Moderate Utility*

Junnan He
and
Paulo Natenzon†

Updated: December 2022

Abstract

Hotelling's and Salop's spatial competition models as well as nested logit, covariance probit, elimination-by-aspects, and several other well-known discrete choice models belong to the class of *moderate utility models*, where binary choices are a function of the ratio between utility difference and a product differentiation index satisfying the properties of a distance metric. We provide a behavioral foundation for this class of models. Our main result establishes that moderate utility has a single, directly testable implication: choice probabilities are moderately transitive. We use our characterization to show how the model achieves a useful compromise between explanatory power and predictive power.

*A previous version circulated under the title "Moderate Expected Utility".

†Junnan He, Department of Economics, Sciences Po. Paulo Natenzon, Olin Business School, Washington University in St. Louis. Contact email: `pnatenzon@wustl.edu`. We are grateful to Faruk Gul, Ryota Iijima, Tomasz Strzalecki, and audience participants at the Barcelona GSE Workshop on Stochastic Choice, the CIREQ Montreal Workshop on the Foundations of Discrete Choice, Princeton University, UBC, McGill University, UC Berkeley, UC Davis, RUD Paris, BRIC Aarhus, AMES Xiamen, the University of Rochester, and the University of Paris 1, for helpful comments and suggestions. First version posted in September 2018. Please check `http://pnatenzon.wustl.edu` for the latest version.

1 Introduction

A wide variety of discrete choice models used by economists —including Hotelling's and Salop's spatial competition models, as well as the classic logit, nested logit, probit, covariance probit, and elimination-by-aspects — are special cases of the *moderate utility model* proposed in psychology by Halff (1976). In this model, the probability that an option x is preferred to another option y is given by:

$$\rho(x,y) = F\left(\frac{u(x)-u(y)}{d(x,y)}\right) \tag{MUM}$$

The utility function u in the MUM represents the value of each option. The distance metric d captures the *differentiation* between the options and reflects their substitutability and their comparability. The ratio $[u(x) - u(y)]/d(x,y)$ can be interpreted as the strength of preference for option x over option y. The increasing transformation F maps strength of preference to choice probabilities and satisfies $F(t) = 1 - F(-t)$ for all t.

Despite its ubiquity and wide-ranging application, the consequences of the MUM formula for observable choice behavior —that is, a set of necessary and sufficient conditions that exhaust the testable implications of the model— have not been established. Our main contribution is to fully determine the empirical content of the MUM formula, providing behavioral foundations for binary discrete choice with differentiation.

We identify a single, directly testable, non-parametric condition that fully characterizes the model. Theorem 1 shows that a binary choice rule ρ over a finite set of alternatives can be represented by the MUM if and only if it is *moderately transitive.* This condition requires that if x is preferred to y with probability larger than one-half and, in turn, y is preferred to z with probability larger than one-half, then the probability that x is preferred to z must lie above the *minimum* of those two probabilities. For example, if $\rho(x,y) = 0.6$ and $\rho(y,z) = 0.8$, moderate transitivity requires that $\rho(x,z) > 0.6$. Moderate transitivity is a less studied, intermediate condition between the two well-known postulates of *weak transitivity* (which for the same antecedent requires the less demanding conclusion $\rho(x,z) \geq 0.5$ in the example above) and *strong transitivity* (which requires the more demanding conclusion $\rho(x,z) > 0.8$).

The transitivity postulates above are directly testable in data by checking simple moment inequalities. In Section 2, we present the formal definitions and apply them to data examples drawn not only from economics, but also from psychology and biology (Examples 2–4). We use these examples throughout the paper for illustrative purposes.

They belong to a robust class of empirical phenomena in which strong transitivity fails, but moderate transitivity still holds. By Theorem 1, all these examples can be accommodated by the MUM. In addition, we show that moderately transitive models retain significantly more empirical bite and predictive power out-of-sample than weakly transitive models. Thus, Theorem 1 implies that the MUM provides the analyst with a valuable modeling compromise between flexibility and predictive power.

We contribute to an important literature that seeks to establish the equivalence between binary stochastic choice models and their testable implications (see Figure 1). Existing characterizations by Debreu (1958), Luce (1959), Tversky and Russo (1969), Fudenberg et al. (2015) study models which ignore the role of differentiation and impose strong transitivity. Theorem 1 is the first result to provide behavioral foundations for binary discrete choice with differentiation.

Theorem 1 reveals how parametric restrictions commonly used by practitioners in applied work map on to restrictions on observable choice behavior. For instance, specializing the distance metric in the MUM to the discrete metric is equivalent to going from moderately transitive behavior to acyclic* behavior. Conversely, we use Theorem 1 to show that the triangle inequality property of the distance metric in the MUM is exactly responsible for the 'gap' between weakly and moderately transitive choice behavior.

Some special cases of the MUM are also instances of the random utility model (RUM) characterized by Block and Marschak (1959) and Falmagne (1978). We show that, despite having a non-empty intersection, neither the MUM nor the RUM nest each other.

The paper is organized as follows. Section 2 introduces the setup and the transitivity postulates. Sections 3 presents the moderate utility model and contains our main result. Section 4 compares our result to previous existing characterizations and obtains a generalization. Section 5 concludes by showing the MUM is not nested nor nests the classic random utility model.

2 Stochastic choice and transitivity

We begin with some definitions. Let Z be a finite set of choice options. A (binary, stochastic) *choice rule* on Z is a function $\rho : Z^2 \to [0,1]$ such that $\rho(i,j) + \rho(j,i) = 1$

*Acyclicity, which we define in Section 4, is slightly more restrictive than strong transitivity (Fudenberg et al., 2015).

for every $i, j \in Z$. When ρ represents demand in a population of standard rational consumers, the number $\rho(i, j)$ is the proportion of the population that prefers i to j. When ρ represents individual stochastic choice, $\rho(i, j)$ is the probability that the decision maker selects option i in a binary comparison against j.

A choice rule may satisfy one of several probabilistic versions of the classic transitivity postulate. The two best known notions of stochastic transitivity for binary choice data are weak transitivity and strong transitivity. A choice rule ρ satisfies *weak transitivity* when for any $i, j, k \in Z$,

$$\min\{\rho(i, j)\, \rho(j, k)\} \geq 1/2 \text{ implies } \rho(i, k) \geq 1/2.$$

A choice rule ρ satisfies *strong transitivity* when for all $i, j, k \in Z$,

$$\min\{\rho(i, j)\, \rho(j, k)\} \geq (>)\, 1/2 \text{ implies } \rho(i, k) \geq (>)\, \max\{\rho(i, j), \rho(j, k)\}.$$

where a strict inequality in the hypothesis implies a strict inequality in the conclusion. Weak and strong transitivity are well-studied in the literature (see Rieskamp et al., 2006). In this paper we focus on a less studied, intermediate form of transitivity. A choice rule ρ satisfies *moderate transitivity* when for all $i, j, k \in Z$,

$$\min\{\rho(i, j)\, \rho(j, k)\} \geq 1/2 \text{ implies } \begin{cases} \rho(i, k) > \min\{\rho(i, j), \rho(j, k)\} \\ \text{or} \\ \rho(i, k) = \rho(i, j) = \rho(j, k) \end{cases}$$

It is clear from the definitions above that strong implies moderate and moderate implies weak transitivity.

The three transitivity postulates are directly testable in choice data by checking simple moment inequalities. We now show that moderate transitivity achieves a useful compromise between explanatory power and predictive power. First, we show that moderate transitivity often holds in empirical settings where strong transitivity is systematically violated. We then show that a model that imposes moderate transitivity affords the analyst significantly more predictive power out-of-sample than just imposing weak transitivity.

