UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Topics in Voting and Choice Theory

Permalink
https://escholarship.org/uc/item/8fj6j7mr

Author
Brady, Richard Lee

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Topics in Voting and Choice Theory

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Economics

by

Richard Lee Brady

Committee in charge:

Professor Christopher P. Chambers, Chair
Professor Mark J. Machina
Professor Craig R. M. McKenzie
Professor David A. Schkade
Professor Joel Sobel

2016

The Dissertation of Richard Lee Brady is approved and is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
                                                    Chair

University of California, San Diego

2016

DEDICATION

To Jess, Mom, Dad, Lauren, Carol, and Leo

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

2005	Bachelor of Science, Christian Brothers University
2006	Master of Science, University of Memphis
2011	Master of Arts, Duke University
2016	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Huesch, Marco D., and Richard Brady. "Allowing repeat winners." *Judgment and Decision Making* 5.5 (2010): 374.

Brady, Richard L., and Christopher P. Chambers. "Spatial implementation." *Games and Economic Behavior* 94 (2015): 200-205.

Brady, Richard, and John Rehbeck. "Menu-Dependent Stochastic Feasibility." *Econometrica*, forthcoming.

Brady, Richard L., and Christopher P. Chambers. "A Spatial Analogue of May's Theorem." *Social Choice and Welfare*, forthcoming.

ABSTRACT OF THE DISSERTATION

Topics in Voting and Choice Theory

by

Richard Lee Brady

Doctor of Philosophy in Economics

University of California, San Diego, 2016

Professor Christopher P. Chambers, Chair

This dissertation contains three essays on choice and voting theory. In Chapter 1, a model of stochastic choice is proposed and characterized. Randomness in choice is driven by variability in the availability of alternatives. Chapter 2 addresses the classic implementation problem in a spatial setting. A social choice rule is recommended that can be viewed as a spatial analogue to majority rule as originally axiomatized by May (1952). Chapter 3 investigates testability, identification, and estimation in a large class of models used in modeling voting behavior. The models have relevance to the growing number of online voting platforms.

Chapter 1

Menu-Dependent Stochastic Feasibility

Abstract: We examine the role of stochastic feasibility in consumer choice using a *random conditional choice set rule* (RCCSR) and uniquely characterize the model from conditions on stochastic choice data. Feasibility is modeled to permit correlation in availability of alternatives, which provides a natural way to examine substitutability/complementarity. We show that an RCCSR generalizes the *random consideration set rule* of Manzini and Mariotti (2014). We then relate this model to existing literature. In particular, an RCCSR is not a random utility model.

1.1 Introduction

We investigate the role of stochastic feasibility in consumer choice. Consider a researcher with scanner data on a consumer's purchases from repeated visits to a grocery store. In addition, the store supplies the researcher with the list of offered alternatives. However, there is random variation of alternatives that are *available* to consumers that is unknown to the researcher. For example, the researcher may not know if a delivery is delayed, food is spoiled, or some alternatives are out of stock.[1] In each case, a rational consumer's choices will depend on the available alternatives. Therefore, random variation in feasibility causes a rational consumer's choices to appear stochastic to the researcher. Hence, stochastic feasibility induces a stochastic choice function.[2]

The events mentioned above may cause correlation in availability of alternatives. For example, a delivery truck carrying meat and dairy may be delayed, a disease can spoil certain fruits, and stock-outs may depend on similar products being offered. When feasibility is driven by stock-outs, correlation provides a natural way to discuss substitutability/complementarity.[3] For example, we say two alternatives are substitutes if there is negative correlation in feasibility because one alternative is less likely to be available in the presence of the other.[4]

We model stochastic feasibility using a *Random Conditional Choice Set Rule* (RCCSR). An RCCSR assumes the agent has deterministic preferences while feasibility is driven by an exogenous stochastic process. In particular, the proba-

---

[1]This could also be a limited attention model where the researcher does not know the subset of alternatives the consumer considered.

[2]We are grateful to Doron Ravid for suggesting this interpretation.

[3]See Manzini, Mariotti, and Ülkü (2015) for a discussion on the appeal of using correlation to identify substitutes/complements.

[4]If feasibility is driven by consideration, correlation captures substitutability/complementarity of their consideration.

bility of a particular set being feasible is conditioned on the offered menu. This feature permits correlation in availability which facilitates discussion of substitutability/complementarity. We model the possibility that the feasible set is empty with a default option. We show that an RCCSR is uniquely characterized from conditions on stochastic choice data (Theorem 1.3.1). Further, we demonstrate how an RCCSR generalizes the *random consideration set rule* of Manzini and Mariotti (2014) (henceforth MM) and provide a new characterization (Theorem 1.3.3).

The rest of the paper proceeds as follows: Section 1.2 introduces notation and defines an RCCSR. Section 1.3 uniquely characterizes an RCCSR using conditions on stochastic choice data. Section 1.4 demonstrates how an RCCSR differs from existing models.

1.2 Definitions and notation

Let X be a non-empty finite set of alternatives and $\mathcal{D}$ a domain of menus which are subsets of X. We assume that the domain satisfies the following richness condition: $\{a, b\} \in \mathcal{D}$ for all distinct $a, b \in X$ and $B \in \mathcal{D}$ whenever $A \in \mathcal{D}$ and $B \subseteq A$.[5] Let the *default option* be $x^* \notin X$. The default option is available for each menu and can be interpreted as choosing nothing or not choosing from a particular class of alternatives.[6] We use the notation $X^* = X \cup \{x^*\}$ and $A^* = A \cup \{x^*\}$ for all $A \in \mathcal{D}$.

Definition 1.2.1. A *random choice rule* is a map $P : X^* \times \mathcal{D} \to [0, 1]$ such that: for all $A \in \mathcal{D}$, $\sum_{a \in A^*} P(a, A) = 1$; for all $a \notin A^*$, $P(a, A) = 0$; and for all $A \in \mathcal{D} \setminus \emptyset$,

---

[5]This domain assumption captures two important special cases: classical stochastic choice framework and classical binary stochastic choice.

[6]Outside of an experimental study, it may be difficult to observe a consumer "choosing" nothing. This can be ameliorated if one is interested in consumer choice within a class of alternatives. For example, if the researcher is concerned about the purchase of fruit, the default option could be interpreted as "did not buy fruit".

for all $a \in A^*$, $P(a, A) \in (0, 1)$.

In the above definition, $P(a, A)$ is the probability that alternative a is chosen from A^*. When the menu is empty, the default option x^* is always chosen, so $P(x^*, \emptyset) = 1$. For all $A \in \mathcal{D}$ and $B \subseteq A^*$, we denote $P(B, A) = \sum_{b \in B} P(b, A)$.

We investigate the behavior of an agent whose preferences are given by a strict total ordering $\succ$ on X.[7] For any $A \in \mathcal{D}$, we denote the set of feasible alternatives as $F(A) \subseteq A$. We call $F(A)$ the *feasible set*. An agent's choice is made by maximizing $\succ$ over alternatives in $F(A)$. We allow $F(A)$ to be empty, in which case the agent chooses the default option x^*. Therefore, $P(x^*, A)$ is the probability that $F(A)$ is empty.

For a *random conditional choice set rule* (RCCSR), we consider a full support probability distribution π on $\mathcal{D}$. Thus, there is a positive probability each $A \in \mathcal{D}$ is feasible.[8] When $\mathcal{D} = 2^X$, $\pi(A)$ represents the probability that A is feasible in X. For a menu A, the probability of facing the feasible set $B \subseteq A$ is

$$Pr(F(A) = B) = \frac{\pi(B)}{\sum_{C \subseteq A} \pi(C)}.$$

If B is not a subset of A, then $Pr(F(A) = B) = 0$. Thus, the probability of facing a given feasible set is conditioned on the offered menu.[9] For a menu $A \in \mathcal{D}$ and $a \in A$, let $A_a = \{B \subseteq A \mid a \in B \text{ and } \forall b \in B \setminus \{a\}\ a \succ b\}$. A_a is the set of subsets of A where a is the most preferred alternative. We now formally define an RCCSR.

---

[7] A strict total ordering is an asymmetric, transitive, and weakly connected binary relation. A binary relation $\succ$ on a set X is asymmetric if for all $x, y \in X$, $x \succ y$ implies that $y \succ x$ does not hold. The relation $\succ$ is transitive if for all $x, y, z \in X$, $[x \succ y$ and $y \succ z]$ implies $x \succ z$. The relation $\succ$ is weakly connected if for all $a, b \in X$ such that $a \neq b$, then $a \succ b$ or $b \succ a$.

[8] We find this condition reasonable for the feasibility interpretation. However, this is hardly defensible for consideration sets. We refer the reader to Appendix C of the Supplemental Material for a model where an agent considers at most a pair of alternatives.

[9] We discuss an alternative type of conditioning in Appendix B of the Supplemental Material.

Definition 1.2.2. A *random conditional choice set rule* (RCCSR) is a random choice rule $P_{\succ,\pi}$ for which there exists a pair $(\succ, \pi)$, where $\succ$ is a strict preference ordering on X and $\pi : \mathcal{D} \to (0,1)$ a full support probability distribution over $\mathcal{D}$, such that for all $A \in \mathcal{D}$ and for all $a \in A$

$$P_{\succ,\pi}(a, A) = \frac{\sum_{B \in A_a} \pi(B)}{\sum_{C \subseteq A} \pi(C)}.$$

Thus, $P_{\succ,\pi}(a, A)$ is the probability that a is the best feasible alternative when offered menu A. Menu-dependence is clear since $Pr(F(A) = B)$ is conditioned on the subsets of the offered menu. Further, an RCCSR incorporates correlation in availability of alternatives.

