random utility model* if there exists a probability measure P and there exist real-valued jointly distributed random variables $(\mathbf{U}_c)_{c \in \mathcal{C}}$ such that $\forall i, j \in \mathcal{C}, i \neq j$,

$$P_{ij} = P(\mathbf{U}_i > \mathbf{U}_j). \tag{6}$$

In words, this is as follows: The probability that the respondent chooses i, when offered i versus j in a ternary paired comparison

---

[2] Similar ideas have been used in signal detection theory with multiple response categories (Macmillan & Creelman, 2005).

trial, is the (marginal) total probability that the random utility of i exceeds the random utility of j.[3]

To see of how all of these ideas come together and how the model generates predicted ternary paired comparison data, consider Figure 1. The box titled Random Utility shows the values of five hypothetical utility functions as well as the probability of each function. For example, the utility function in the first row has a probability of .15 and assigns utilities 100, 50, . . . 0 to $a, b, \ldots e$. Each utility function induces a strict weak ordering of the alternatives. For example, the first two functions induce the weak order $a > b > c > d > e$; hence, that strict weak order has a probability of $.15 + .25 = .40$. Let $i \sim j$ denote that both $i \nsucc j$ and $j \nsucc i$. To illustrate the mixture model in Equation 5 consider the weak orders in which $a > b$, namely, $a > b > c > d > e$ and $a > d > c > b \sim e$, whose probabilities sum to .5, hence, the entry $P_{ab} = 0.5$ in the bottom left matrix. Notice that $P_{ab} + P_{ba} = .5 + .3 = .8 \neq 1$ because the weak order $a \sim b > c > d > e$ has a positive probability of .2.

To illustrate the random utility model in Equation 6, notice that P_{ab} is also given by summing the probabilities $.15 + .25 + .1$ of the utility functions with $U(a) > U(b)$. We can now establish that the models in Equations 5 and 6 are mathematically equivalent; hence, we can later focus on the formulation in Equation 5. The following theorem (proven in the online supplement) follows directly from related prior work (Heyer & Niederée, 1992; Niederée & Heyer, 1997; Regenwetter, 1996; Regenwetter & Marley, 2001). This theorem is extremely important: It shows that any algebraic theory that uses a numerical function $U(\cdot)$, say cumulative prospect theory, can be extended quite naturally into a probabilistic model that induces a probability distribution over strict weak orders.

> *Theorem: A system of ternary paired comparison probabilities P_{ij} satisfies the weak order model in Equation 5 if and only if it satisfies the distribution-free random utility model in Equation 6.*

Characterizing models like Equation 5 or Equation 6 is a nontrivial task that has received much attention, for example, in mathematics, operations research and mathematical psychology. The most promising approach tackles the problem by characterizing the weak order model stated in Equation 5 with tools from convex geometry. In that terminology, the collection of all ternary paired comparison probabilities P_{ij} consistent with the model in Equation 5 forms a geometric object called a "convex polytope" in a vector space whose coordinates are the probabilities P_{ij}. A *convex polytope* is a high-dimensional generalization of (two-dimensional) polygons and of (three- dimensional) polyhedra.[4] We can test the weak order model if we understand its polytope.

A triangular area in two-dimensional space (in a plane) can be described fully by the area between the three points that form its corners (so-called *vertices*) or it can be described equivalently as the area between its three sides (so-called *facets*). This logic extends to high dimensional spaces. A convex polytope can likewise be described in two equivalent ways: One can describe it as either the "convex hull of a set of vertices" or as the solution to a system of "affine inequalities" (Ziegler, 1995) that define its facets (for examples, see $WO_1 - WO_3$, below). The model formulation in Equation 5 directly yields the *vertex representation* as the convex hull of a collection of vertices: For each strict weak order $>$ in $\mathcal{WO_C}$, create a *vertex* with the following 0/1-coordinates:

$$P_{ij} = \begin{cases} 1 & \text{if } i > j, \\ 0 & \text{if } i \nsucc j. \end{cases} \quad (7)$$

Each vertex of the polytope represents one strict weak order. Intuitively, if a person has deterministic preferences in the form of one single weak order, then this person's choice probabilities are given by the vertex coordinates of that weak order: If the person always prefers i over j then, according to the model in Equation 5, she chooses i over j with probability 1. If a person never prefers i to j, then, according to the model in Equation 5, she chooses i over j with a probability of 0. While each vertex can be thought of as a degenerate distribution that places all probability mass on a single strict weak order, the entire polytope forms a geometric representation of all possible probability distributions over strict weak orders. That is precisely the model stated in Equation 5. The online supplement illustrates these geometric concepts with a simple two-dimensional example and three illustrative figures.