An early theoretical example suggests that strong transitivity is too stringent of a requirement to describe choice behavior with pairs of options that vary in their degree of differentiation and comparability:

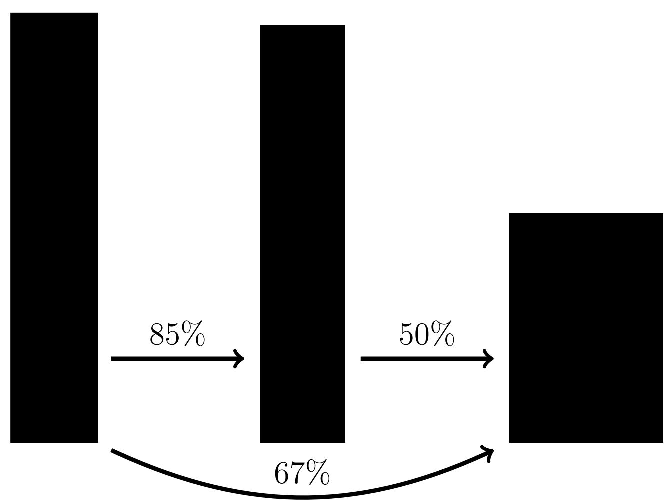
Example 1 (Georgescu-Roegen (1958)). *A consumer is equally likely to choose either option between two consumption bundles* $A = (a_1, a_2, \cdots, a_n)$ *and* $B = (b_1, b_2, \cdots, b_n)$ *because they involve several hard-to-evaluate tradeoffs across the n commodities. That is* $\rho(A, B) = 1/2$*. Now consider a small improvement in the first bundle* $A' = A + \Delta$ *where* $\Delta \geq 0$*. It is reasonable to expect the consumer to clearly prefer the improved bundle over the original bundle* $\rho(A', A) = 1$*. It is not reasonable, however, to expect strong transitivity to hold in this case, since it would imply that all the difficulty must be resolved by the small improvement* $\rho(A', B) \geq \max\{\rho(A', A), \rho(A, B)\} = 1$*. Moderate transitivity is more reasonable, requiring only that* $\rho(A', B) > \min\{\rho(A', A), \rho(A, B)\} = 1/2$*.*

The example above is also attributed by Luce and Suppes (1965) to L. J. Savage. In Savage's rendition, a child has difficulty choosing between a bicycle (option A) and a pony (option B). A store owner brings in a second bicycle that is identical to the first bicycle but clearly better in minor ways, such as having a bell (option A'). The child still hesitates. An impatient parent forces the child to choose between the two bicycles, and the child immediately picks the slightly better bicycle. The same example is also recast by Tversky (1972a) with a trip to Paris (option A), a trip to Paris enhanced by a €1 bonus (option A') and a trip to Rome (option B).

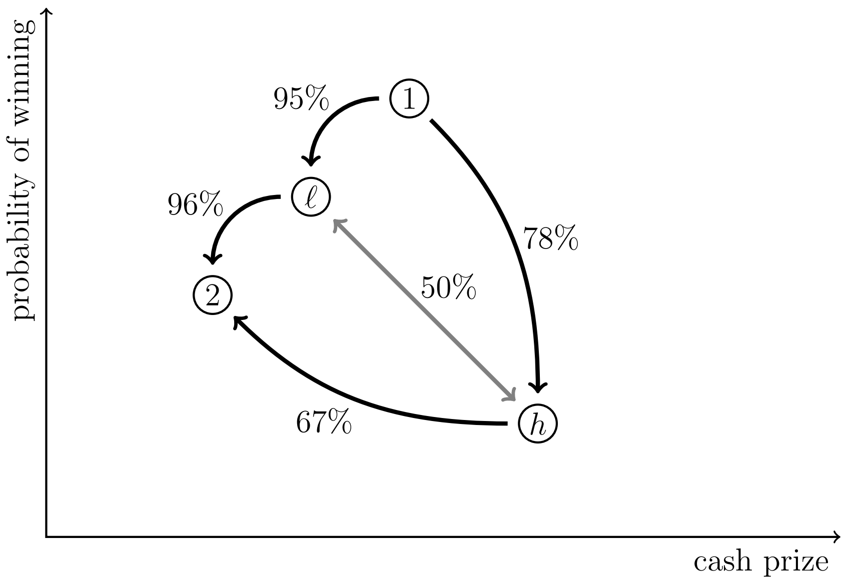
Example 1 suggests that strong transitivity should be violated systematically in real-world choice data and, in fact, the accumulated evidence is very robust. Reviewing some of this evidence, Mellers, Chang, Birnbaum and Ordonez (1992, p. 348) note that "weak and moderate stochastic transitivity are often satisfied, although a few exceptions have been noted," while "[s]trong stochastic transitivity is frequently violated." Chipman (1960) provides perhaps the earliest empirical demonstration of the intuition behind Example 1 in economics.

Tversky and Russo (1969) provide a visually compelling demonstration of this same intuition in psychology (see Example 2). Note that the empirical violation of strong transitivity in Example 2 follows the recipe suggested by Example 1. Like options A and A', the first and second rectangles in Example 2 are less differentiated and therefore easier to compare. And like option B, the third rectangle is more differentiated from the other options and therefore harder to compare.

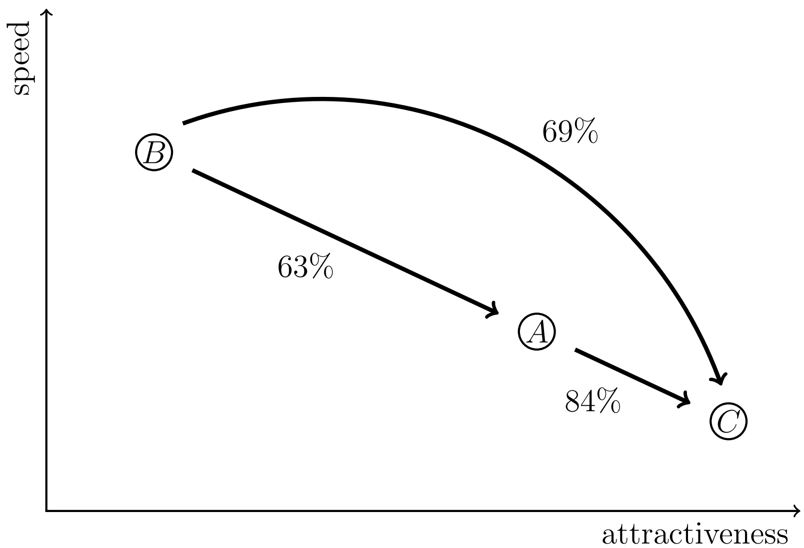
Likewise, the different levels of differentiation between lotteries over money also drive violations of strong transitivity in Example 3 below. Finally, Example 4 is taken from animal mating choice experiments in biology, further broadening the empirical reach of the intuition behind Example 1.



Example 2 (Perceptual task data)**.** *Tversky and Russo (1969) recorded hundreds of decisions by prison inmates in Michigan in perceptual choice tasks. Subjects were shown many different pairs of rectangles and were asked to pick the rectangle with the largest area in each pair. Three of these rectangles and their pairwise relative frequencies of choice are shown above. The middle and right rectangles have equal areas, so that each one is chosen 50% of the time in a binary comparison. The left rectangle is slightly larger than the others. The same difference in area was more easily detected in pairs with less differentiated shapes (85% correct answers) than in pairs with more differentiated shapes (67% correct answers). These frequencies violate strong transitivity but satisfy moderate transitivity.*



Example 3 (Choice over lotteries). *Soltani, De Martino and Camerer (2012) recorded thousands of choices by 21 male Caltech undergraduates using simple lotteries that pay a cash prize of m dollars with probability p in the lab. A high risk lottery h and a low risk lottery ℓ, depicted above, were fine-tuned to each individual to be approximately indifferent, (i.e., equally likely to be chosen in a binary comparison). Slightly perturbed versions of h and ℓ were then offered for comparison against several types of 'decoy' lotteries. Above we depict the relative location of two decoy lotteries* 1 *and* 2 *with respect to h and ℓ. Decoy lottery* 1 *dominates ℓ and was chosen 95% of the time against ℓ but only 78% of the time against h. Thus, choice frequencies violate strong transitivity in the direction $1 \to \ell \to h$. Decoy lottery* 2*, on the other hand, is dominated by ℓ and was chosen 4% of the time against ℓ and 33% of the time against h. Hence, choice frequencies also violate strong transitivity in the direction $h \to \ell \to 2$. It is easy to verify that moderate transitivity holds in both cases.*



Example 4 (Animal studies)**.** *Lea and Ryan (2015) recorded hundreds of mating decisions by female túngara frogs. Female túngara frogs choose mates based on the sound of their call. Above we depict how the calls of the three male options A, B and C were differentiated along two desirable attributes. The horizontal axis represents a measure of static attractiveness, and the vertical axis represents speed measured in calls per second. In binary comparisons, option B was chosen in 63% of the trials against A; option A was chosen in 84% of the trials against C; and option B was chosen in 69% of the trials against C. Choices therefore satisfy moderate transitivity but violate strong transitivity.*

These examples show that strong transitivity is too strong to be descriptive: relaxing it helps address a robust range of empirical phenomena. Conversely, we now argue that the weak transitivity postulate is too weak: it allows that $\rho(i,k) = .51$, for example, even if we observe that $\rho(i,j) = \rho(j,k) = .95$. Imposing moderate transitivity, in this case, leads to the sharper and arguably more sensible prediction $\rho(i,k) \geq 95$.

