Online Supplement:

Behavioral Variability of Choices

Versus

Structural Inconsistency of Preferences

Michel Regenwetter¹ Clintin P. Davis-Stober²

¹ University of Illinois at Urbana-Champaign, USA

² University of Missouri, USA

1 Proof of the Theorem.

Denote by W_0C the collection of all strict weak orders on a given finite set C of choice alternatives. A collection $(P_{ij})_{i,j\in\mathcal{C}\atop i\neq j}$ 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 $WO_{\mathcal{C}}$

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

$$
\succ \mapsto P_{\succ},
$$

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

$$
P_{ij} = \sum_{\substack{\succ \in W \mathcal{O}_C \\ i \succ j}} P_{\succ}.\tag{1}
$$

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 $\forall i, j \in \mathcal{C}, i \neq j$,

$$
P_{ij} = \mathcal{P}(\mathbf{U}_i > \mathbf{U}_j). \tag{2}
$$

THEOREM. A system of ternary paired comparison probabilities P_{ij} satisfies the weak order model in (1) if and only if it satisfies the distribution-free random utility model in (2).

PROOF: By Observation 3 of Regenwetter (1996), for any probability distribution P over strict weak orders there exist random variables $(\mathbf{U}_c)_{c \in \mathcal{C}}$ and a probability measure \mathcal{P} , and, vice versa, for any random variables $(U_c)_{c \in \mathcal{C}}$ and a probability measure $\mathcal P$ there exists a probability distribution P over strict weak orders, such that

$$
\sum_{\lambda \in W_{c}^{\mathcal{O}_{\mathcal{C}}}} P_{\lambda}
$$
\n
$$
= \sum_{\lambda \in W_{c}^{\mathcal{O}_{\mathcal{C}}}} \mathcal{P} \left(\bigcap_{\substack{f: c \to \mathbb{R} \\ f(k) > f(l) \Leftrightarrow k \succ l}} \bigcap_{c \in \mathcal{C}} U_{c} = f(c) \right) = \sum_{\lambda \in W_{c}^{\mathcal{O}_{\mathcal{C}}}} \mathcal{P} \left(\bigcap_{\substack{k,l \in \mathcal{C} \\ k \succ l \\ m \neq n}} U_{m} \leq U_{n} \right)
$$
\n
$$
= \mathcal{P} \left(U_{i} > U_{j} \right).
$$

QED.

The Theorem is illustrated by the example at the center of Figure 1 in the article.

2 Characterizing the Weak Order Mixture Model us-

ing Convex Geometry.

Figure 1: Vertices of the strict weak order polytope for $C = \{a, b\}.$

Figure 1 illustrates the vertex representation in the very simple two-element case where $C = \{a, b\}$. In that case, there are only three strict weak orders, and these are represented in the figure by three vertices with $0/1$ -coordinates. The horizontal axis indicates P_{ab} , the vertical axis indicates P_{ba} . There are three possible strict weak orders: $a \succ b$, $b \succ a$, and "no preference," i.e., neither $a \succ b$, nor $b \succ a$. These three strict weak orders are represented as three vertices with $0/1$ coordinates. If a person always prefers a to b, we can write this as $P_{ab} = 1, P_{ba} = 0$, hence yielding the coordinates $(1, 0)$ for the strict weak order $a \succ b$. The other two vertices have similar interpretations. The point (1,1) does not represent a strict weak order. In our analysis with five choice alternatives, we represent 541 strict weak orders as 541 such $0/1$ vertices in a 20-dimensional space. The remaining million or so $0/1$ vertices in 20-dimensional space, like (1,1) in this figure, do not represent weak orders.

While each vertex can be thought of as a degenerate probability that places all 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 strict weak order mixture model. Figure 2 illustrates this also when $\mathcal{C} = \{a, b\}$. In that case, the strict weak order polytope is a triangle, or more precisely, it is the convex hull of three $0/1$ vertices. Every probability distribution over the three strict weak orders can be represented as a point in the convex polytope (here the shaded triangle). The figure shows various examples of such probability distributions. Each distribution is represented by one point in the triangle. In our analysis with five choice alternatives, we consider the convex hull of 541 such $0/1$ vertices in a 20-dimensional space.