We now define the *random consideration set rule* of Manzini and Mariotti (2014) (MM) which we re-characterize in Section 1.3.3.[10]

Definition 1.2.3. A *random consideration set rule* is a random choice rule $P_{\succ,\gamma}$ for which there exists a pair $(\succ, \gamma)$, where $\succ$ is a strict preference ordering on X and γ is a map $\gamma : X \to (0,1)$, such that for all $A \in \mathcal{D}$ and for all $a \in A$ that

$$P_{\succ,\gamma}(a, A) = \gamma(a) \prod_{b \in A : b \succ a} (1 - \gamma(b)).$$

The random consideration set rule is a simple model with only $|X|$ parameters which represent how likely an object is considered. Setting $\pi(A) = \prod_{b \in X \setminus A}(1 - \gamma(b)) \prod_{a \in A} \gamma(a)$ gives $P_{\succ,\pi} = P_{\succ,\gamma}$. Hence, a random consideration set rule is a special case of an RCCSR.

---

[10]Horan (2014) provides a characterization of a random consideration set rule without a default alternative. See Section 1.3.1 for a discussion on difficulties with removing the default option.

1.3 Characterization

1.3.1 Revealed preference and limited data

The revealed preference relation of our model is based on a sequential independence condition. We say that alternative b is *sequentially independent* from alternative a in menu $A \in \mathcal{D}$ for menus $|A| \geq 2$ denoted bI_Aa if

$$P(b, A) = P(b, A \setminus \{a\})P(A^* \setminus \{a\}, A).$$

Assuming an agent faces random feasible sets and has a deterministic preference $\succ$ with $a \succ b$, then b will be chosen only if a is not available. Thus, it seems reasonable that the agent chooses b independent of a not being available. However, the term $P(A^* \setminus \{a\}, A)$ is the probability a is not available in A. Thus, sequential independence is the case described. In contrast, the most preferred option is chosen when available with any other alternatives. Hence, removal of a sub-optimal alternative may cause non-independent changes to the choice probability of a.

We define the revealed preference relation $\succ$ by $a \succ b$ if and only if bI_Aa for some menu A with $a, b \in A$. In contrast, the revealed preference relation $\tilde{\succ}$ of MM is given by $a \tilde{\succ} b$ if and only if $P(b, A) < P(b, A \setminus \{a\})$ for some menu A, so the revealed preference relation $\succ$ implies $\tilde{\succ}$.

Upon rearranging, one sees that *sequential independence* is a hazard rate condition. For example, bI_Aa if and only if the probability of choosing b in $A \setminus \{a\}$ is the hazard rate

$$P(b, A \setminus \{a\}) = \frac{P(b, A)}{1 - P(a, A)}.$$

We see that the probability b is chosen from the set $A \setminus \{a\}$ is the same as the

probability b is chosen from A conditional on a being sold out. This relaxes the "stochastic path independence" of a random consideration set choice rule in MM.[11]

Now suppose we observe stochastic choice data of all alternatives from only two menus $A, B \in \mathcal{D}$ generated by an RCCSR. What can we infer about π and $\succ$? Suppose $B = A \setminus \{b\}$ for some alternative $b \in A$. Then, we can determine b's rank relative to all alternatives in A. To see this, note that any $c \in A \setminus \{b\}$ satisfying

$$\frac{P(c, A \setminus \{b\})}{P(c, A)} = \frac{P(x^*, A \setminus \{b\})}{P(x^*, A)}$$

must also satisfy $b \succ c$. This is because the choice frequencies of all goods inferior to b change by the same proportion as the change in choice frequency of x^* once b is removed from the menu. All alternatives a such that the equality does not hold satisfy $a \succ b$. Thus, b's rank among the alternatives is established.

Further, it is possible to find the probability that b is feasible in A since

$$Pr(b \in F(A)) = \frac{\sum\limits_{B \subseteq A | b \in B} \pi(B)}{\sum_{C \subseteq A} \pi(C)} = 1 - \frac{P(x^*, A)}{P(x^*, A \setminus \{b\})}.$$

We can then use that $\sum\limits_{B \subseteq A | b \in B} \pi(B) \leq P(b \in F(A))$ to place bounds on π with limited data.

Now consider an RCCSR when the feasible set must be nonempty, so π is a probability distribution over $\mathcal{D} \backslash \emptyset$. As in MM, an RCCSR lacks unique identification once the default option is removed. For example, let $X = \{a, b\}$ and suppose $P(a, \{a, b\}) = \alpha$ and $P(b, \{a, b\}) = \beta$.[12] This is consistent with $a \succ b$

---

[11] One might also expect similarities between an RCCSR and the regular perception-adjusted Luce model from Echenique, Saito, and Tserenjigmid (2014) which imposes conditions on hazard rates. However, these models differ in many ways which we discuss in Section 1.4 and Appendix D of the Supplemental Material.

[12] Note that $P(a, \{a\}) = P(b, \{b\}) = 1$ in this framework.

and $\pi(\{a\}) + \pi(\{a, b\}) = \alpha$ and $\pi(\{b\}) = \beta$ or with $b \succ a$ and $\pi(\{a\}) = \alpha$ and $\pi(\{b\}) + \pi(\{a, b\}) = \beta$. However, if $\mathcal{D} = 2^X$ we can still identify a revealed preference ordering which is unique up to the two least preferred alternatives. That is, for any distinct $a, b, c \in X$ we can identify the most preferred alternative among them by evoking sequential independence on the menu $\{a, b, c\}$. Whether the default option can be removed by a process similar to Horan (2014) remains an open question.

1.3.2 Characterization of RCCSR

We now characterize an RCCSR using conditions on stochastic choice data.

ASI : (Asymmetric Sequential Independence) For all distinct $a, b \in X$, exactly one of the following holds:

$$a I_{\{a,b\}} b \quad \text{or} \quad b I_{\{a,b\}} a.$$

ASI assumes that the alternatives are asymmetric in sequential independence. The intuition for this condition was argued earlier when discussing the revealed preference relation.

TSI: (Transitive Sequential Independence) For all distinct $a, b, c \in X$,

$$a I_{\{a,b\}} b \quad \text{and} \quad b I_{\{b,c\}} c \quad \Rightarrow \quad a I_{\{a,c\}} c.$$

TSI says that if a is chosen independently when b is not feasible and b is chosen independently when c is not feasible in their respective binary menus, then a is chosen independently when c is not feasible in menu $\{a, c\}$. This condition imposes that the I relation is an ordering over alternatives in binary menus.

ESI: (Expansive Sequential Independence) For all $a \in X$ and all menus

$A, B \in \mathcal{D}$ such that $a \in A \cap B$, if

$$\forall b \in A \setminus \{a\} \quad bI_Aa \quad \text{and} \quad \forall c \in B \setminus \{a\} \quad cI_Ba \quad \Rightarrow \quad \forall d \in A \cup B \setminus \{a\} \quad dI_{A \cup B}a.$$

ESI expands sequential independence from binary to arbitrary menus. It says if an agent chooses alternatives independently when a is not feasible in different menus, then they are still chosen independently when a is not feasible in the union of the menus.

We introduce some new notation for the following condition. For all $A \in \mathcal{D} \setminus \emptyset$, let $O_A = \frac{P(A,A)}{P(x^*,A)}$ be the odds of the feasible set being nonempty in menu A. For the empty set, let $O_\emptyset = 0$. For $A, B \in \mathcal{D}$, we define $\Delta_B O_A = O_A - O_{A \setminus B} = \frac{P(A,A)}{P(x^*,A)} - \frac{P(A \setminus B, A \setminus B)}{P(x^*, A \setminus B)}$. Let $\mathcal{B} = \{B_1, \ldots, B_n\}$ be any collection of sets such that $B_i \in \mathcal{D}$. Let $\Delta_\mathcal{B} O_A = \Delta_{B_n} \ldots \Delta_{B_1} O_A = \Delta_{B_n} \ldots \Delta_{B_2} O_A - \Delta_{B_n} \ldots \Delta_{B_2} O_{A \setminus B_1}$ be the successive marginal differences of feasible odds.

IFO: (Increasing Feasible Odds) For any $A \in \mathcal{D} \setminus \emptyset$, $|A| \geq 2$, and for any finite collection $\mathcal{B} = \{B_1, \ldots, B_n\}$ with $B_i \in \mathcal{D}$,

$$\Delta_\mathcal{B} O_A > 0.$$

IFO states that enlarging the menu decreases the odds the default option is chosen at an increasing rate.[13] Aguiar (2015b) further examines successive difference conditions on choice probabilities to study the role of capacities in stochastic choice. We note this condition is equivalent to a multiplicative version of the Block-Marschak polynomials on the default option. Thus, choice of the default

---

[13]Note this defines a capacity from the odds that the feasible set is non-empty. Making this inequality weak characterizes a model with $\{A \in \mathcal{D} \mid |A| \leq 2\} \subseteq \text{support}(\pi)$. Removing this condition, we would characterize a model where $\pi(\cdot)$ represents set intensities on choice which could be negative.

option behaves as if in a random utility model. In particular, IFO is equivalent to the condition that for all $A \in \mathcal{D}$ such that $|A| \geq 2$,

$$\sum_{B \subseteq A} (-1)^{|A \setminus B|} \prod_{C \subseteq A : C \neq B} P(x^*, C) > 0.$$

One can make other restrictions on how choice probabilities of the default option behave when removing alternatives. For example, if we instead require that the choice frequency of the default option exhibits a menu-independent marginal effect when adding an alternative, we arrive at the random consideration set model of MM (Theorem 1.3.3). We also characterize a model where π has limited support in Appendix B of the Supplemental Material. We now present the main result.