To quantify this additional predictive power, let ρ be a weakly transitive choice rule. Enumerate the options in $Z = \{1,\ldots,n\}$ in such a way that $\rho(i,j) \geq 1/2$ whenever $i \leq j$. For the sake of simplicity, let us assume that choice probabilities differ whenever possible, so that the set $\{\rho(i,j) \in [0,1] : i \neq j\}$ has maximum cardinality with $n(n-1)$ elements.

When $Z = \{1,2,3\}$ has three alternatives, weak transitivity allows ρ to have six strict orderings:

$$\text{weakly transitive}\left\{\begin{array}{l} \left.\begin{array}{l} \rho(1,3) > \rho(1,2) > \rho(2,3) \\ \rho(1,3) > \rho(2,3) > \rho(1,2) \\ \rho(1,2) > \rho(1,3) > \rho(2,3) \\ \rho(2,3) > \rho(1,3) > \rho(1,2) \end{array}\right\}\text{moderately transitive} \\ \rho(1,2) > \rho(2,3) > \rho(1,3) \\ \rho(2,3) > \rho(1,2) > \rho(1,3) \end{array}\right.$$

Moderate transitivity rules out the last two of the six strict orderings, where $\rho(1,3) < \min\{\rho(2,3), \rho(1,2)\}$. Let $\text{Weak}(n) = [n(n-1)/2]!$ denote the number of strict orderings allowed by weak transitivity when Z has n options, and likewise, let $\text{Moderate}(n)$ denote the number of strict orderings allowed by moderate transitivity. The ratio $\text{Moderate}(n)/\text{Weak}(n)$ provides a measure of how restrictive is moderate compared to weak transitivity. In the case $n = 3$ we just showed the ratio $\text{Moderate}(3)/\text{Weak}(3)$ is equal to $2/3$. This ratio decreases to less than $1/4$ when $n = 4$ and less than $1/17$ when $n = 5$. In fact, the ratio is arbitrarily small when n is large:

Proposition 1. $\lim_{n\to\infty} \mathit{Moderate}(n)/\mathit{Weak}(n) = 0.$

We prove Proposition 1 in the Appendix. For completeness, we also show in the Appendix that the ratio between the number of strongly transitive orderings and moderately transitive orderings goes to zero when n is large.

To summarize, a parametric choice model that spans the range of moderately transitive choice behavior may be useful in two ways: first, it provides the flexibility that is needed to accommodate empirical violations of strong transitivity. Second, it imposes significant restrictions out of sample —allowing the analyst to make sharper predictions— than the more lenient weak transitivity. We describe such a model in the next section.

3 Moderate utility model

A choice rule ρ on a finite set Z is a *moderate utility model (MUM)* if there is a utility function $u : Z \to \mathbb{R}$, a distance metric $d : Z^2 \to \mathbb{R}_+$ and a strictly increasing function F, such that for all $i \neq j$,

$$\rho(i,j) = F\left(\frac{u(i) - u(j)}{d(i,j)}\right) \tag{1}$$

where F satisfies $F(t) = 1 - F(-t)$ for all t. This last requirement automatically follows from $F(t) = \rho(i,j) = 1 - \rho(j,i) = 1 - F(-t)$. The utility u represents the value of each option. It is easy to see that $\rho(i,j) \geq 1/2$ if and only if $u(i) \geq u(j)$. The distance d measures the *differentiation* between i and j. For any fixed utility difference, increasing the differentiation $d(i,j)$ drives choice probabilities closer to $1/2$, capturing the fact that more differentiated objects are less substitutable and harder to compare. The ratio $[u(i) - u(j)]/d(i,j)$ can be interpreted as the strength of preference for option i over option j, while the function F maps strength of preference to choice probabilities.

Many models proposed in the discrete choice literature to address empirical phenomena related to the comparability and the substitutability between different options turn out to be a MUM. While these models may appear to take very different forms, they all represent the differentiation between the options with a distance metric, which influences binary comparisons through a special case of formula (1). The following examples are perhaps the most familiar.

Example 5. *The probit model (Thurstone, 1927) is a MUM. The model is described by a Gaussian vector $(X_1, \ldots, X_n)$, each coordinate X_i corresponding to an option $i \in Z$, such that $\rho(i,j) = \mathbb{P}\{X_i > X_j\}$. Note that*

$$\rho(i,j) = \mathbb{P}\left\{\frac{X_i - X_j - \mathbb{E}[X_i - X_j]}{\sqrt{Var(X_i - X_j)}} > \frac{\mathbb{E}[X_i - X_j]}{\sqrt{Var(X_i - X_j)}}\right\} = \Phi\left(\frac{\mathbb{E}[X_i - X_j]}{\sqrt{Var(X_i - X_j)}}\right)$$

which is a special case of the MUM formula (1) with utility $u(i) = \mathbb{E}[X_i]$, distance metric $d(i,j) = \sqrt{Var(X_i - X_j)}$ (we allow correlation but rule out perfectly correlated *variables), and $F = \Phi$ the standard Gaussian cdf. The 'standard' probit model is a special case of this model with independent variables, and therefore also a MUM.*

Example 6. *The nested logit model (McFadden, 1978) is a MUM. The set of alternatives is partitioned into K disjoints nests $B_1 \cup \cdots \cup B_K = Z$. The utility of each option i is a random variable $U(i) = u(i) + \varepsilon_i$ where $u(i) \in \mathbb{R}$ is the deterministic part of utility, and the random taste shocks ε_i have the joint cumulative distribution*

$$G(\varepsilon_1, \ldots, \varepsilon_n) = \exp\left\{-\sum_{k=1}^{K}\left[\sum_{j\in B_k}\exp\left(-\varepsilon_j/\lambda_k\right)\right]^{\lambda_k}\right\}$$

with parameters $\lambda_1, \ldots, \lambda_K \in (0, 1]$. To see the nested logit is a MUM, note the probability that option i is preferred to option j is

$$\rho(i,j) = \mathbb{P}\{U(i) > U(j)\} = \left[1 + \exp\left(\frac{u(j) - u(i)}{d(i,j)}\right)\right]^{-1}$$

where $d(i,j) = \lambda_k$ whenever i and j belong to nest B_k, and $d(i,j) = 1$ when i and j belong to different nests. Writing the nested logit as a MUM allows for a simpler interpretation of the λ_k parameters: they measure the distance between alternatives within nest k. When all $\lambda_k = 1$ we obtain the standard logit model (Luce, 1959; McFadden, 1974) as a special case. Hence, standard logit is also a MUM.

Example 7. *The elimination-by-aspects model (Restle, 1961; Tversky, 1972a,b) is a MUM. The model assumes that each option i has a set of aspects $A(i)$. There is a measure m defined over the set of aspects such that*

$$\rho(i,j) = \frac{m\left[A(i)\right] - m\left[A(j)\right]}{m\left[A(i) \setminus A(j)\right] + m\left[A(j) \setminus A(i)\right]}.$$

This formula is a special case of MUM where utility $u(i) = m[A(i)]$ is the measure of the set of option i's aspects, distance $d(i,j) = m\left[A(i) \setminus A(j)\right] + m\left[A(j) \setminus A(i)\right]$ is the measure of the set of aspects that do not overlap between the i and j, and with $F(t) = 1/2 + t/2$.