To test the model, we wish to determine whether a set of data was generated by a vector of ternary paired comparison probabilities that belongs to the polytope. To achieve this goal, it is useful to find the alternative representation of the polytope, that is, to find a system of inequalities whose solution set is the polytope. We can then test whether the ternary paired comparison probabilities satisfy or violate those "facet-defining inequalities" (FDIs).

When $|\mathcal{C}| = 3$, say, $\mathcal{C} = \{a, b, c\}$, the polytope is known to have 13 vertices, representing the 13 different strict weak orders over three objects. These vertices "live" in a six-dimensional vector space spanned by the probabilities P_{ab}, P_{ac}, P_{ba}, P_{bc}, P_{ca}, and P_{cb}. The following three families of inequalities completely characterize the strict weak order polytope when $|\mathcal{C}| = 3$:

---

[3] Notice that we do *not* require *noncoincidence,* that is, we do not require $P(\mathbf{U}_i = \mathbf{U}_j) = 0$ for $(i \neq j)$. Random utility models in econometrics often satisfy noncoincidence because they rely on, say, multivariate normal or extreme value distributions. Such distributional assumptions are known idealizations that serve the purpose of making modeling and statistical estimation tractable. One can show that noncoincidence effectively forces "completeness" of binary preferences because indifference statements have probability zero, according to noncoincidence; hence, incomplete relations $>$ have zero probability in Equation 5. This is precisely the assumption that we do not want to make when testing weak order preferences and/or testing numerical utility. The most interesting applications of these models are when $P(\mathbf{U}_i = \mathbf{U}_j) > 0$; thus, $P_{ij} + P_{ji} < 1$.

We also do *not* assume that the random variables $\mathbf{U}$ are mutually independent, as do some Thurstonian models in psychology. Because we strive to avoid technical convenience assumptions, especially distributional assumptions that would imply noncoincidence or independence assumptions in Equation 6, our approach differs from standard off-the-shelf measurement models as well as signal-detection models that relate subjective percepts to an objective ground truth (Macmillan & Creelman, 2005). In our data analysis, we do require that the data result from an independent and identically distributed (iid) random sample, due to a lack of statistical tools for non-iid samples at this time. However, the iid sampling assumption neither assumes nor implies independence among the random variables $(\mathbf{U}_c)_{c \in \mathcal{C}}$ in the model for Equation 6.

[4] Triangles, parallelograms, and so on are examples of convex polygons in two space; cubes, pyramids, and so on are examples of convex polyhedra in three space.

WO_1: $P_{ij} \geq 0$, for all distinct i, j in $\mathcal{C}$;

WO_2: $P_{ij} + P_{ji} \leq 1$, for all distinct i, j in $\mathcal{C}$;

WO_3: $P_{ij} + P_{jk} - P_{jk} \leq 1$ for all distinct i, j, k in $\mathcal{C}$.

WO_1 states that probabilities cannot be negative. WO_2 reflects the empirical paradigm: The probabilities of choosing either i or j may possibly sum to less than one because the decision maker may use the indifference option with positive probability. Constraints WO_1 and WO_2 automatically hold in a ternary paired comparison paradigm and are thus not subject to testing, whereas WO_3 is testable. Convex geometry allows one to prove that WO_1 through WO_3 are nonredundant and that they fully characterize the family of all conceivable ternary paired comparison probabilities for $|\mathcal{C}| = 3$ that can be represented using Equation 5. This means that the constraints WO_1 through WO_3 form a *minimal* description, that is, the shortest possible list of nonredundant constraints that fully describe the strict weak order model for $|\mathcal{C}| = 3$ (Fiorini & Fishburn, 2004). Such constraints are called *facet-defining* inequalities. Each of these inequalities defines a facet of the polytope, that is, a face of maximal dimension.[5]

Obtaining a minimal description of convex polytopes can be computationally prohibitive. Our experiment used three stimulus sets of five alternatives each. The minimal description of the polytope in terms of facet-defining inequalities for the case of interest to us, where $|\mathcal{C}| = 5$, was previously unknown. We determined this description using public domain software, PORTA, the POlyhedron Representation Transformation Algorithm, by Christof and Löbel (1997).[6] That strict weak order polytope has 541 vertices in a vector space of dimension 20. This polytope occupies one two-thousandth of the space of trinomial probabilities. It is characterized by 75,834 facet-defining inequalities (see the online supplement).[7] The following is one of these 75,834 constraints. It interrelates all 20 choice probabilities for $\mathcal{C} = \{a, b, c, d, e\}$.