Theorem 1.3.1. *A random choice rule satisfies ASI, TSI, ESI, and IFO if and only if it is an RCCSR $P_{\succ, \pi}$. Moreover, both $\succ$ and π are unique, that is, for any RCCSR with $P_{\succ, \pi} = P_{\succ', \pi'}$ we have that $(\succ, \pi) = (\succ', \pi')$.*

All proofs can be found in Appendix A. We give intuition for showing sufficiency. We first show the revealed preference relation $\succ$ described previously is a strict preference ordering using ASI, TSI, and ESI. Next, we show that the probability of choosing the default option has the an RCCSR representation on the domain. We then prove the representation holds for arbitrary alternatives on singleton and binary menus. Finally, we extended the representation to all other menus via induction. We then define a valid probability distribution π using IFO and a Möbius inversion formula.

We now present a lemma used in the proof of the main result.

Lemma 1.3.1. *If ASI, TSI, and ESI hold, then for any $A \in \mathcal{D}$ such that $|A| \geq 2$ and $\tilde{a} \in A$ such that $\forall b \in A \setminus \{\tilde{a}\}$ $\tilde{a} \succ b$ we have that $x^* I_A \tilde{a}$.*

Lemma 1.3.1 shows that these conditions restrict choice of the default option to satisfy sequential independence. Therefore, a model where the default option is more preferred than some alternative would require a different characterization.

An RCCSR's appeal is being able to exhibit menu-dependent feasibility without assuming menu-dependent parameters. A counterpart to an RCCSR is a model with menu-dependent feasibility parameters. We define a *menu-dependent random conditional choice set rule* as a random choice rule $P_{\succ,\nu}$ for which there exists a pair $(\succ, \nu)$, where $\succ$ is a strict total order on X and ν is a map $\nu : \mathcal{D} \times \mathcal{D} \backslash \emptyset \to (0, 1)$, such that

$$P_{\succ,\nu}(a, A) = \frac{\sum_{B \in A_a} \nu(B, A)}{\sum_{C \subseteq A} \nu(C, A)} \quad \forall A \in \mathcal{D}, \forall a \in A.$$

However, a model with menu-dependent feasibility parameters has no empirical content.

Theorem 1.3.2. *For every strict total order $\succ$ on X and for every random choice rule P, there exists a menu-dependent random conditional choice set rule $P_{\succ,\nu}$ such that $P = P_{\succ,\nu}$.*

1.3.3 Characterization of random consideration set rule

We now obtain the random consideration set rule of Manzini and Mariotti (2014) by replacing IFO with a constant marginal effects condition on the choice probability of the default option.

MIDO : (Menu Independent Default Option) For all $a \in X$ and for all $A, B \in \mathcal{D}$ such that $a \in A \cap B$ then

$$\frac{P(x^*, A \setminus \{a\})}{P(x^*, A)} = \frac{P(x^*, B \setminus \{a\})}{P(x^*, B)}$$

This is MM's i-Independence on the default option. It restricts how the

choice frequency of the default option changes once an alternative is removed from the menu. Specifically, the condition requires the effect to be menu independent. This condition is similar to the independence condition of Luce (1959a) and is one reasonable way to restrict choice of the default option.

Theorem 1.3.3. *A random choice rule satisfies ASI, TSI, ESI, and MIDO if and only if it is a random consideration set rule $P_{\succ,\gamma}$. Moreover, both $\succ$ and γ are unique, that is, for any random consideration set rule with $P_{\succ,\gamma} = P_{\succ',\gamma'}$ we have that $(\succ, \gamma) = (\succ', \gamma')$.*

1.4   Comparison to related models

Although an RCCSR has a strong structure and a rational agent, it allows for deviations from a standard model of choice and permits correlation among feasible alternatives. One could think of using data on feasibility and stated preferences to generate predictions from the model which could be tested against observed consumer choice frequencies. Additionally, since the random consideration set rule is a special case of an RCCSR, an RCCSR can exhibit choice frequency reversals and violations of stochastic transitivity.

We examine the i-Asymmetry condition required for the random considera-tion set rule of MM. i-Asymmetry states that

$$\frac{P(a, A \setminus \{b\})}{P(a, A)} \neq 1 \quad \Rightarrow \quad \frac{P(b, A \setminus \{a\})}{P(b, A)} = 1.$$

This says that if removing b affects the choice probability of a in a menu, then removing a cannot affect the choice probability of b in the same menu. However, it is reasonable that removal of either alternative may affect the other's choice

probability within a menu. We will show that an RCCSR allows violations of i-Asymmetry.

First, return to the story of a researcher following an agent's choices in the introduction. Suppose in addition that the researcher has knowledge of alternatives that were available at the time of choice. It would be reasonable to think correlation exists between which objects are feasible. Correlation would mean that $Pr(a \in F(A) \mid b \in F(A)) \neq Pr(a \in F(A))$ for some $a, b \in A$ with $a \neq b$. We note that a random consideration set rule does not allow these effects. The following example details a situation in which an RCCSR generates choice frequencies which violate i-Asymmetry and alternatives have correlation in availability.

Example 1. (Grocery Store) Consider a researcher with scanner data of a consumer's purchases from several grocery stores. The alternatives of interest are apples (a), bananas (b), and carrots (c). Here the set of alternatives is $X = \{a, b, c\}$ and $\mathcal{D} = 2^X$. Suppose we observe choice from all possible nonempty menus given by Table 1.1.

Table 1.1. Grocery store stochastic choice data

	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	7/20	1/3	1/2	0	1/2	0	0
b	11/20	1/2	0	11/13	0	3/4	0
c	1/20	0	1/4	1/13	0	0	1/2

One can use this data and the revealed preference relation to find that $a \succ b \succ c$ and that the π system is given by

$$\pi(\emptyset) = \frac{1}{20} \quad \pi(\{a\}) = \frac{1}{20} \quad \pi(\{b\}) = \frac{3}{20} \quad \pi(\{c\}) = \frac{1}{20}$$

$$\pi(\{a,b\}) = \frac{1}{20} \quad \pi(\{a,c\}) = \frac{1}{20} \quad \pi(\{b,c\}) = \frac{8}{20} \quad \pi(\{a,b,c\}) = \frac{4}{20}.$$

Looking at the pair a and b, we see that

$$\frac{P(a,\{a,c\})}{P(a,X)} = \frac{10}{7} \quad \text{and} \quad \frac{P(b,\{b,c\})}{P(b,X)} = \frac{20}{13}$$

which is a violation of i-Asymmetry. Here we see that a and b both benefit from the other's removal. Next, suppose that a researcher observes b is available when the agent chooses from X. Now, the researcher can back out the probability that a was also in the feasible set since

$$P(a \in F(X) \mid b \in F(X)) = \frac{\pi(\{a,b\}) + \pi(\{a,b,c\})}{\pi(\{a,b\}) + \pi(\{a,b,c\}) + \pi(\{b,c\}) + \pi(\{b\})} = \frac{5}{16}$$

but $P(a \in F(X)) = 7/20$. As discussed earlier, if menus are subject to stock-outs this may suggest that apples and bananas are substitutes since apples are less likely to be available given bananas are still available.

We now consider how an RCCSR compares to other models in the literature.[14] Block and Marschak (1960) considered a class of stochastic choice functions known as random utility models. A random utility model is described by a probability measure over preference orderings, where the agent selects the maximal alternative available according to the randomly assigned preference ordering. Random utility models obey the regularity condition that $P(a, B) \geq P(a, A)$ for any $a \in B \subseteq A$. However, Example 1 violates this condition as seen by examining the choice probabilities of a in $\{a, b\}$ and X. Therefore, RCCSRs are not nested in random utility models.

---

[14]Like many models of stochastic choice, we do not explicitly include measurement or feasibility errors in our characterization based on choice probabilities. However, it may be interesting to see if choice behavior similar to an RCCSR could be generated by a profit maximizing firm choosing a costly technology which yields a stochastic menu of goods to rational consumers.

Moreover, random utility models are not nested in RCCSRs. To see this, consider the model from Luce (1959a), which is a special case of a random utility model. The Luce model is of the form $P(a, A) = \frac{u(a)}{\sum_{b \in A} u(b)}$ for a strictly positive utility function u and is characterized by the IIA condition.[15] However, an RCCSR will necessarily violate IIA when $A \in \mathcal{D}$ and $|A| \geq 3$ since the ratio of the probability of choosing the most preferred alternative over the probability of choosing the least preferred alternative will necessarily decrease once the middle-ranked alternative is removed from the menu. Lastly, we note that there are models which are both a random utility model and an RCCSR such as the random consideration set rule.

A recent model which appears similar to an RCCSR is the regular perception-adjusted Luce model (rPALM) from Echenique, Saito, and Tserenjigmid (2014). In fact, both an RCCSR and rPALM use conditions on hazard rates to characterize the models. Furthermore, both an RCCSR and rPALM accommodate violations of regularity, IIA, and stochastic transitivity. Nonetheless, an RCCSR and rPALM are distinct.[16] One strong prediction of an RCCSR is that choice frequency of the default alternative decreases as alternatives are added to a menu. However in an rPALM, default alternative choice probabilities need not systematically increase or decrease since they are driven by a menu dependent parameter. An rPALM also requires ratios of hazard rates to be constant across menus and satisfy a regularity condition. Both of these conditions are difficult to interpret, but neither is required in an RCCSR. This suggests several ways to discern which model is appropriate from data.