Example 8. *Hotelling's spatial differentiation model (Hotelling, 1929) is a MUM. Consumers are uniformly distributed on the unit interval. A consumer $\ell \in [0, 1]$ chooses a firm $x \in [0, 1]$ at price p_x when it maximizes $w(x, p_x, \ell) = -(x-\ell)^2 - p_x$, which takes into account the disutility from price and from traveling to the firm's location. In a duopoly,* *the demand is split between firms x and y at the indifference point*

$$\ell^* = \frac{x+y}{2} + \frac{p_y - p_x}{2(y-x)}$$

which is a special case of MUM with utility $u(x, p_x) = x - x^2 - p_x$, distance $d(x, y) = |x-y|$ and transformation $F(t) = 1/2 + t/2$.

Other special cases of MUM abound in the literature, including the ideal point model of Coombs et al. (1961), the consumer address model of Salop (1979), the random coefficients model of Hausman and Wise (1978), the additive perturbed utility of Fudenberg et al. (2015), the Bayesian learning model of Natenzon (2019), the weighted linear model of Chambers et al. (2021), and the moderate linear model of He and Natenzon (2022).

Models in the MUM class take advantage of the differentiation metric d to address several empirical phenomena related to the substitutability and the comparability of choice options. Our main characterization result, below, implies the MUM can fit the all the choice data in Examples 1–4.

Theorem 1. *A choice rule ρ is a MUM if and only if it satisfies moderate transitivity.*

Halff (1976) proposed the MUM formula, and showed the first step we need to prove the necessity part of Theorem 1:

Lemma 1 (Halff, 1976)**.** *The MUM satisfies* $\min\{\rho(i,j), \rho(i,k)\} \geq 1/2 \Rightarrow \rho(i,k) \geq 1/2$.

Proof. Suppose that $\rho(i,j) \geq 1/2$ and $\rho(j,k) \geq 1/2$ but, contrary to the claim, $\rho(i,k) < \min\{\rho(i,j), \rho(j,k)\}$. A MUM representation and the triangle inequality property of the distance d would imply the contradiction

$$\begin{aligned} u(i) - u(k) &< d(i,k) \min\left\{\frac{u(i)-u(j)}{d(i,j)}, \frac{u(j)-u(k)}{d(j,k)}\right\} \\ &\leq [d(i,j) + d(j,k)] \min\left\{\frac{u(i)-u(j)}{d(i,j)}, \frac{u(j)-u(k)}{d(j,k)}\right\} \\ &\leq d(i,j)\frac{u(i)-u(j)}{d(i,j)} + d(j,k)\frac{u(j)-u(k)}{d(j,k)} \\ &= u(i) - u(k). \end{aligned}$$

□

To complete the necessity part of the proof, suppose $\rho(i,j) > \rho(j,k) = \rho(i,k) \geq 1/2$.

The MUM representation would require

$$\frac{u(i)-u(j)}{d(i,j)} > \frac{u(j)-u(k)}{d(j,k)} = \frac{u(i)-u(k)}{d(i,k)},$$

which, in turn, would imply a contradiction to the triangle inequality:

$$d(i,k) = \frac{u(i)-u(j)+u(j)-u(k)}{u(j)-u(k)} d(j,k) > d(i,j) + d(j,k)$$

and hence every MUM must satisfy moderate transitivity.

The proof of sufficiency in Theorem 1 is more involved and left for the Appendix. For any given moderately transitive choice rule ρ, we explicitly construct the utility u, the distance d and the transformation F; we show that d satisfies the properties of a metric (the key property being the triangle inequality); and we show that the MUM representation holds.

The next proposition shows the extent to which choices reveal ordinal information about the MUM parameters u and d:

Proposition 2. *Let ρ be a MUM with parameters u, d, and F. Then*

(i) $\rho(i,j) \geq 1/2$ *if and only if* $u(i) \geq u(j)$*;*

(ii) $\rho(i,j) > \rho(i,k) > \rho(j,k) \geq 1/2$ *implies* $d(i,j) < d(i,k)$*.*

Proof. From the MUM formula (1) it follows that $\rho(i,j) > 1/2$ if and only if $[u(i) - u(j)]/d(i,j) > 0$ if and only if $u(i) > u(j)$ proving (i). Suppose the assumption in item (ii) holds. Then, item (i) implies $u(j) \geq u(k)$ hence $u(i) - u(j) \leq u(i) - u(k)$. The MUM formula (1) implies $[u(i) - u(j)]/d(i,j) > [u(i) - u(k)]/d(i,k)$, hence $d(i,k) > d(i,j)$. □

Item (i) in Proposition 2 shows choices in a MUM reveal a complete and transitive ranking over the options represented by the utility parameter u. Item (ii) shows how every violation of strong transitivity is explained by the differentiation parameter d. To illustrate, consider the choice data from Example 4. By Proposition 2, the analyst concludes that any MUM that generates this data must satisfy $u(B) > u(A) > u(C)$ and $d(A,C) < d(B,C)$. Likewise, every MUM that generates the data in Example 3 must satisfy $u(1) > u(\ell) = u(h) > u(2)$, $d(2,\ell) < d(2,h)$ and $d(1,\ell) < d(1,h)$.

It is worth noting that the inequalities $d(A,C) < d(B,C)$, $d(2,\ell) < d(2,h)$ and $d(1,\ell) < d(1,h)$ revealed by choice data agree with the inequalities an analyst would obtain by applying the standard Euclidean distance, angle distance, or Manhattan distance to the vectors of measurable attributes in Example 3 and Example 4. That is, options that are revealed to be 'close' in the subjective parameter d are in fact 'close' in the space of observable attributes. The best way to map the abstract utility and differentiation parameters to the observable attributes must be determined empirically in any given application.[†] For an illustration, consider the example of angular distance used in Hausman and Wise (1978). In Example 3, options 1 and ℓ in form a small angle with respect to the origin, so that $d(1, \ell)$ is small, while options 1 and h form a wider angle with respect to the origin, so that $d(1, h)$ is large. Hence, a decision maker may ascribe to h and ℓ the same utility values, and yet have an easier time comparing 1 to ℓ than to h. In Example 4, options A and C form a smaller angle with respect to the origin than options B and C. Even if options A and B are close in value, the frogs find option C much easier to compare to A than to B. As these examples show, a MUM can address both situations in which the ease of comparison involves dominance (Example 3) and non-dominance (Example 4) in the attribute space.

4 Some restrictions and a generalization

The MUM characterized in Theorem 1 generalizes several nested models of stochastic binary choice in the literature, which we represent in order of generality in Figure 1. The most restrictive model, at the very bottom in Figure 1, is the binary *Logit model* in which choice probabilities are given by

$$\rho(i, j) = \frac{e^{u(i)}}{e^{u(i)} + e^{u(j)}} = \frac{1}{1 + e^{-[u(i)-u(j)]}} \tag{2}$$

for some utility function $u : Z \to \mathbb{R}$. Luce (1959) showed formula (2) is equivalent to the *product rule*

$$\rho(i, j)\rho(j, k)\rho(k, i) = \rho(i, k)\rho(k, j)\rho(j, i)$$

which can be interpreted as saying that the probability of observing a choice cycle in the direction $i \succ j \succ k \succ i$ is always equal to the probability of observing a choice cycle in the opposite direction. Luce (1959) obtains this equivalence under the mild assumption of *positivity*, which requires that $\rho(i, j) > 0$ for all i, j.

[†]See Apesteguia and Ballester (2018) for some important issues that may arise when the analyst maps the abstract parameters of a random choice model to the attribute space.