$$3(P_{ab} - P_{ba} + P_{db} - P_{bd} + P_{ce} + P_{ec}) + P_{ac} + P_{ca} + P_{bc} - P_{cb}$$
$$- P_{de} + P_{ed} + 3P_{be} - P_{eb} - 3P_{ae} + P_{ea} - 3P_{da} - P_{ad} - 3P_{dc}$$
$$- P_{cd} \leq 4. \quad (8)$$

How can one test thousands of joint constraints like Inequality 8? First, it is possible that the data fall inside the tiny polytope, namely, if the observed choice proportions themselves satisfy all 75,000+ constraints (see Figure 1). When they do, the observed choice frequencies are precisely within the range of predicted values, and we do not need to compute statistical goodness-of-fit.[8] When the observed choice proportions fall outside the polytope, then writing N_{ij} for the observed frequency with which a person chose object i when offered the ternary paired comparison among i and j altogether K many times, and assuming independent and identically distributed (iid) sampling, the likelihood function is given by

$$\prod_{\substack{i, j \in \mathcal{C} \\ i \neq j}} \frac{K!}{N_{ij}! \, N_{ji}! (K - N_{ij} - N_{ji})!} P_{ij}^{N_{ij}} P_{ji}^{N_{ji}} (1 - P_{ij} - P_{ji})^{K - N_{ij} - N_{ji}}.$$

It can be shown that the log-likelihood ratio test statistic, G^2, need not have an asymptotic chi-square distribution because the

model in Equation 5 imposes inequality constraints on the probabilities P_{ij} and P_{ji}. The asymptotic distribution of G^2 is a mixture of chi-square distributions, with the mixture weights depending on the structure of the polytope in the neighborhood around the maximum likelihood point estimate (Davis-Stober, 2009; Myung, Karabatsos, & Iverson, 2005; Silvapulle & Sen, 2005). Depending on the location of the point estimate on the weak order polytope, we need to employ some of the 75,000+ facet-defining inequalities to derive a suitable mixture of chi-square distributions. Next, we use Davis-Stober's (2009) methods, implemented in a Matlab (Version 7.0) program, to test the model on new data.

Experiment and Data Analysis

Our experimental study built on Tversky (1969) and used stimuli similar to Regenwetter et al. (2011). The main change was that decisions were ternary paired comparisons. We investigated the behavior of 30 participants, and we secured high statistical power by collecting 45 trials per gamble pair and per respondent. The gambles were organized into four sets: Gamble Set I, Gamble Set II, Gamble Set III, and Distractors. We provide details of the experiment, including recruitment, stimuli, and our data, in the online supplement.

Table 1 provides our results for all 30 participants and the three gamble sets. When the ternary paired-comparison proportions are in the interior of the weak order polytope, we indicate this with a checkmark ($\sqrt{}$). Data outside the polytope have positive G^2 values, with large p values indicating statistically nonsignificant violations. Each G^2 needs to be evaluated against a specific "Chi-bar-squared" distribution, thus individual G^2 values cannot be compared across results. Significant violations are marked in boldfaced font.

In Gamble Set I, 22 respondents generated choice proportions that satisfied all 75,000+ (nonredundant) inequality constraints that the model in Equation 5 imposes on the choice probabilities. In other words, 22 of 30 respondents generated a "perfect fit." Given that the parameter space of the model occupies only a tiny fraction of the outcome space, this is a remarkable discovery. One out of 30 participants violated the model at a significance level of 5%. This is well within Type-I error range. In Gamble Set II, which created more difficult trade-offs between probabilities of winning and outcome, 20 of 30 respondents yielded a perfect fit. This is, again, an extraordinary finding. Among the remaining 10 respondents, two significantly violated the constraints of the model

---

[5] For example, a three-dimensional cube has eight square-shaped facets, and a tetrahedron has four triangular facets. The edges of these polyhedra are one-dimensional faces and, hence, not facets.

[6] Available at http://www.iwr.uni-heidelberg.de/groups/comopt/software/PORTA/. Translating the vertex representation into a minimal description took three weeks of continuous computation on a powerful PC.

[7] In particular, it can be shown that Equation 5 again implies the family of so-called "triangle inequalities," labeled WO_3 above. WO_3 has previously surfaced for two-alternative forced choice because it is directly related to transitivity (Block & Marschak, 1960). Here, WO_3 is only one of thousands of testable constraints that characterize the model. Unlike $WO_1 - WO_3$, these constraints are not so straightforward to interpret.

[8] Here, we take a frequentist view. We comment on Bayesian approaches in the Discussion and Conclusion section.