The recent work of Gul, Natenzon, and Pesendorfer (2014) axiomatizes an attribute rule where the decision maker first randomly chooses an attribute

---

[15]IIA states that $\frac{P(a,A)}{P(b,A)} = \frac{P(a,B)}{P(b,B)}$ for any a, b and menus A, B such that $a, b \in A \cap B$.

[16]In particular, an rPALM cannot generate the choice frequencies exhibited in Example 1 (Appendix D of Supplemental Material).

from all perceived attributes and then randomly selects an alternative containing the selected attribute. Every attribute rule is a random utility model and the Luce model is a special case. Therefore, an RCCSR and an attribute rule are not equivalent from the discussion of random utility models.

There are other works worthy of mention which take a different approach than those mentioned here. The models of Machina (1985), Mattsson and Weibull (2002), and Fudenberg, Iijima, and Strzalecki (2015) assume an agent has deterministic preferences over lotteries and chooses a probability distribution to maximize utility on a menu. Therefore, these models induce "stochastic choice" from deterministic preferences on lotteries. We refer the reader to the survey by Rieskamp, Busemeyer, and Mellers (2006) for a survey of other related works.

Chapter1 will appear in a forthcoming issue of Econometrica and was coauthored with John Rehbeck. The copyright of this article is held by the Econometric Society.

1.5 Appendix A: Main results

We present a series of lemmas which characterize the preference relation, properties on larger menus, and the proof of Lemma 1.3.1. We then present a statement of the Möbius inverse formula used in the proof of Theorem 1.3.1. The proof of Theorem 1.3.1 follows.

Lemma 1.5.1. *If ASI and TSI hold, then there exists a strict total order of X such that for any $a, b \in X$*

$$a \succ b \quad \Leftrightarrow \quad P(b, \{a, b\}) = P(b, \{b\})P(\{b, x^*\}, \{a, b\}).$$

Proof. The relation $\succ$ is asymmetric and weakly connected since by ASI for distinct $a, b \in X$ we have that exactly one of the below is true

$$P(a, \{a, b\}) = P(a, \{a\})P(\{a, x^*\}, \{a, b\})$$

or

$$P(b, \{a, b\}) = P(b, \{b\})P(\{b, x^*\}, \{a, b\}).$$

Suppose for $a, b, c \in X$ that $a \succ b$ and $b \succ c$. By definition we have

$$P(b, \{a, b\}) = P(b, \{b\})P(\{b, x^*\}, \{a, b\})$$

and

$$P(c, \{b, c\}) = P(c, \{c\})P(\{c, x^*\}, \{b, c\}).$$

By TSI we have that $P(c, \{a, c\}) = P(c, \{c\})P(\{c, x^*\}, \{a, c\})$ so by definition of $\succ$ we have $a \succ c$, so that $\succ$ is transitive. $\square$

Lemma 1.5.2. *If ASI, TSI, and ESI hold, then for any menu $A \in \mathcal{D}$ there exists an $\tilde{a} \in A$ such that for all $b \in A \setminus \{\tilde{a}\}$ we have*

$$P(b, A) = P(b, A \setminus \{\tilde{a}\})P(A^* \setminus \{\tilde{a}\}, A).$$

Proof. By Lemma 1.5.1 we know $\succ$ is strict, so for any $A \in \mathcal{D}$ there exists an $\tilde{a}$ such that $\tilde{a} \succ b$ for all $b \in A \setminus \{\tilde{a}\}$. The result obviously holds for binary menus by ASI so assume $|A| = 3$ with $A = \{\tilde{a}, b, c\}$. By definition of $\succ$ we know

$$P(b, \{\tilde{a}, b\}) = P(b, \{b\})P(\{b, x^*\}, \{\tilde{a}, b\}) \text{ and}$$

$$P(c, \{\tilde{a}, c\}) = P(c, \{c\})P(\{c, x^*\}, \{\tilde{a}, c\}).$$

By ESI we have

$$P(b, \{\tilde{a}, b, c\}) = P(b, \{b, c\})P(\{b, c, x^*\}, \{\tilde{a}, b, c\})$$

with an analogous statement for c. For $|A| > 3$ the result holds by induction. $\square$

Proof of Lemma 1.3.1. Lemma 1.5.2 established the existence of a maximal alternative in any menu, so let $A \in \mathcal{D}$ with $|A| \geq 2$ and let $\tilde{a} \in A$ be the maximal alternative. Using some basic algebra and the sequential independence result from

Lemma 1.5.2 we have that

$$P(x^*, A) = 1 - P(\tilde{a}, A) - \sum_{b \in A \setminus \{\tilde{a}\}} P(b, A)$$

$$= 1 - P(\tilde{a}, A) - \sum_{b \in A \setminus \{\tilde{a}\}} P(b, A \setminus \{\tilde{a}\}) P(A^* \setminus \{\tilde{a}\}, A)$$

$$= 1 - P(\tilde{a}, A) - P(A^* \setminus \{\tilde{a}\}, A) \sum_{b \in A \setminus \{\tilde{a}\}} P(b, A \setminus \{\tilde{a}\})$$

$$= 1 - P(\tilde{a}, A) - P(A^* \setminus \{\tilde{a}\}, A)(1 - P(x^*, A \setminus \{\tilde{a}\}))$$

$$= 1 - P(\tilde{a}, A) - P(A^* \setminus \{\tilde{a}\}, A) + P(x^*, A \setminus \{\tilde{a}\}) P(A^* \setminus \{\tilde{a}\}, A)$$

$$= P(x^*, A \setminus \{\tilde{a}\}) P(A^* \setminus \{\tilde{a}\}, A).$$

$\square$

Möbius inversion has been used in economics since Shapley (1953). In particular, the result of Falmagne (1978) that the Block-Marschak polynomials are sufficient for a random utility model was proved by Fiorini (2004) using Möbius inversion. In general, it is a powerful tool to move between two functions when there is a partial order. Here we use the partial order over sets. We now present a version of Möbius inversion from Shafer (1976).

Theorem 1.5.1. *(Möbius inversion (Shafer, 1976)) If Θ is a finite set with f and g functions on 2^Θ then*

$$f(A) = \sum_{B \subseteq A} g(B)$$

for all $A \subseteq \Theta$ if and only if

$$g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B)$$

for all $A \subseteq \Theta$.

Using the expression from Section 1.3.1, this will hold for an RCCSR for $\Theta = X$ with $g(A) = \pi(A)$ and $f(A) = \frac{P(x^*, X)}{P(x^*, A)}$.

Corollary 1.5.1. *If $P = P_{\succ, \pi}$ is a RCCSR with $\mathcal{D} = 2^X$ then for all $A \subseteq 2^X$ we have that*

$$\frac{P(x^*, X)}{P(x^*, A)} = \sum_{B \subseteq A} \pi(B)$$

For the proof of the main result, we have that $\mathcal{D}$ may not be the power set. However, this will affect the above intuition by changing only a scaling factor. We now present the proof of the main result.

Proof of Theorem 1.3.1 . That an RCCSR satisfies ASI, TSI, ESI, and IFO is simple to check and is omitted here.

Now, suppose $|X| = N \geq 1$ and P is a random choice rule that satisfies ASI, TSI, ESI, and IFO. From Lemma 1.5.1 and $\mathcal{D}$ rich, we can define an ordering $\succ$ on X which is a total order. We want to show that the $P(\cdot, \cdot)$ is an RCCSR. We prove the representation inductively on menu size. Let $M = \max_{A \in \mathcal{D}} |A|$ be the largest order of sets in $\mathcal{D}$. Let $D_M = \operatorname{argmax}_{A \in \mathcal{D}} |A|$ be the elements of $\mathcal{D}$ with maximal order. First, define $\lambda : \mathcal{D} \to \mathbb{R}$ such that for $A \in \mathcal{D}$ we have that

$$\lambda_A = \lambda(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \frac{1}{P(x^*, B)}.$$

Note that this imposes $\lambda_\emptyset = \frac{1}{P(x^*, \emptyset)} = 1$. Moreover, we have that for $A = \{a\}$ for $a \in X$ that

$$\lambda_a = \frac{1}{P(x^*, \{a\})} - 1 > 0$$

since $P(x^*, \{a\}) \in (0, 1)$ by definition of a random choice rule. For $A \in \mathcal{D}$ with

$|A| \geq 2$ since IFO is equivalent to positivity of the multiplicative polynomial

$$\sum_{B \subseteq A}(-1)^{|A \setminus B|} \prod_{C \subseteq A: C \neq B} P(x^*, C) > 0 \Leftrightarrow \sum_{B \subseteq A}(-1)^{|A \setminus B|} \frac{1}{P(x^*, B)} > 0 \Leftrightarrow \lambda(A) > 0$$

by dividing the polynomial by $\prod_{C \subseteq A} P(x^*, C)$.