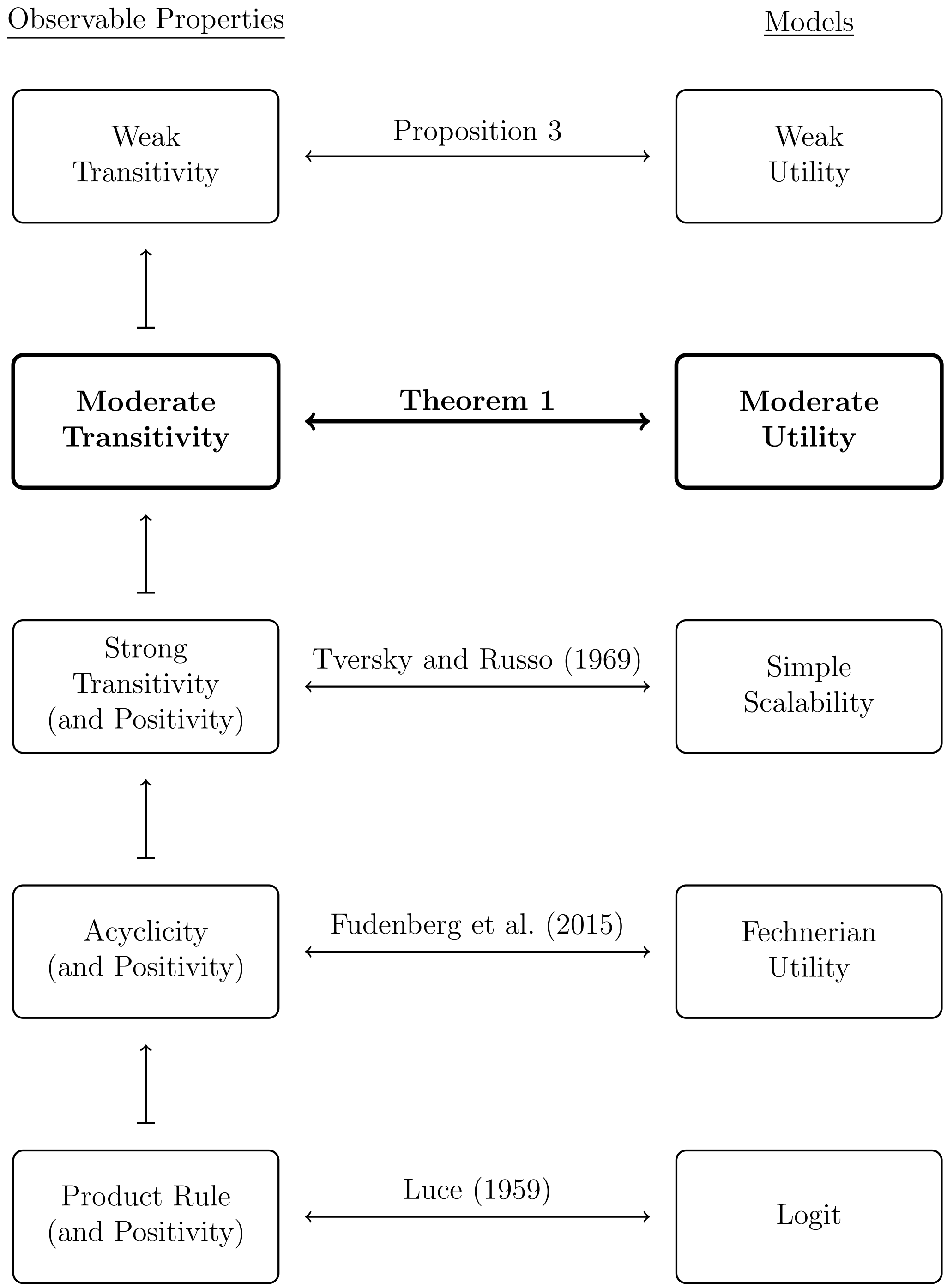


Figure 1: Relationship between models and postulates on choice probabilities for binary stochastic choice over a finite set of options. A double arrow ($\leftrightarrow$) indicates equivalence while an arrow ($\mapsto$) indicates implication in the direction of the arrow and failure of implication in the opposite direction.

Formula (2) is a special case of the *Fechnerian utility model* from psychophysics (Fechner, 1859; Debreu, 1958; Davidson and Marschak, 1959) where

$$\rho(i, j) = F(u(i) - u(j)) \tag{3}$$

for some utility function $u : Z \to \mathbb{R}$ and a strictly increasing $F : \mathbb{R} \to (0, 1)$. Fudenberg et al. (2015) show that (3) is equivalent, under the assumption of positivity, to the postulate of *acyclicity*. This postulate rules out cycles of the form $\rho(w^i, x^i) \geq \rho(y^i, z^i)$ for all $i = 1, \ldots, n$ with at least one strict inequality, whenever $\{w^i, x^i\} = \{y^{f(i)}, z^{f(i)}\}$ and $w^i = y^{g(i)}$ for some permutations $f, g : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Formula (3) is a special case of *simple scalability* (Krantz, 1964) which requires

$$\rho(i, j) = F(u(i), u(j)) \tag{4}$$

for some utility function u and a real valued function F which is strictly increasing in the first argument and strictly decreasing in the second. Tversky and Russo (1969) showed that simple scalability is equivalent to strong transitivity.

A quick comparison of the formulas shows that (2) $\Rightarrow$ (3) $\Rightarrow$ (4). To see that (4) is a special case MUM, note that strong transitivity immediately implies moderate transitivity, and the result follows from Theorem 1. The failure of the reverse implications is easily seen by examples.

Above, we imposed restrictions on the parameters of the MUM to obtain several special cases in the literature. Conversely, we now relax the triangle inequality property in the distance metric of the MUM to obtain a more general model, and we show this model is equivalent to the weak transitivity postulate.

A *semimetric* on Z is a function $s : Z^2 \to \mathbb{R}_+$ satisfying $s(i, j) = 0$ if and only if $i = j$, and $s(i, j) = s(j, i)$ for all i, j. A semimetric does not need to satisfy the triangle inequality. A choice rule ρ on a finite set Z is a *weak utility model (WUM)* if there is a utility function $u : Z \to \mathbb{R}$, a semimetric $s : Z^2 \to \mathbb{R}_+$ and a strictly increasing function F, such that for all $i \neq j$,

$$\rho(i, j) = F\left(\frac{u(i) - u(j)}{s(i, j)}\right) \tag{WUM}$$

where F satisfies $F(t) = 1 - F(-t)$ for all t.

Proposition 3. *A choice rule ρ is a WUM if and only if it is weakly transitive.*

We prove this result in the Appendix. Proposition 3 appears at the top of Figure 1.

Combining Theorem 1 and Proposition 3, we conclude that the triangle inequality property of the distance metric in the MUM is exactly responsible for the restriction from weakly transitive to moderately transitive choice behavior.

5 Relation to random utility models

A choice rule ρ on a finite Z is a *random utility model* (RUM) if there exists a probability measure μ over the strict orderings on Z such that $\rho(i,j)$ equals the probability under μ of the event in which i beats j. Block and Marschak (1959) and Falmagne (1978) characterize the set of RUMs in an abstract setting of choice options when choice data for all finite menus is available. A review of the literature that tackles the characterization of binary choice RUMs is provided by Fishburn (1992). The MUM and RUM families have a non-empty intersection which includes Examples 5 and 6 above. Next, we show that neither MUM nor RUM nest each other.

Example 9. *We modify an example given in de Souza (1983) to obtain a choice rule that is a MUM but not a RUM. Let $Z = \{1, 2, 3, 4, 5, 6\}$, let $0 < \varepsilon < 3/46$ and let the choice rule ρ on Z be given by*

$$\begin{aligned} \rho(4,5) = \rho(4,6) = \rho(2,5) = \rho(2,3) = \rho(1,6) = \rho(1,3) &= 1 - \varepsilon \\ \rho(2,6) = \rho(1,5) &= \frac{1}{2} + \varepsilon \\ \rho(2,4) = \rho(1,4) = \rho(3,5) = \rho(3,6) &= \frac{1}{2} + \frac{\varepsilon}{2} \\ \rho(3,4) = \rho(1,2) = \rho(5,6) &= \frac{1}{2} + \frac{\varepsilon}{3} \end{aligned}$$

It is straightforward to verify that ρ is moderately transitive, and therefore a MUM by Theorem 1. Now suppose ρ is a RUM generated by the probability μ on the set of strict orderings over Z. Since $\rho(2,3) = \rho(4,6) = 1 - \varepsilon$, the probability of the event $\{2 \succ 3\} \cap \{4 \succ 6\}$ is larger or equal to $1 - 2\varepsilon$. By transitivity, the event $\{2 \succ 3\} \cap \{3 \succ 4\} \cap \{4 \succ 6\}$ is contained in the event $\{2 \succ 6\}$. Hence the event $\{3 \succ 4\} \cap \{6 \succ 2\}$ has at most probability 2ε. By the same reasoning, $\{3 \succ 4\} \cap \{5 \succ 1\}$ has at most probability 2ε. And likewise $\{6 \succ 2\} \cap \{5 \succ 1\}$ has at most probability 2ε. Since μ is a probability, this implies $\rho(3,4) + \rho(5,1) + \rho(6,2) \leq 1 + 6\varepsilon$. But instead we have $\rho(3,4) + \rho(5,1) + \rho(6,2) = 3/2 - 5\varepsilon/3 > 1 + 6\varepsilon$ and therefore ρ cannot be a RUM.