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 weak order polytope, i.e., to figure out a system of affine inequalities whose solution set is the convex polytope. We can then attempt to test whether the ternary paired comparison probabilities satisfy or violate that system of inequalities. The step of finding such inequalities often poses a formidable mathematical challenge. The two-element case, where $\mathcal{C} = \{a, b\}$, however, is trivial, so we show it in Figure 3. The same triangle, that Figure 2 described as the convex hull of the

Figure 2: Strict weak order polytope for $C = \{a, b\}$, vertex representation as a convex hull of 0/1 vertices.

points $(0,0), (0,1),$ and $(1,0),$ can also be described as the region that satisfies the inequalities

$$
P_{ab} \ge 0, \quad P_{ba} \ge 0, \quad P_{ab} + P_{ba} \le 1. \tag{3}
$$

We simply exploit the fact that a triangle can be described either as the area between the three 'corners' or as the area between the three sides. While the case of $C = \{a, b\}$ is very simple and can be drawn in 2-dimensional space, the general case becomes very complicated as the number of objects in $\mathcal C$ increases. The convex polytope (here the triangle) is the region that simultaneously satisfies three inequalities, namely $P_{ab} \ge 0, P_{bc} \ge 0$, and $P_{ab} + P_{bc} \le 0$,

Figure 3: Strict weak order polytope for $C = \{a, b\}$, as a system of affine inequality constraints.

that are indicated visually with black arrows. Each of these three inequalities are facetdefining, and in this two-dimensional case, all three facet-defining inequalities automatically hold for all possible ternary paired comparison probabilities. In our analysis with five choice alternatives, rather than obtaining 3 inequalities in 2-dimensional space, we obtain over 75,000 testable inequalities in a 20-dimensional space. While the triangle in this figure occupies the entire space of ternary paired comparison probabilities, the polytope in our analysis occupies only $\frac{1}{2,000^{th}}$ of the space of ternary paired comparison probabilities.

In our illustration, Figures 1-3 illustrate a case with two choice alternatives, three vertices,

and three facet-defining inequalities in 2-dimensional space. The geometry of this example is straightforward: The convex polytope is a triangle, and its facets simply form the three sides of the triangle. The inequalities in (3) form a complete minimal description because 1) knowing the three sides suffices in order to completely describe the triangle, and 2) if we drop one of the sides we no longer have a triangle, so the three inequalities are nonredundant.

MODEL COMPLEXITY/PARSIMONY: To characterize the level of parsimony in our predictions, we wanted to know the volume of the strict weak order polytope for five choice alternative, relative to the 20-dimensional space of all conceivable data generating trinomials (each vector of trinomials is a point in the sample space). We independently drew points from the space, using a uniform distribution and saw what proportion of points lie in the polytope. So, we estimated what proportion of trinomials fall within the model. Specifically, we drew 200,000 such points for a high level of accuracy. Of those, roughly 100 lay in the polytope.

3 Experiment.

Our experimental study was an expansion of the most important landmark study of intransitivity of preferences, namely Tversky's (1969) main experiment. Tversky had eight participants take altogether 20 separate two-alternative forced choice (2AFC) decisions for each pair among five gambles. We used similar gambles as stimuli, but decisions were ternary paired comparisons, and the experiment involved 30 participants. We also secured high statistical power by collecting 45 trials per gamble pair and per respondent.

Each participant attended three sessions, and each session took approximately one hour

to complete. Each session consisted of the participant working at a computer, making ternary paired comparisons among gambles. The computer program iteratively displayed pairs of gambles, and the participant used the computer mouse to click on the gamble in the pair that he/she preferred. If the participant could not decide between the two gambles, he/she clicked the third option, a picture of a pan balance which they had been instructed to interpret that they felt "indifferent" among the two offered gambles. After a few warm-up choices at the beginning of each session, each participant made $15 \times 4 \times 10 = 600$ choices between gambles per session. The participants were allowed to take short breaks whenever necessary to reduce fatigue during a session.