Using the Möbius inversion formula Theorem 1.5.1 gives us that

$$\frac{1}{P(x^*, A)} = \sum_{B \subseteq A} \lambda_B.$$

We first examine the choice of the default option option from any menu $A \in \mathcal{D}$. Here

$$\begin{aligned} P(x^*, A) &= \left(\frac{1}{P(x^*, A)}\right)^{-1} \\ &= \lambda_\emptyset \left(\sum_{B \subseteq A} \lambda_B\right)^{-1} \\ &= \frac{\lambda_\emptyset}{\sum_{B \subseteq A} \lambda_B}. \end{aligned}$$

Now, singleton menus are in $\mathcal{D}$ by richness. Thus, focusing on singleton menus $A = \{a\}$ for $a \in X$ the above result can be used to show that

$$\begin{aligned} P(a, \{a\}) &= 1 - P(x^*, \{a\}) \\ &= 1 - \frac{\lambda_\emptyset}{\lambda_\emptyset + \lambda_{\{a\}}} \\ &= \frac{\lambda_{\{a\}}}{\lambda_\emptyset + \lambda_{\{a\}}} \end{aligned}$$

Next, binary menus are in $\mathcal{D}$ by richness, so suppose that the menu is binary i.e. $A = \{a, b\}$ for $a, b \in X$ and that $a \succ b$ without loss of generality. By definition

of $\succ$ we have $P(b, \{a, b\}) = P(b, \{b\})P(\{b, x^*\}, \{a, b\})$ and by Lemma 1.3.1 that $P(x^*, \{a, b\}) = P(x^*, \{b\})P(\{b, x^*\}, \{a, b\})$. Combining these two gives us

$$\frac{P(b, \{a, b\})}{P(x^*, \{a, b\})} = \frac{P(b, \{b\})P(\{b, x^*\}, \{a, b\})}{P(x^*, \{b\})P(\{b, x^*\}, \{a, b\})} = \frac{P(b, \{b\})}{P(x^*, \{b\})}.$$

However, after a simple rearrangement we have that

$$\frac{P(b, \{a, b\})}{P(b, \{b\})} = \frac{P(x^*, \{a, b\})}{P(x^*, \{b\})} = \frac{\left(\sum_{B \subseteq \{a,b\}} \lambda_B\right)^{-1}}{\left(\sum_{B \subseteq \{b\}} \lambda_B\right)^{-1}} = \frac{\sum_{B \subseteq \{b\}} \lambda_B}{\sum_{B \subseteq \{a,b\}} \lambda_B}$$

This relates the ratios of probabilities to the weight function λ defined earlier. Moreover, using this result and Lemma 1.5.2 and Lemma 1.3.1 it is clear that for any menu A and alternative b which is not the maximal element $\tilde{a}$ of $\succ$ in A, we have that

$$\frac{P(b, A)}{P(b, A \setminus \{\tilde{a}\})} = \frac{\sum_{B \subseteq A \setminus \{\tilde{a}\}} \lambda_B}{\sum_{B \subseteq A} \lambda_B}. \tag{1.1}$$

Using this and the earlier result from singleton menus it follows that

$$\begin{aligned}
P(b, \{a, b\}) &= P(b, \{b\})\frac{P(b, \{a, b\})}{P(b, \{b\})} \\
&= \left(\frac{\lambda_{\{b\}}}{\sum_{B \subseteq \{b\}} \lambda_B}\right)\left(\frac{\sum_{B \subseteq \{b\}} \lambda_B}{\sum_{B \subseteq \{a,b\}} \lambda_B}\right) \\
&= \frac{\lambda_{\{b\}}}{\sum_{B \subseteq \{a,b\}} \lambda_B}.
\end{aligned}$$

Next, examining the choice probability of a, the most preferred alternative in $\{a, b\}$,

$$P(a, \{a, b\}) = 1 - P(b, \{a, b\}) - P(x^*, \{a, b\})$$

$$= 1 - \frac{\lambda_{\{b\}}}{\sum_{B \subseteq \{a, b\}} \lambda_B} - \frac{\lambda_{\emptyset}}{\sum_{B \subseteq \{a, b\}} \lambda_B}$$

$$= \frac{\lambda_{\{a\}} + \lambda_{\{a, b\}}}{\sum_{B \subseteq \{a, b\}} \lambda_B}.$$

Summarizing, we have for any singleton or binary menu A and $a \in A$ that

$$P(a, A) = \frac{\sum_{B \in A_a} \lambda_B}{\sum_{B \subseteq A} \lambda_B}$$

where as before $A_a = \{B \subseteq A \mid a \in B \text{ and } \forall b \in B \backslash \{a\} \ a \succ b\}$. Assume inductively that the representation holds for menus of size $m - 1 < M$ in $\mathcal{D}$. Let $A \in \mathcal{D}$ be a menu such that $|A| = m$. Recall $B \in \mathcal{D}$ for all $B \subseteq A$ from richness. Thus, for all $a \in A$ then $P(\cdot, A \backslash \{a\})$ satisfies the representation. From Lemma 1.5.2 we have that there is a unique maximal element $\tilde{a} \in A$. Therefore, for any $b \in A \backslash \{\tilde{a}\}$ we have

$$P(b, A) = P(b, A \backslash \{\tilde{a}\}) \left(\frac{P(b, A)}{P(b, A \backslash \{\tilde{a}\})} \right)$$

$$= \left(\frac{\sum_{B \in (A \backslash \{\tilde{a}\})_b} \lambda_B}{\sum_{B \subseteq A \backslash \{\tilde{a}\}} \lambda_B} \right) \left(\frac{\sum_{B \subseteq A \backslash \{\tilde{a}\}} \lambda_B}{\sum_{B \subseteq A} \lambda_B} \right)$$

$$= \frac{\sum_{B \in A_b} \lambda_B}{\sum_{B \subseteq A} \lambda_B}.$$

where the second equality follows by the induction hypothesis and (1.1) and the third equality follows because $A_b = (A \backslash \{\tilde{a}\})_b$ since $\tilde{a} \succ b$. Now examining the

choice probability of $\tilde{a}$ in A we see that

$$
\begin{aligned}
P(\tilde{a}, A) &= 1 - \sum_{b \in A^* \setminus \{\tilde{a}\}} P(b, A) \\
&= 1 - \sum_{b \in A^* \setminus \{\tilde{a}\}} \frac{\sum_{B \in A_b} \lambda_B}{\sum_{B \subseteq A} \lambda_B} \\
&= \frac{\sum_{B \in A_{\tilde{a}}} \lambda_B}{\sum_{B \subseteq A} \lambda_B}
\end{aligned}
$$

where the second equality follows by the previous result and the third equality follows since $\sum_{B \subseteq A} \lambda_B - \sum_{b \in A^* \setminus \{\tilde{a}\}} \sum_{B \in A_b} \lambda_B = \sum_{B \in A_{\tilde{a}}} \lambda_B$.

We now have the appropriate definition but λ is not necessarily a probability. We define $\tilde{\lambda} : \mathcal{D} \to \mathbb{R}$ by $\tilde{\lambda}(A) = \frac{\lambda(A)}{\sum_{B \in \mathcal{D}} \lambda(B)}$. Thus, we have that

$$
\sum_{A \in \mathcal{D}} \tilde{\lambda}_A = \sum_{A \in \mathcal{D}} \frac{\lambda(A)}{\sum_{B \in \mathcal{D}} \lambda(B)} = 1
$$

and for $A \in \mathcal{D}$ that $\tilde{\lambda_A} \in (0, 1)$. Therefore, $\tilde{\lambda}$ forms a valid full support probability distribution on $\mathcal{D}$ which is related to λ by a constant.

Therefore, we have that P is an RCCSR with $\succ$ as defined from Lemma 1.5.1 and $\tilde{\lambda} = \pi$.

To show that $(\succ, \pi)$ is unique, suppose that there exists a $(\succ', \pi')$ such that $P_{\succ, \pi} = P_{\succ', \pi'}$. First, we note for singleton menus $\{a\} \in \mathcal{D}$ that

$$
\frac{P_{\succ, \pi}(a, \{a\})}{P_{\succ, \pi}(x^*, \{a\})} = \frac{\pi(\{a\})}{\pi(\emptyset)} = \frac{\pi'(\{a\})}{\pi'(\emptyset)} = \frac{P_{\succ', \pi'}(a, \{a\})}{P_{\succ', \pi'}(x^*, \{a\})} \quad \Rightarrow \quad \pi'(\emptyset) = \pi(\emptyset) \frac{\pi'(\{a\})}{\pi(\{a\})}.
$$

This means that $\pi'(\emptyset) = \alpha \pi(\emptyset)$ for $\alpha > 0$ and $\pi'(\{a\}) = \alpha \pi(\{a\})$ for any singleton

menu. Then, since $\succ' \neq \succ$ we know that there exist $a \succ b$ and $b \succ' a$ so

$$
\begin{aligned}
P_{\succ,\pi}(a, \{a, b\}) &= \frac{\pi(\{a\}) + \pi(\{a, b\})}{\pi(\{a, b\}) + \pi(\{a\}) + \pi(\{b\}) + \pi(\emptyset)} \\
&= \frac{\pi'(\{a\})}{\pi'(\{a, b\}) + \pi'(\{a\}) + \pi'(\{b\}) + \pi'(\emptyset)} = P_{\succ',\pi'}(a, \{a, b\}).
\end{aligned}
$$

However, cross multiplying equations, using the scale relation, and eliminating variables

$$
\pi'(\{a, b\})(\pi(\{a\}) + \pi(\{a, b\})) + \alpha(\pi(\{b\}) + \pi(\emptyset))\pi(\{a, b\}) = 0.
$$

This is a contradiction since all of the quantities are positive. Therefore we have that $\succ = \succ'$. The uniqueness of π follows immediately since $\succ$ is uniquely defined then $\pi' = \alpha\pi$. However, for π' to be a probability requires $\alpha = 1$. Therefore, the pair $(\succ, \pi)$ is unique for each RCCSR. $\qquad\square$

Proof of Theorem 1.3.2. The proof follows immediately from Theorem 2 in Manzini and Mariotti (2014) by letting $\nu(B, A) = \prod_{b \in B} \delta(b, A) \prod_{a \in A \setminus B}(1 - \delta(a, A))$ for δ defined as in the proof of Manzini and Mariotti (2014). $\qquad\square$

Proof of Theorem 1.3.3. That a random consideration set rule satisfies ASI, TSI, ESI, and MIDO is simple to check and is omitted here.