A converse example based on the well-known Condorcet paradox shows that RUM models can violate weak transitivity (and hence also moderate transitivity). By Theorem 1 these RUM models are not MUMs. Let μ assign equal probability to three strict orderings $i \succ j \succ k$, $j \succ k \succ i$ and $k \succ i \succ j$ over the options i, j and k. Then the binary choice rule ρ generated by μ has $\rho(i,j) = \rho(j,k) = \rho(k,i) = 2/3$ which violates weak transitivity. Similarly, some recent models proposed in the random choice literature including the random consideration set rule (Manzini and Mariotti, 2014), the attribute rule (Gul et al., 2014), the single-crossing random utility rule (Apesteguia et al., 2017), the deliberately stochastic choice rule (Cerreia-Vioglio et al., 2019) and the focus-then-compare procedure (Ravid and Steverson, 2019) can be easily verified to violate weak transitivity and, therefore, Theorem 1 implies their binary choice restrictions are not nested by MUM.

A Proofs

Proof of Proposition 1

Let $Z = \{1, 2, \ldots, n\}$ be the finite set of alternatives. Consider the set of weakly transitive choice rules ρ on Z with $\rho(i,j) \geq 1/2$ whenever $i \leq j$ and for which the set $\{\rho(i,j) \in [0,1] : i \neq j\}$ has maximum cardinality with $n(n-1)$ elements. Each such ρ induces a strict ordering $\succ_\rho$ of the $n(n+1)/2$ pairs $P_n := \{(i,j) : 1 \leq i < j \leq n\}$ given by $(i,j) \succ_\rho (k,\ell)$ if and only if $\rho(i,j) > \rho(k,\ell)$. This set of choice rules ρ induces $\text{Weak}(n) = [n(n-1)/2]!$ different strict orderings $\succ_\rho$ on P_n.

Moderate transitivity allows Moderate(n) different strict orderings over P_n. Now consider the addition of alternative $n+1$ to the set Z.

Lemma 2. *Moderate*$(n+1) \leq [n(n-1)/2+1]^n$ *Moderate*(n)

Proof. Take a single strict ordering over P_n compatible with moderate transitivity. There are multiple ways to extend this strict ordering to incorporate the new pairs $(1, n+1)$, $(2, n+1)$, $\ldots$, $(n, n+1)$ and obtain a strict ordering over P_{n+1} that is still moderately transitive. Since the original ordering has $n(n-1)/2$ pairs, there are $n(n-1)/2+1$ different positions to include $(n, n+1)$. In this way we obtain $n(n-1)/2+1$ different strict orderings, all of which respect moderate transitivity. The total number of strict orderings over $P_n \cup \{(n, n+1)\}$ that satisfy moderate transitivity is therefore $[n(n-1)/2+1]\,\text{Moderate}(n)$. Now we take one such strict ordering and extend it to incorporate a second pair $(n-1, n+1)$. This pair can in principle be added into $n(n-1)/2 + 2$ different positions, but placing it in the very last position would violate moderate transitivity, since it requires $\rho(n-1, n+1) > \min\{\rho(n-1, n), \rho(n, n+1)\}$. The total number of strict orderings over $P_n \cup \{(n, n+1), (n-1, n+1)\}$ which satisfy moderate transitivity must therefore be smaller or equal to $[n(n-1)/2+1]^2\,\text{Moderate}(n)$. A simple inductive argument completes the proof. $\square$

Lemma 3. $\lim_{n\to\infty}\left[\prod_{k=1}^{n}\frac{n(n-1)/2+k}{n(n-1)/2+1}\right] = e$

Proof. The result can be shown by verifying that, for each n,

$$\left(1+\frac{1}{n}\right)^{n-1} \leq \left[\prod_{k=1}^{n}\frac{n(n-1)/2+k}{n(n-1)/2+1}\right] \leq \left(1+\frac{1}{n}\right)^{n}$$

and taking the limit as $n \to \infty$. We leave the details to the reader. $\square$

Lemma 2 implies that

$$\begin{aligned}\frac{\text{Moderate}(n+1)}{\text{Weak}(n+1)} &\leq \frac{\text{Moderate}(n)}{\text{Weak}(n)}\frac{[n(n-1)/2]!}{[n(n+1)/2]!}[n(n-1)/2+1]^n \\ &= \frac{\text{Moderate}(n)}{\text{Weak}(n)}\left[\prod_{k=1}^{n}\frac{n(n-1)/2+1}{n(n-1)/2+k}\right]\end{aligned}$$

and by Lemma 3 the last expression in brackets goes to $1/e$ when n goes to infinity, where $e \approx 2.718$ is the base of the natural logarithm. Hence for all sufficiently large n the ratio $\text{Moderate}(n+1)/\text{Weak}(n+1)$ is less than half the ratio $\text{Moderate}(n)/\text{Weak}(n)$, completing the proof.

Finally, we prove the additional claim that

$$\lim_{n\to\infty} \text{Strong}(n)/\text{Moderate}(n) = 0.$$

The probability $\rho(1, n)$ must be the highest in every strongly transitive ρ. For each strict ordering satisfying strong transitivity, there exist at least $n-2$ strict orderings which violate strong transitivity but satisfy moderate transitivity: for each $k = 2, 3, \ldots, n-1$ change the value of $\rho(1, n)$ to $\max\{\rho(1, k), \rho(k, n)\} - \varepsilon$ for $\varepsilon > 0$ sufficiently small. It follows that each resulting ranking violates strong transitivity. To see that moderate transitivity still holds, note that every inequality required by strong transitivity holds, except those involving $\rho(1, n)$. In addition, strong transitivity implies that for each

$k, j = 2, \ldots, n-1$, $\max\{\rho(1,k), \rho(k,n\}) > \min\{\rho(1,j), \rho(j,n\})$ hence for ε small we have $\rho(1,n) > \min\{\rho(1,j), \rho(j,n)\}$. Thus, Strong($n$)/Moderate($n$) $\leq 1/(n-1) \to 0$ when $n \to \infty$. $\square$

Proof of Theorem 1

Necessity is shown in the main text. For sufficiency, suppose ρ satisfies moderate transitivity. In particular, ρ satisfies weak transitivity, and hence, by letting $x \succcurlyeq y$ if and only if $\rho(x,y) \geq 1/2$, we obtain a complete and transitive relation $\succcurlyeq$ over the finite set of options Z. The relation $\succcurlyeq$ induced by ρ divides the n alternatives in Z into $k \leq n$ indifference classes. Therefore, there exists a utility function $u : Z \to \{1, \ldots, k\}$ that is onto and represents $\succcurlyeq$, that is, $u(x) \geq u(y)$ if and only if $x \succcurlyeq y$ if and only if $\rho(x,y) \geq 1/2$.

Let $Y := \{\{x,y\} \subset Z : \rho(x,y) \neq 1/2\}$, and let m be the cardinality of the set $\{|\rho(x,y) - 1/2| : \{x,y\} \in Y\}$. Partition the set Y into m disjoint sets $Y_1 \cup Y_2 \cup \cdots \cup Y_m = Y$ such that for any two pairs $\{w,x\}$ and $\{y,z\}$ in Y we have $\{w,x\} \in Y_i$ and $\{y,z\} \in Y_j$ with $i \geq j$ if and only if $|\rho(w,x) - 1/2| \leq |\rho(y,z) - 1/2|$. Thus, the pairs in Y_1 have the highest value of $|\rho(x,y) - 1/2|$, while the pairs in Y_m have the lowest value of $|\rho(x,y) - 1/2|$ in Y.