Table 1
Goodness-of-Fit for 30 Individual (ID) Respondents (1–30) and Three Gamble Sets (Set I, Set II, and Set III)

ID	Set I: G^2 (p)	Set II: G^2 (p)	Set III: G^2 (p)	ID	Set I: G^2 (p)	Set II: G^2 (p)	Set III: G^2 (p)
1	√	√	0.23 (.91)	16	√	√	√
2	√	√	2.23 (.55)	17	√	√	√
3	√	1.11 (.77)	2.82 (.32)	18	√	√	4.77 (.27)
4	√	√	10.78 (**.01**)	19	√	√	√
5	√	0.17 (.89)	√	20	√	√	√
6	√	√	√	21	√	√	1.87 (.46)
7	√	√	√	22	5.01 (.10)	14.90 (<**.01**)	3.63 (.42)
8	√	√	√	23	0.16 (.95)	0.04 (.89)	0.54 (.62)
9	0.52 (.82)	5.95 (.18)	√	24	√	√	√
10	2.82 (.47)	√	√	25	√	√	6.42 (.08)
11	√	√	√	26	1.93 (.48)	1.30 (.75)	0.06 (.92)
12	√	√	√	27	√	√	√
13	√	√	√	28	√	√	√
14	8.01 (**.02**)	1.22 (.38)	12.09 (.05)	29	6.04 (.26)	14.81 (**.01**)	√
15	√	0.80 (.40)	0.40 (.84)	30	1.03 (.54)	0.04 (.88)	√

Note. A checkmark (√) indicates a perfect fit. Positive G^2 values are accompanied with ps in parentheses. Significant ps (< .05), that is, significant violations of the weak order model, are marked in bold.

in Equation 5 at a significance level of 5%. Again, two significant violations is within a reasonable Type-I error range. For Gamble Set III, 18 of 30 participants generated choice proportions that satisfied all 75,000+ inequalities. The remaining 12 participants included one significant violation at the 5% level. This is yet again within Type-I error range.

Discussion and Conclusions

A number of descriptive, prescriptive, and normative theories of decision making share the structural assumption that individual preferences are strict weak orders. These include the high-profile class of theories of decision making that rely on numerical representations of preferences, such as (subjective) expected utility and (cumulative) prospect theory, among many others. Noncompensatory models, such as lexicographic semiorders (Tversky, 1969) or the priority heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006) violate such structural properties.

To consider the future direction of decision research, we propose to carry out a thorough "triage" at a general level. If preferences are consistent with strict weak orders, then this has implications for the type of process models that can work, and it eliminates the large class of existing or potential theories that violate weak orders, for example, because they predict intransitive preferences. On the other hand, if preferences are not consistent with strict weak orders, then we may have to give up modeling choice through numerical representations. This would have far-reaching consequences, for example, in modeling economic behavior.

We believe that much confusion can be eliminated if we find successful means to distinguish variability of overt choice from structural inconsistency of hypothetical, latent preference. We have shown how to formulate a probabilistic model of weak order preference and, equivalently, a probabilistic model of numerical utility. While the basic concepts underlying these mathematical models, by themselves, are not new, the models have never been fully characterized nor have they ever been tested empirically in their full generality using quantitative methods. We leverage new

mathematical and statistical results to offer the first direct, parsimonious and quantitative test of weak order preferences and numerical utility representations to date.

Our approach is furthermore guided by three key criteria for the rigorous development and testing of psychological theories (S. Roberts & Pashler, 2000; Rodgers & Rowe, 2002). (a) Researchers should understand the full scope of a theory's predictions and focus on models that make highly restrictive predictions. (b) Researchers should understand and control or limit the sampling variability in their data. (c) Researchers should more often ask themselves: "what would disprove my theory?" (S. Roberts & Pashler, 2000, p. 366). The a priori expectation should be that the theory will be rejected; hence, a good fit should be a surprise.

Scope and Parsimony

The weak order model makes predictions on stimuli like our noncash prospects *A* and *C*, for which other theories, such as the priority heuristic and parametric versions of cumulative prospect theory are mute. We have fully characterized the model for five stimuli. It forms a convex polytope in a 20 dimensional space, characterized by precisely 75,834 nonredundant mathematical constraints that define a volume of just one two-thousandth of the space of all possible ternary paired comparison probabilities. We derive these restrictive predictions using a minimum of mathematical assumptions, that is, we expressly avoid common auxiliary assumptions that are, at best, idealizations of a more complex reality.