Now, suppose $|X| = N \geq 1$ and P is a random choice rule that satisfies ASI, TSI, ESI, and MIDO. By Lemma 1.5.1 and $\mathcal{D}$ rich, we can define an ordering $\succ$ on X which is a total order. Let $M = \max_{A \in \mathcal{D}} |A|$ be the largest order of sets in $\mathcal{D}$. Let $D_M = \operatorname{argmax}_{A \in \mathcal{D}} |A|$ be the elements of $\mathcal{D}$ with maximal order. We want to show that the $P(\cdot, \cdot)$ is a random consideration set rule. We prove the representation inductively on menu size.

For all $a \in X$ we define $\lambda_a = \lambda(a) = P(a, \{a\})$. Now examining the choice of the default option in any menu $A \in \mathcal{D}$ and any alternative $a \in A$, we have by MIDO that

$$\frac{P(x^*, A \setminus \{a\})}{P(x^*, A)} = \frac{P(x^*, \emptyset)}{P(x^*, a)} = \frac{1}{P(x^*, a)} \implies P(x^*, A) = P(x^*, \{a\})P(x^*, A \setminus \{a\}).$$

Since the above argument was for a generic alternative and menu, we have

$$P(x^*, A) = \prod_{a \in A} P(x^*, \{a\}) = \prod_{a \in A}(1 - P(a, \{a\})) = \prod_{a \in A}(1 - \lambda_a)$$

For singleton menus, the representation trivially holds. Next we examine the case of choice in binary menus. For $a, b \in X$ we suppose without loss of generality that $a \succ b$. By ASI and $\mathcal{D}$ rich , we have

$$
\begin{aligned}
P(b, \{a, b\}) &= P(b, \{b\})P(\{x^*, b\}, \{a, b\}) \\
&= P(b, \{b\})\frac{P(x^*, \{a, b\})}{P(x^*, \{b\})} \\
&= \lambda_b \frac{(1 - \lambda_b)(1 - \lambda_a)}{(1 - \lambda_b)} \\
&= \lambda_b(1 - \lambda_a).
\end{aligned}
$$

Where the second equality is by Lemma 1.3.1 and the third equality is by the representation of the choice of the default option in terms of λ. Then for the best

alternative a,

$$P(a, \{a, b\}) = 1 - P(b, \{a, b\}) - P(x^*, \{a, b\})$$

$$= 1 - \lambda_b(1 - \lambda_a) - (1 - \lambda_a)(1 - \lambda_b)$$

$$= \lambda_a.$$

Now assume that the representation holds for all sets of size $m - 1 < M$ so if $|A| < m$

$$P(a, A) = \lambda_a \prod_{b \in A | b \succ a} (1 - \lambda_b).$$

For the a menu A with $|A| = m$ we have by Lemma 1.5.2 that there is a maximal element $\tilde{a} \succ b$ for all $b \in A$. Now looking at the choice of $b \in A$ such that $\tilde{a} \succ b$ we have that

$$P(b, A) = P(b, A \setminus \{a\})P(A^* \setminus \{\tilde{a}\}, A)$$

$$= P(b, A \setminus \{a\})\frac{P(x^*, A)}{P(x^*, A \setminus \{\tilde{a}\})}$$

$$= \left(\lambda_b \prod_{c \in A \setminus \{\tilde{a}\} | c \succ b} (1 - \lambda_c) \right) (1 - \lambda_{\tilde{a}})$$

$$= \lambda_b \prod_{c \in A | c \succ b} (1 - \lambda_c).$$

Where the first equality follows from Lemma 1.5.2 and definition of $\succ$, the second equality follows from Lemma 1.3.1, and the last equality follows since $\tilde{a} \succ b$. Now

examining the choice of $\tilde{a}$, we have that

$$
\begin{aligned}
P(\tilde{a}, A) &= 1 - P(A^* \setminus \{\tilde{a}\}, A) \\
&= 1 - \frac{P(x^*, A)}{P(x^*, A \setminus \{\tilde{a}\})} \\
&= 1 - (1 - \lambda_{\tilde{a}}) \\
&= \lambda_{\tilde{a}}.
\end{aligned}
$$

Therefore, P is a random consideration set rule with preference $\succ$ and attention parameters $\gamma(a) = \lambda(a)$. Where for all $a \in X$ we have $\gamma(a) \in (0, 1)$ since P is a random choice rule and $\gamma(a) = P(a, \{a\})$. That this representation is unique follows immediately from Theorem 1 in Manzini and Mariotti (2014). $\qquad\square$

Appendix B contains a discussion on an alternative form of conditioning for feasible set probabilities. Appendix C replaces IFO with a condition which restricts the support of π to sets of at most two elements. Appendix D examines differences between the regular Perception Adjusted Luce Model and an RCCSR.

1.6 Appendix B: Alternative conditioning

Recall that throughout our analysis, we have considered a model where the probability of obtaining a feasible set is given by

$$
Pr(F(A) = B) = \frac{\pi(B)}{\sum_{C \subseteq A} \pi(C)}.
$$

However, there are other ways that one could condition to obtain a feasible set. In particular, one could consider the model given by

$$Pr(F(A) = B) = \sum_{C \in \mathcal{D}: C \cap A = B} \pi(C),$$

where the default option is chosen if $B = \emptyset$. This alternative conditioning formula is used in Barberà and Grodal (2011) to characterize a preference for flexibility over menus.

We prefer the conditioning formula used in an RCCSR for two main reasons. First, suppose that a feasible set is generated by what items an agent considers from a menu. In this case, an RCCSR says an agent first looks at the menu, then considers a set of alternatives from the menu, and lastly makes a choice. If we used the alternative conditioning formula, it will change the timing of these actions. In particular, the alternative formulation says an agent first considers a set of alternatives, then looks at the menu and further restricts the considered objects, and finally makes a choice.[17] Therefore, in this alternative formulation an agent could be thinking of a better/worse alternative when choosing from the menu. The alternative formulation also seems ill suited for the case of general feasibility. For example, it would seem surprising that the probability an alternative is out of stock in a menu depends on alternatives not offered.

Secondly, we prefer the formulation used in an RCCSR for its identifiability and flexibility. The alternative formulation makes it difficult to identify π completely. In addition, this alternative conditioning formula produces choice probabilities consistent with a random utility model.

---

[17]Using the "in the mood" interpretation, an RCCSR conditioning says a consumer sees the menu and draws a random mood which is consistent with the offered alternatives. In the alternative formulation, the consumer receives a mood before looking at the offered menu.

1.7 Appendix C: Binary support

We can also characterize some models which have limited support by replacing IFO with other conditions. Here we still assume that $\mathcal{D}$ is rich.

BIFO : (Binary Increasing Feasible Odds) For all distinct $a, b \in X$,

$$\Delta_a \Delta_b O_{\{a,b\}} > 0$$

This condition restricts IFO to binary menus. In a consideration set framework, this would mean that adding acceptable alternatives draws consideration away from the default option.

CMD : (Constant Marginal Differences) For all distinct $a, b \in X$ and $A, B \in \mathcal{D}$ with $a, b \in A \cap B$ then

$$\frac{P(a, A)}{P(x^*, A)} - \frac{P(a, A \setminus \{b\})}{P(x^*, A \setminus \{b\})} = \frac{P(a, B)}{P(x^*, B)} - \frac{P(a, B \setminus \{b\})}{P(x^*, B \setminus \{b\})}$$

This condition states that the marginal effect on the odds ratio with respect to the default option of removing an alternative is constant across menus. Replacing IFO with these conditions, we get a model with $|X| + \binom{|X|}{2}$ parameters. We now define a *binary random choice set rule*.

Definition 1.7.1. A *binary random choice set rule* (BRCSR) is a random choice rule $P_{\succ, \alpha}$ for which there exists a pair $(\succ, \alpha)$, where $\succ$ is a strict preference ordering on X and $\alpha : \mathcal{D} \to [0, 1)$ a distribution with $\alpha(A) > 0$ for sets $A \in \mathcal{D}$ with $|A| \leq 2$ and zero otherwise, such that for all $A \in \mathcal{D}$ and for all $a \in A$

$$P_{\succ, \alpha}(a, A) = \frac{\alpha(\{a\}) + \sum_{b \in A | a \succ b} \alpha(\{a, b\})}{\sum_{C \subseteq A : |C| \leq 2} \alpha(C)}.$$

Theorem 1.7.1. *A random choice rule satisfies ASI, TSI, ESI, BIFO, and CMD if and only if it is a BRCSR $P_{\succ,\alpha}$. Moreover, both $\succ$ and α are unique, that is, for any BRCSR with $P_{\succ,\alpha} = P_{\succ',\alpha'}$ we have that $(\succ, \alpha) = (\succ', \alpha')$.*

Proof. Note that a BRCSR satisfies ASI, TSI, ESI, BIFO, and CMD is simple to check and omitted here.

Now, suppose $|X| = N \geq 1$ and P is a random choice rule that satisfies ASI, TSI, ESI, BIFO, and CMD. By Lemma 1.5.1 and $\mathcal{D}$ rich, we can define an ordering $\succ$ on X which is a total order. Let $M = \max_{A \in \mathcal{D}} |A|$ be the largest order of sets in $\mathcal{D}$. Let $D_M = \text{argmax}_{A \in \mathcal{D}} |A|$ be the elements of $\mathcal{D}$ with maximal order. We want to show that the $P(\cdot, \cdot)$ has the BRCSR representation. We prove the representation inductively on menu size.