The result is trivial when Z has $n \leq 2$ alternatives so suppose $n \geq 3$. Define a constant $C = (n-1)^{[n(n-1)/2+1]} > 0$ and define the sequence $D_1, D_2, \ldots, D_m$ by:

$$D_1 = 0; D_j = (n-1)^{j-2} \text{ for } j = 2, \ldots, m.$$

Let $d : Z \times Z \to [0, \infty)$ be defined as follows:

$$d(x,y) = \begin{cases} 0, & \text{if } x = y \\ C, & \text{if } x \neq y \text{ and } \rho(x,y) = 1/2 \\ (C/2 + D_j)\,|u(x) - u(y)|\,, & \text{if } \{x,y\} \in Y_j \end{cases} \tag{5}$$

From (5) it is immediate that d satisfies (i) $d(x,y) \geq 0$; (ii) $d(x,y) = 0$ if and only if $x = y$; and (iii) $d(x,y) = d(y,x)$ for all $x, y \in Z$. To show d is a metric, it remains to verify the triangle inequality: $d(x,z) \leq d(x,y) + d(y,z)$. The inequality trivially holds when any two options among x, y, z are equal. Consider three distinct options $x, y, z \in Z$.

Case 1: $u(x) = u(y) = u(z)$. By the definition of u we have $\rho(x,y) = \rho(y,z) =$

$\rho(x,z) = 1/2$. By the definition of d we have $d(x,z) = C < 2C = d(x,y) + d(y,z)$.

Case 2: $u(x) \neq u(y) = u(z)$. The definitions of u and d imply

$$\begin{aligned} d(x,y) + d(y,z) - d(x,z) &= (C/2 + D_i)\,|u(x) - u(y)| + C - (C/2 + D_j)\,|u(x) - u(z)| \\ &= (D_i - D_j)\,|u(x) - u(z)| + C \\ &\geq -(n-1)^{m-2}(n-1) + C \\ &= (n-1)^{[n(n-1)/2+1]} - (n-1)^{m-1} \\ &> 0 \end{aligned}$$

where the last inequality follows from the fact that we defined m to be the cardinality of $\{|\rho(x,y) - 1/2| : \{x,y\} \in Y\}$ which is smaller or equal to $n(n-1)/2$.

Case 3: $u(y) \neq u(x) = u(z)$. The definitions of u and d imply

$$\begin{aligned} d(x,y) + d(y,z) - d(x,z) &= (C/2 + D_i)\,|u(x) - u(y)| + (C/2 + D_j)\,|u(y) - u(z)| - C \\ &= (C + D_i + D_j)\,|u(y) - u(z)| - C \\ &\geq 0. \end{aligned}$$

Case 4: $u(z) \neq u(x) = u(y)$. Same argument as Case 2.

Case 5: $u(x) > u(y) > u(z)$. By the definition of u we have $\{x,y\} \in Y_i$, $\{y,z\} \in Y_j$, and $\{x,z\} \in Y_\ell$, for some i, j, ℓ. The definition of d implies

$$\begin{aligned} d(x,y) + d(y,z) - d(x,z) &= (C/2 + D_i)\,|u(x) - u(y)| + (C/2 + D_j)\,|u(y) - u(z)| \\ &\quad - (C/2 + D_\ell)\,|u(x) - u(y) + u(y) - u(z)| \\ &= (D_i - D_\ell)\,|u(x) - u(y)| + (D_j - D_\ell)\,|u(y) - u(z)| \end{aligned}$$

The definition of u implies $\rho(x,y) > 1/2$ and $\rho(y,z) > 1/2$. By moderate transitivity we have either $\rho(x,y) = \rho(y,z) = \rho(x,z)$ or $\rho(x,z) > \min\{\rho(x,y), \rho(y,z)\}$. The first case implies $D_i = D_j = D_\ell$ and therefore $d(x,y) + d(y,z) - d(x,z) = 0$. The second case implies $D_\ell < \max\{D_i, D_j\}$. If $D_\ell \leq \min\{D_i, D_j\}$ then both $(D_i - D_\ell)$ and $(D_j - D_\ell)$ are positive and the desired inequality holds. It remains to show the inequality holds when $\min\{D_i, D_j\} < D_\ell < \max\{D_i, D_j\}$, which implies

$$\begin{aligned}
d(x,y) + d(y,z) - d(x,z) &\geq (\max\{D_i, D_j\} - D_\ell)\, 1 + (\min\{D_i, D_j\} - D_\ell)\,(n-2) \\
&\geq (n-1)^{\ell-1} - (n-1)^{\ell-2} + [0 - (n-1)^{\ell-2}](n-2) \\
&= 0.
\end{aligned}$$

Case 6: $u(x) > u(z) > u(y)$. By the definition of u we have $\{x, y\} \in Y_i$, $\{y, z\} \in Y_j$, and $\{x, z\} \in Y_\ell$, for some i, j, ℓ. The definition of d implies

$$\begin{aligned}
d(x,y) + d(y,z) - d(x,z) &= (C/2 + D_i)\,[u(x) - u(z) + u(z) - u(y)] \\
&\quad + (C/2 + D_j)\,[u(z) - u(y)] - (C/2 + D_\ell)\,[u(x) - u(z)] \\
&= (D_i - D_\ell)\,[u(x) - u(z)] + (C + D_i + D_j)\,[u(z) - u(y)] \\
&\geq \left(0 - (n-1)^{m-2}\right)(n-2) + (C + 0 + 0)\, 1 \\
&= -(n-1)^{m-1} + (n-1)^{m-2} + (n-1)^{n(n-1)/2+1} \\
&> 0.
\end{aligned}$$

Case 7: $u(y) > u(x) > u(z)$. By the definition of u we have $\{x, y\} \in Y_i$, $\{y, z\} \in Y_j$, and $\{x, z\} \in Y_\ell$, for some i, j, ℓ. The definition of d implies

$$\begin{aligned}
d(x,y) + d(y,z) - d(x,z) &= (C/2 + D_i)\,[u(y) - u(x)] \\
&\quad + (C/2 + D_j)\,[u(y) - u(x) + u(x) - u(z)] \\
&\quad - (C/2 + D_\ell)\,[u(x) - u(z)] \\
&= (C + D_i + D_j)\,[u(y) - u(x)] + (D_j - D_\ell)\,[u(x) - u(z)] \\
&> 0.
\end{aligned}$$

Case 8: $u(y) > u(z) > u(x)$. Similarly to Case 7,

$$d(x,y) + d(y,z) - d(x,z) = (C + D_i + D_j)\,[u(y) - u(z)] + (D_i - D_\ell)\,[u(z) - u(x)] > 0.$$

Case 9: $u(z) > u(x) > u(y)$. Similarly to Cases 7 and 8,

$$d(x,y) + d(y,z) - d(x,z) = (C + D_i + D_j)\,[u(x) - u(y)] + (D_j - D_\ell)\,[u(z) - u(x)] > 0.$$

Case 10: $u(z) > u(y) > u(x)$. Since $d(x,y) + d(y,z) \leq d(x,z)$ if and only if $d(y,x) + d(z,y) \leq d(z,x)$, the inequality follows from the same argument as in Case 5.