Data Variability

The choice frequencies differ dramatically between respondents. We did not aggregate across participants because such aggregated data need not be representative of even a single person. Even within person, we went beyond the standard approach that often merely considers modal choices. We provided and tested highly specific quantitative predictions about the predicted choice proportions for each individual. By analyzing individual participants

separately, we controlled for interindividual differences. By collecting 45 observations per trinomial for each person, we reduced sampling variability substantially, especially compared with studies with one observation per person per gamble pair.

How Surprising Is the Model Fit?

By including stimuli with a historic reputation for allegedly yielding intransitive preferences, we biased the experimental paradigm against the weak order model. Even disregarding the nature of the stimuli themselves, for five choice alternatives, 99% of asymmetric binary preference relations violate weak orders, and only one two-thousandth of all conceivable ternary paired comparison probabilities satisfy the polytope. Using simulation methods, we estimated the average power of our test to be about .96.

Yet, our data analysis revealed a remarkable picture: Over all three experimental conditions combined, 60 out of 90 data sets generated choice proportions that lie inside of the tiny space allowed by the model. Hence, two thirds of our data sets yielded a perfect fit. A perfect fit means that the model generated predictions that are so accurate that no matter how small the significance level α, we cannot reject the model. Combining all three conditions, only five out of 90 data sets violated the strict weak order model at a 5% significance level. This is just about at the usual Type-I error level. We also cross-replicated the excellent model performance within individuals across three stimulus sets, including noncash gambles: Of 30 individuals, 13 consistently fit perfectly across all three gamble sets, and none generated a significant violation for more than a single gamble set.

We plan to complement this work with a Bayesian analysis along the lines sketched by Myung et al. (2005). By the likelihood principle and thanks to our large sample size, we expect the Bayesian viewpoint to reinforce our substantive conclusions, unless one employs prior distributional assumptions that give little weight to the empirical data. Since a Bayesian approach does not allow for a perfect fit, we predict that our perfect fit results will translate into very high Bayes' factors in favor of the weak order model over an unconstrained model (that permits all ternary choice probabilities).

Variable choice behavior is often cited as a reason to reject structural properties of algebraic theories, such as, here, weak orders and numerical utility representations. We show that variability of preference (or uncertainty about preference) can be separated from violations of structural properties, even without requiring an error theory. In our model, all strict weak orders are allowable preference states, that is, we allow for maximal variability in latent preference and overt choice. But we also require perfect adherence to the structural properties of strict weak orders in latent preference. Adding a random error component to the model would relax our constraints by inflating the weak order polytope and would automatically further improve the already excellent fit.

Common sense tells us that the variability of, say, restaurant choice, should not be mistaken for incoherence in, say, food preference. Probabilistic models like the weak order model allow us to tease apart variability of choice from inconsistency of preference. Future work should replicate our findings in other labs and expand the analysis to other stimulus sets.

References

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

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

Christof, T., & Löbel, A. (1997). Porta: Polyhedron representation transformation algorithm [Software]. Retrieved from http://www2.iwr.uni-heidelberg.de/groups/comopt/

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

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

Fiorini, S., & Fishburn, P. C. (2004). Weak order polytopes. *Discrete mathematics, 275,* 111–127. doi:10.1016/S0012-365X(03)00101-8

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. doi:10.1016/0165-4896(92)90036-5

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291. doi:10.2307/1914185

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92,* 805–824. doi:10.2307/2232669

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

Macmillan, N., & Creelman, C. (2005). *Detection theory: A user's guide.* Mahwah, NJ: Erlbaum. doi:10.1037/h0043158

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

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

Niederée, R., & Heyer, D. (1997). Generalized random utility models and the representational theory of measurement: A conceptual link. In A. A. J. Marley (Eds.), *Choice, decision and measurement: Essays in Honor of R. Duncan Luce* (pp. 155–189). Mahwah, NJ: Erlbaum.

Pfeiffer, G. (2004). Counting transitive relations. *Journal of Integer Sequences, 3,* 04.3.2.

Regenwetter, M. (1996). Random utility representations of finite m-ary relations. *Journal of Mathematical Psychology, 40,* 219–234. doi:10.1006/jmps.1996.0022

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

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

Roberts, F. S. (1979). *Measurement theory.* London, England: Addison-Wesley.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367. doi:10.1037/0033-295X.107.2.358

Rodgers, J. L., & Rowe, D. C. (2002). Theory development should begin (but not end) with good empirical fits: A comment on Roberts and Pashler (2000). *Psychological Review, 109,* 599–603. doi:10.1037/0033-295X.109.3.599

Savage, L. J. (1954). *The foundations of statistics.* New York, NY: Wiley.

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

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297–323. doi:10.1007/BF00122574

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Ziegler, G. M. (1995). *Lectures on polytopes.* Berlin, Germany: Springer-Verlag.