First, define $\lambda : \mathcal{D} \to \mathbb{R}$ such that for $A \in \mathcal{D}$ we have that

$$\lambda_A = \lambda(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \frac{1}{P(x^*, B)}.$$

This is related to a Möbius inversion formula. Theorem 1.5.1 gives us that

$$\frac{1}{P(x^*, A)} = \sum_{B \subseteq A} \lambda_B.$$

First, note that for singleton menus $\{a\} \in 2^X$ that

$$\lambda_{\{a\}} = \frac{1}{P(x^*, \{a\})} - 1 > 0$$

since $P(x^*, \{a\}) < 1$ by definition of a random choice rule. Next, for binary menus

$\{a, b\} \in \mathcal{D}$ assume without loss of generality that $a \succ b$. Then BIFO implies

$$\Delta_a \Delta_b O_{\{a,b\}} = \sum_{B \subseteq \{a,b\}: B \neq \emptyset} (-1)^{|\{a,b\} \setminus B|} \frac{P(B, B)}{P(x^*, B)}$$

$$= \sum_{B \subseteq \{a,b\}} (-1)^{|\{a,b\} \setminus B|} + \sum_{B \subseteq \{a,b\}: B \neq \emptyset} (-1)^{|\{a,b\} \setminus B|} \frac{P(B, B)}{P(x^*, B)}$$

$$= \sum_{B \subseteq \{a,b\}} (-1)^{|\{a,b\} \setminus B|} \left(1 + \frac{P(B, B)}{P(x^*, B)}\right)$$

$$= \sum_{B \subseteq \{a,b\}} (-1)^{|\{a,b\} \setminus B|} \frac{1}{P(x^*, B)} > 0$$

where we used that $\sum_{B \subseteq \{a,b\}} (-1)^{|\{a,b\} \setminus B|} = \sum_{i=0}^{2} (-1)^i \binom{2}{i} = 0$.

Therefore, we have

$$P(x^*, \{a, b\})^{-1} - P(x^*, \{b\})^{-1} - P(x^*, \{a\})^{-1} + 1 > 0$$

where we used Lemma 3.1 to get to the second equality. Thus we have

$$\lambda_{\{a,b\}} = \sum_{B \subseteq \{a,b\}} (-1)^{|\{a,b\} \setminus B|} \frac{1}{P(x^*, B)} > 0.$$

Now, we show a result on how the λ terms relate to $P(\cdot, \cdot)$ under CMD and then show for all $A \in \mathcal{D}$ such that $|A| \geq 3$ that $\lambda_A = 0$. Note for $A = \{a, b, c\}$ such that $a \succ b$ and $a \succ c$ then we have by CMD that

$$\frac{P(a, A)}{P(x^*, A)} - \frac{P(a, \{a, c\})}{P(x^*, \{a, c\})} = \frac{P(a, \{a, b\})}{P(x^*, \{a, b\})} - \frac{P(a, \{a\})}{P(x^*, \{a\})}.$$

First, looking at the left side of the equality and using Lemma 3.1

$$\frac{P(a, A)}{P(x^*, A)} - \frac{P(a, \{a, c\})}{P(x^*, \{a, c\})} = \frac{1 - P((A \setminus \{a\})^*, A)}{P(x^*, A)} - \frac{1 - P(\{c, x^*\}, \{a, c\})}{P(x^*, \{a, c\})}$$

$$= P(x^*, A)^{-1} - P(x^*, \{b, c\})^{-1}$$

$$- P(x^*, \{a, c\})^{-1} + P(x^*, \{c\})^{-1}.$$

Similarly, looking at the right side of the equality and using Lemma 3.1

$$\frac{P(a, \{a, b\})}{P(x^*, \{a, b\})} - \frac{P(a, \{a\})}{P(x^*, \{a\})} = P(x^*, \{a, b\})^{-1} - P(x^*, \{b\})^{-1} - P(x^*, \{a\})^{-1} + 1.$$

$$= \lambda_{\{a,b\}}$$

Rearranging the equality we see that

$$P(x^*, A)^{-1} = \lambda_{\{a,b\}} + P(x^*, \{b, c\})^{-1} + P(x^*, \{a, c\})^{-1} - P(x^*, \{c\})^{-1}$$

$$= \lambda_{\{a,b\}} + \left(P(x^*, \{b, c\})^{-1} - P(x^*, \{b\})^{-1} - P(x^*, \{c\})^{-1} + 1\right)$$

$$+ P(x^*, \{a, c\})^{-1} + \left(P(x^*, \{b\})^{-1} - 1\right)$$

$$= \lambda_{\{a,b\}} + \lambda_{\{b,c\}} + \lambda_{\{b\}} + P(x^*, \{a, c\})^{-1}$$

$$= \lambda_{\{a,b\}} + \lambda_{\{b,c\}} + \lambda_{\{b\}} + \left(P(x^*, \{a, c\})^{-1} - P(x^*, \{a\})^{-1}\right)$$

$$- P(x^*, \{c\})^{-1} + 1) + \left(P(x^*, \{a\})^{-1} - 1\right) + \left(P(x^*, \{c\})^{-1} - 1\right) + 1$$

$$= \sum_{B \subsetneq A} \lambda_B$$

Therefore, we have that $\frac{1}{P(x^*, A)} = \sum_{B \subsetneq A} \lambda_B$ for all $|A| = 3$. However, using the Möbius inversion result we know that

$$\sum_{B \subsetneq A} \lambda_B = \frac{1}{P(x^*, A)} = \sum_{B \subseteq A} \lambda_B \quad \Rightarrow \quad \lambda_A = 0.$$

Now, suppose that $\frac{1}{P(x^*,A)} = \sum_{B \subsetneq A} \lambda_B$ holds for sets $A \in \mathcal{D}$ with $|A| = m - 1$ and $3 \leq m - 1 < M$. For $A \in \mathcal{D}$ such that $|A| = m$ and $\forall c \in A \setminus \{a, b\}$ such that $a \succ b \succ c$, we have

$$\frac{P(a, A)}{P(x^*, A)} - \frac{P(a, A \setminus \{b\})}{P(x^*, A \setminus \{b\})} = \frac{P(a, \{a, b\})}{P(x^*, \{a, b\})} - \frac{P(a, \{a\})}{P(x^*, \{a\})}.$$

We can perform the same substitutions using Lemma 3.1 as in the three element case so

$$P(x^*, A)^{-1} - P(x^*, A \setminus \{a\})^{-1} - P(x^*, A \setminus \{b\})^{-1} + P(x^*, A \setminus \{a, b\})^{-1} = \lambda_{\{a,b\}}.$$

Since $A \setminus \{a\}$ and $A \setminus \{b\}$ are $m - 1$ element sets, we can use our induction step and then rearrange so

$$P(x^*, A)^{-1} = \lambda_{\{a,b\}} + \sum_{B \subseteq A \setminus \{a\} : |B| \leq 2} \lambda_B + \sum_{B \subseteq A \setminus \{b\} : |B| \leq 2} \lambda_B - \sum_{B \subseteq A \setminus \{a,b\} : |B| \leq 2} \lambda_B$$

$$= \lambda_{\{a,b\}} + \sum_{B \subseteq A \setminus \{a\} : |B| \leq 2} \lambda_B + \sum_{B \subseteq A \setminus \{b\} : |B| = 2 \text{ and } a \in B} \lambda_B + \lambda_{\{a\}}$$

$$= \sum_{B \subseteq A : |B| \leq 2} \lambda_B.$$

We restrict looking at weights λ_B with $|B| \leq 2$ since the inductive step makes other λ terms zero. Performing subtraction of the rightmost terms leads to the second equality. The third equality comes by collecting all terms. Thus, we have that $\frac{1}{P(x^*,A)} = \sum_{B \subseteq A : |B| \leq 2} \lambda_B = \sum_{B \subsetneq A} \lambda_B$ since $\lambda_B = 0$ for all $B \subsetneq A$ with $|B| \geq 3$ by induction. Using the Möbius inversion formula, $\sum_{B \subsetneq A} \lambda_B = \sum_{B \subseteq A} \lambda_B$ so that $\lambda_A = 0$. Therefore, we have shown by induction that $\lambda_A = 0$ for all $A \in \mathcal{D}$ with $|A| \geq 3$. The representation now holds immediately from the proof of Theorem 3.1.

and letting $\alpha = \tilde{\lambda}$.

$\square$

1.8   Appendix D: Comparison to PALM

The regular perception-adjusted Luce model (rPALM) of Echenique, Saito, and Tserenjigmid (2014) is described by a pair $(\succsim_P, u)$ where $\succsim_P$ is a weak order on X and $u : 2^X \to \mathbb{R}$ is a function such that

$$P_{\succsim_P, u}(a, A) = \mu(a, A) \prod_{\alpha \in A/\succsim_P : \alpha \succ_P a} (1 - \sum_{c \in A : c \in \alpha} \mu(c, A))$$

where

$$\mu(a, A) = \frac{u(a)}{\sum_{b \in A} u(b) + u(A)}$$

and

$$u(c) \geq u(\{a, b\}) - u(\{a, b, c\})$$

for all $a, b, c \in X$ with strict inequality with if $b \nsim_P c$.

The notation $A/\succsim_P$ is for the set of equivalence classes according to $\succsim_P$ that partition A, so the product is over all classes of alternatives that are ordered ahead of a. In rPALM, $\succsim_P$ is interpreted as a perception priority relation, and the authors attribute all violations of IIA to perception priority. More specifically, when $a, b \in X$ do not violate IIA, then we have $a \sim_P b$.

One of the distinguising features of an RCCSR relative to an rPALM is that the choice frequency of the default alternative must obey monotonicity with respect to set inclusion under an RCCSR: $B \subset A \Rightarrow P(x^*, B) > P(x^*, A)$. In the context of availability, this restriction is logical in that larger menus are more likely to have an alternative available. An rPALM places no such consistency restrictions

on choice frequency of the default alternative. This is one potential way in which the two models can be distinguished from choice data.