The gambles were organized into four sets: Gamble Set I, Gamble Set II, Gamble Set III, and Distractors. The gambles in these sets are described below. There are five gambles in each set, so each set's gambles can be paired ten distinct ways. The computer cycled through the four sets as it chose gamble pairs for the participant, i.e., one pair of gambles from each set was shown to the participant over a four-trial cycle, and then the cycle repeated in the same order with new gamble pairs from each set. The gamble pairs were drawn from the sets in a quasi-random fashion with the conditions that the same gamble pair did not appear in consecutive cycles, and each gamble pair (except from the Distractor set) was shown to the participant 15 times per session. The stimuli are listed in Table 1.

Payoffs included gift cards redeemable for the following prizes from local shops: 15 sandwiches, 40 movie rentals, 4 CDs, 40 coffees, or 7 books. The gift cards were worth about \$75, but participants were not informed of their exact cash values. At the beginning of the experiment, participants ranked these prizes 1-5 in order of preference. The first-ranked prize was associated with the lowest probability of winning, the second-ranked prize associated Table 1: Three Gamble Sets. Gamble Set I matches Tversky's stimuli in 2006 dollar equivalents. (Data collection occurred in 2006.)

Gamble	Α	В	\mathcal{C}		E,
Probability of Winning	7/24	8/24	9/24	$10/24$ $11/24$	
Payoff		$$26.32 \mid $25.00 \mid $23.68 \mid $22.36 \mid 21.04			

Gamble Set I

Gamble Set III

with the second lowest probability, and so on. Figures 4 and 5 show screen shots of ternary paired comparison trials from Gamble Set I and Gamble Set III, respectively.

Distractors

These gambles were randomly generated, as follows. The payoffs in this set were randomly chosen from a uniform distribution on the 15 payoffs in the three sets above. The probabilities of winning were randomly generated from a uniform distribution on the interval [0.02, 0.42].

Figure 4: Screenshot of a ternary paired comparison using Gamble Set I.

Participant Information and Recruiting

30 participants completed all three sessions. No information was formally recorded in regard to age, gender, or other demographic factors, but we believe that the participants were roughly balanced in terms of gender, and that most, if not all, were college students.

Participants were recruited through posted advertisements in academic buildings and dormitories around the University of Illinois campus. Participants were informed that the study involved decision making, and that they would be paid for their participation in the study. As the research applies to members of the general adult population who are able to read and perform a simple choice or ranking task on the computer, no particular participant

Figure 5: Screenshot of a ternary paired comparison using Gamble Set III.

groups were targeted or excluded. After a short instructional presentation in the first session, participants signed a consent form and were given the opportunity to try the task and decide whether they were able and willing to perform it.

Payment

At the end of each session a participant completed, he/she received a \$5 minimum payment. In addition to this baseline payment, a gamble was randomly selected (using a uniform distribution) from a collection of cash gambles that the participant had chosen during the session, and the participant played this gamble for real money. For gamble pairs where the participant clicked 'indifferent,' a gamble from the pair was chosen at random and added to the collection of chosen gambles (only for the purpose of payment). This gave participants an incentive to choose gambles carefully, as they might eventually play any gamble they chose during the session, and to avoid overusing the indifference option, since it could result in either gamble in a pair entering the collection of potential gambles that participants might ultimately play for real money. Since cash gambles ranged in value from \$20-\$31.43, participants earned \$5-\$36.43 per session. To add further incentive for the participants to complete all three sessions, at the end of the third session, a noncash gamble (such as a chance for free movie rentals) was also played in addition to the cash gamble. This gamble was selected in a similar fashion as the cash gambles.

4 Data.

We provide the raw ternary paired-comparison frequencies of all 30 participants across all three gamble sets. For each participant and gamble set, we also indicate whether the choice proportions were inside the weak order polytope (indicated by "inside WOP"), i.e., lead to perfect fit. Whenever they were not inside WOP, we provide the p-value of the violation. This p-value was computed using the technique developed in Davis-Stober (2009). A small p-value indicates a significant violation of the model, whereas a large p-value indicates a violation that can be attributed to sampling variability. Davis-Stober's technique yields the appropriate "Chi-bar-square" distribution for the asymptotic distribution of the $G²$ statistic, based on the local geometric structure of the weak order polytope in a neighborhood around the maximum likelihood point estimate.

References

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

Regenwetter, M. (1996). Random utility representations of finite m-ary relations. Journal of Mathematical Psychology, 40:219—234.