By Cases 1 to 10 above, d satisfies the triangle inequality and is a metric. Now we show u and d provide an ordinal representation for ρ,

$$\rho(w,x) \geq \rho(y,z) \text{ if and only if } \frac{u(w)-u(x)}{d(w,x)} \geq \frac{u(y)-u(z)}{d(y,z)}. \tag{6}$$

First, $\rho(w,x) \geq \rho(y,z) > 1/2$ if and only if $\rho(w,x) > 1/2$, $\rho(y,z) > 1/2$, and $|\rho(w,x) - 1/2| \geq |\rho(y,z) - 1/2|$; if and only if $u(w) > u(x)$, $u(y) > u(z)$, $d(w,x) = (C/2 + D_i)[u(w) - u(x)]$, $d(y,z) = (C/2 + D_j)[u(y) - u(z)]$, and $i \leq j$; if and only if

$$\frac{u(w)-u(x)}{d(w,x)} = \frac{1}{C/2 + D_i} \geq \frac{1}{C/2 + D_j} = \frac{u(y)-u(z)}{d(y,z)} > 0.$$

Second, $1/2 > \rho(w,x) \geq \rho(y,z)$ if and only if $\rho(z,y) \geq \rho(x,w) > 1/2$ and the desired inequality follows from the step above. Finally, $\rho(w,x) \geq 1/2 \geq \rho(y,z)$ if and only if $u(w) - u(x) \geq 0 \geq u(y) - u(z)$ if and only if

$$\frac{u(w)-u(x)}{d(w,x)} \geq 0 \geq \frac{u(y)-u(z)}{d(y,z)}$$

hence the ordinal representation (6) holds. Finding a strictly increasing F such that the cardinal representation (1) holds is then straightforward and left to the reader. □

Proof of Proposition 3

Necessity is straightforward. To show sufficiency, suppose the binary choice rule ρ on a finite set Z satisfies weak transitivity. As in the proof of Theorem 1, weak transitivity implies there is a utility function $u : Z \to \{1, \ldots, m\}$ representing the complete and transitive binary relation given by $i \succcurlyeq j$ if and only if $\rho(i,j) \geq 1/2$. Fix k, ℓ with $\rho(k,\ell) = \max_{i,j \in Z} \rho(i,j)$. Define $s(k,\ell) = 1$ and define $F : [u(\ell) - u(k), u(k) - u(\ell)] \to \mathbb{R}$ by

$$F(t) = \frac{1}{2} + t\left(\frac{\rho(k,\ell) - 1/2}{u(k) - u(\ell)}\right).$$

Finally, for each $i, j \in Z$ define

$$s(i,j) = \frac{u(i) - u(j)}{F^{-1}(\rho(i,j))}.$$

and it is easy to verify that u, s, F represent ρ. $\square$

References

Apesteguia, Jose and Miguel A Ballester, "Monotone stochastic choice models: The case of risk and time preferences," *Journal of Political Economy*, 2018, *126* (1), 74–106.

_ , _ , **and Jay Lu**, "Single-Crossing Random Utility Models," *Econometrica*, 2017, *85* (2), 661–674.

Block, H.D. and Jacob Marschak, "Random Orderings and Stochastic Theories of Response," Cowles Foundation Discussion Papers 66, Cowles Foundation for Research in Economics, Yale University 1959.

Cerreia-Vioglio, Simone, David Dillenberger, Pietro Ortoleva, and Gil Riella, "Deliberately stochastic," *American Economic Review*, July 2019, *109* (7), 2425–45.

Chambers, Christopher, Yusufcan Masatlioglu, and Collin Raymond, "Weighted Linear Discrete Choice," 2021.

Chipman, John S., "Stochastic Choice and Subjective-Probability," *Econometrica*, 1958, *26* (4), 613–613.

Chipman, John S, "Stochastic choice and subjective probability," in Dorothy Willner, ed., *Decisions, Values, and Groups*, Vol. I, Pergamon Press, 1960, pp. 70–95.

Coombs, Clyde H, Marshall Greenberg, and J Zinnes, "A double law of comparative judgment for the analysis of preferential choice and similarities data," *Psychometrika*, 1961, *26* (2), 165–171.

Davidson, Donald and Jacob Marschak, "Experimental tests of a stochastic decision theory," *Measurement: Definitions and theories*, 1959, *17*, 274.

de Souza, Fernando Menezes Campello, "Mixed models, random utilities, and the triangle inequality," *Journal of Mathematical Psychology*, 1983, *27* (2), 183 – 200.

Debreu, Gerard, "Stochastic Choice and Cardinal Utility," *Econometrica*, 1958, *26* (3), 440–444.

Domencich, Thomas A. and Daniel McFadden, *Urban travel demand: a behavioral analysis*, Vol. 93 of *Contributions to Economic Analysis*, North-Holland, 1975.

Falmagne, J. C., "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 1978, *18* (1), 52 – 72.

Fechner, Gustav Theodor, *Elemente der Psychophysik*, Leipzig: Breitkopf & Hartel, 1859.

Fishburn, Peter C, "Induced binary probabilities and the linear ordering polytope: A status report," *Mathematical Social Sciences*, 1992, *23* (1), 67–80.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki, "Stochastic Choice and Revealed Perturbed Utility," *Econometrica*, November 2015, *83* (6), 2371–2409.

Georgescu-Roegen, Nicholas, "The pure theory of consumers behavior," *The Quarterly Journal of Economics*, 1936, *50* (4), 545–593.

_ , "Threshold in Choice and the Theory of Demand," *Econometrica*, 1958, pp. 157–168.

Gul, Faruk, Paulo Natenzon, and Wolfgang Pesendorfer, "Random Choice as Behavioral Optimization," *Econometrica*, September 2014, *82* (5), 1873–1912.

Halff, Henry M, "Choice theories for differentially comparable alternatives," *Journal of Mathematical Psychology*, 1976, *14* (3), 244–246.

Hausman, Jerry A. and David A. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 1978, *46* (2), pp. 403–426.

He, Junnan and Paulo Natenzon, "Random Choice and Differentiation," 2022.

Hotelling, Harold, "Stability in Competition," *The Economic Journal*, March 1929, *39* (153), 41–57.

Krantz, David Harold, "The scaling of small and large color differences." PhD dissertation, University of Pennsylvania 1964.

Lea, Amanda M and Michael J Ryan, "Irrationality in mate choice revealed by túngara frogs," *Science*, 2015, *349* (6251), 964–966.

Luce, R. Duncan, *Individual Choice Behavior: a Theoretical Analysis*, Wiley New York, 1959.

— **and Patrick Suppes**, "Preference, utility, and subjective probability," in R. Duncan Luce, R. R. Bush, and E. H. Galanter, eds., *Handbook of Mathematical Psychology*, Vol. 3, New York: Wiley, 1965, chapter 19, pp. 249–410.

Manzini, Paola and Marco Mariotti, "Stochastic Choice and Consideration Sets," *Econometrica*, May 2014, *82* (3), 1153–1176.

Marschak, Jacob, "Binary-choice constraints and random utility indicators," in K. J. Arrow, S. Karlin, and P. Suppes, eds., *Mathematical methods in the social sciences*, Stanford University Press 1960, pp. 312–329.

McFadden, Daniel, "Conditional logit analysis of qualitative choice behavior," in Paul Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press, 1974, chapter 4, pp. 105–42.

— , "Modeling the choice of residential location," Technical Report 673, Transportation Research Record 1978.

Mellers, Barbara A, Shi jie Chang, Michael H Birnbaum, and Lisa D Ordonez, "Preferences, prices, and ratings in risky decision making.," *Journal of Experimental Psychology: Human Perception and Performance*, 1992, *18* (2), 347.

Natenzon, Paulo, "Random Choice and Learning," *Journal of Political Economy*, February 2019, *1* (127), 419–457.

Ravid, Doron and Kai Steverson, "Focus, Then Compare," *SSRN Working Paper*, March 2019.

Restle, Frank, *Psychology of judgment and choice: A theoretical essay.*, Wiley, 1961.

Rieskamp, Jörg, Jerome R. Busemeyer, and Barbara A. Mellers, "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice," *Journal of Economic Literature*, September 2006, *44* (3), 631–661.

Rubinstein, Ariel, "Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?)," *Journal of economic theory*, 1988, *46* (1), 145–153.

Salop, Steven C, "Monopolistic competition with outside goods," *The Bell Journal of Economics*, 1979, pp. 141–156.

Soltani, Alireza, Benedetto De Martino, and Colin Camerer, "A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence," *PLoS computational biology*, 2012, *8* (7), e1002607.

Thurstone, L. L., "A law of comparative judgment," *Psychological Review*, 1927, *34* (4), 273.

Tversky, A., "Choice by elimination," *Journal of Mathematical Psychology*, 1972, *9* (4), 341–367.

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

Tversky, Amos, "Features of similarity.," *Psychological review*, 1977, *84* (4), 327.

_ **and J. Edward Russo**, "Substitutability and similarity in binary choices," *Journal of Mathematical Psychology*, 1969, *6* (1), 1–12.