Another feature of rPALM is the hazard rate the authors define as

$$q(a, A) = \frac{P_{\succsim_P, u}(a, A)}{1 - P_{\succsim_P, u}(A^a, A)}$$

where $A^a = \{b \in A : b \succ_P a\}$. The authors impose that the hazard rate obeys both IIA ($\frac{q(a, \{a,b\})}{q(b, \{a,b\})} = \frac{q(a, A)}{q(b, A)}$ for all $a, b \in X$ and $A \subseteq X$ such that $a, b \in A$) and regularity ($q(a, \{a, b\}) \geq q(a, \{a, b, c\})$ for all $a, b, c \in X$ and with strict inequality only when $b \nsim_P c$). We will use this to show that an RCCSR is not a special case of rPALM. It is easy to see that an RCCSR can violate hazard rate IIA (in Example 1 it is violated for a, b). Now consider the choice frequencies in Example 1 and note that we have $P(a, \{a, b, c\}) > P(a, \{a, b\})$ and $P(b, \{a, b, c\}) > P(b, \{a, b\})$. An rPALM is unable to generate these choice frequencies. In what follows, let $A = \{a, b, c\}$.

Case 1: $a \succsim_P b, a \succsim_P c, b \nsim_P c$. By regularity we have

$$P_{\succsim_P, u}(a, A) = q(a, A) < q(a, \{a, b\}) = P_{\succsim_P, u}(a, \{a, b\}).$$

Case 2: $a \succ_P b \sim_P c$. By regularity we have

$$P_{\succsim_P, u}(a, A) = q(a, A) = q(a, \{a, b\}) = P_{\succsim_P, u}(a, \{a, b\}).$$

Case 3: $b \succsim_P a, b \succsim_P c, a \nsim_P c$. By regularity we have

$$P_{\succsim_P, u}(b, A) = q(b, A) < q(b, \{a, b\}) = P_{\succsim_P, u}(b, \{a, b\}).$$

Case 4: $b \succ_P a \sim_P c$. By regularity we have

$$P_{\succsim_P, u}(b, A) = q(b, A) = q(b, \{a, b\}) = P_{\succsim_P, u}(b, \{a, b\}).$$

Case 5: $c \succ_P a \succsim_P b$. By regularity we have

$$P_{\succsim_P, u}(a, A) = q(a, A)(1 - P_{\succsim_P, u}(c, A)) < q(a, A) \leq q(a, \{a, b\}) = P_{\succsim_P, u}(a, \{a, b\}).$$

Case 6: $c \succ_P b \succsim_P a$. By regularity we have

$$P_{\succsim_P, u}(b, A) = q(b, A)(1 - P_{\succsim_P, u}(c, A)) < q(b, A) \leq q(b, \{a, b\}) = P_{\succsim_P, u}(b, \{a, b\}).$$

Case 7: $a \sim_P b \sim_P c$. rPALM cannot violate IIA in this case, but

$$\frac{P(a, A)}{P(b, A)} = \frac{7}{11} \neq \frac{2}{3} = \frac{P(a, \{a, b\})}{P(b, \{a, b\})}$$

in Example 1.

Chapter 2

A Spatial Analogue of May's Theorem

Abstract: In a spatial model with Euclidean preferences, we establish that the *geometric median* satisfies Maskin monotonicity, anonymity, and neutrality. For three agents, it is the unique such rule.

2.1 Introduction

An early social-choice theoretic foundation for majority rule was provided by May (1952). In an environment with a group of agents who choose one of two alternatives based on strict preferences, he shows that majority rule is the unique rule satisfying three natural axioms. The first of these axioms, anonymity, requires that the names of the agents do not matter. The second, neutrality, requires that the names of the alternatives do not matter. The third, positive responsiveness, means that in any given profile, if the collection of agents who prefer the chosen alternative increases with respect to set inclusion, then that alternative should remain chosen.

We work in an environment of Euclidean preferences; that is, policy space is given by Euclidean space, and each agent has a "favorite" alternative. The further a policy is from this favored alternative (the ideal or bliss point), the worse it is for the agent. Our aim is to provide a notion of majority rule in such a context analogous to May's. That is, we investigate this environment axiomatically.

In this paper, we show that a concept called the *geometric median* satisfies the natural counterparts of anonymity, neutrality, and positive responsiveness in a spatial framework. We also establish that, for three agents, it is the unique such rule.[1]

Generalizing anonymity to spatial environments is straightforward. Describing the content of the remaining two axioms (neutrality and positive responsiveness) takes some work. With regard to neutrality, we cannot "rename" alternatives arbitrarily. If we were to do so, the resulting preference profile may not be Eu-

---

[1]Other works using the geometric median in economics or political science research include Cervone, Dai, Gnoutcheff, Lanterman, Mackenzie, Morse, Srivastava, and Zwicker (2012), Baranchuk and Dybvig (2009), and Chung and Duggan (2014). In particular, the latter work describes an interesting generalization of the concept to general convex preferences.

clidean. To this end, we interpret neutrality as an equivariance of the social rule to isometries: it does not matter which coordinate system we use to describe ideal points. This axiom formalizes the idea that the names of alternatives should not matter, and is the counterpart of the classical neutrality axiom in our setting.

Positive responsiveness also has a natural generalization to this environment. Maskin Monotonicity (Maskin (1999)) states that, if the chosen alternative moves up in the ranking of all agents, it remains chosen. This is entirely analogous to May's criterion that if the chosen alternative moves up in everybody's ranking, it remains chosen. Thus, Maskin Monotonicity seems an appropriate generalization of May's positive responsiveness condition to spatial environments.

The geometric median (or medians in the case of an even number of agents) is any point which minimizes the aggregate Euclidean distance to the agents' ideal points. We show that, in general, the geometric median satisfies these three properties. Moreover, in the case of three agents, it is the *unique* such rule. More generally, we do not know whether the geometric median is the unique rule satisfying the properties, but it seems plausible that with a high enough dimension of the underlying Euclidean space, it will be.

2.1.1   Related literature

Early attempts to provide a notion of majority rule in spatial environments (such as Plott (1967)) were guided by generalizing the classic results of Hotelling (1929) and Black (1948) to spatial environments. That is, they took the notion of majority rule core as *given* and tried to compute the corresponding equilibrium concept. In fact, this equilibrium, termed the *median in all directions*, very seldom exists.

A related paper is Duggan (2015), who also provides a spatial generalization

of May's Theorem. His work can be distinguished from ours on several grounds: first, he works with more general single-peaked preferences; but he does so only in one-dimensional environments. Second, instead of working with a social choice rule, he works with a rule which aggregates preferences into a single social preference. Third, instead of a positive responsiveness condition, he utilizes different transitivity conditions.

In an earlier paper (Brady and Chambers, 2015) we show that the geometric median is the smallest rule (with respect to set inclusion) satisfying a host of compelling properties appropriate for a variable population model. Although neutrality and anonymity play no role in the variable population framework, Maskin monotonicity is a central property as in the present paper.

We note that there is a large literature on the relationship between Maskin monotonicity and strategy-proof implementation.[2] However, strategy-proofness is in general incompatible with the notion that the outcome of a rule does not depend on choice of coordinates together with any reasonable anonymity properties. Nonetheless, Maskin monotonicity is known to be necessary for Nash implementability, and we briefly discuss the Nash implementability of the geometric median.

The paper is organized as follows. Section 2 presents the model. Section 3 provides the results. Section 4 concludes.

2.2   Model

Let $X = \mathbb{R}^d$ be the policy space. For any $x, y \in X$ let $\|x - y\|$ denote the Euclidean distance. Let $N = \{1, \ldots, n\}$ be a finite set of agents. Each agent $i \in N$ is equipped with a preference relation $\succsim_i$ (with associated strict

---

[2]See, for example, Muller and Satterthwaite (1977); Dasgupta, Hammond, and Maskin (1979); Barbera and Peleg (1990); Berga and Moreno (2009).

preference $\succ_i$). Preferences are assumed to be Euclidean so that for each $i \in N$, $\succsim_i$ can be represented by an "ideal point", $z_i \in X$, with the property that for any $x, y \in X$, $x \succsim_i y$ if and only if $\|x - z_i\| \leq \|y - z_i\|$. We define an *isometry* as any distance preserving mapping $f : X \to X$ so that for all $x, y \in X$ we have $\|x - y\| = \|f(x) - f(y)\|$.[3]

An aggregation or *social choice rule* is a mapping $\varphi : X^N \to X$. Since there is a one-to-one relationship between preference profiles and a set of points in X, we will use the notation $Z \in X^N$ to indicate a preference profile of the agents that is represented by the points $Z = (z_i)_{i \in N}$ where $z_i \in X$ for each i.

For $i \in N$ and $z_i, x \in X$, let $UC_i(z_i, x) = \{y \in X \mid \|y - z_i\| \leq \|x - z_i\|\}$ be the upper contour set for the preference relation represented by the point z_i at the point x. This is simply the set of all outcomes agent i weakly prefers to x. We will say that the preference relation represented by a point z_i' is a monotonic transformation of the preference relation represented by z_i at a point x if $UC_i(z_i', x) \subseteq UC_i(z_i, x)$. Let $MT(z_i, x)$ be the set of all monotonic transformations of the preference relation represented by the point z_i at the point x and $MT(Z, x)$ be the set of all monotonic transformations of a preference profile represented by the set of points Z at a point x.

For $x, y \in X$ we will denote the line segment with x and y as endpoints by $\overline{xy} = \{t \in X \mid \|x - t\| + \|t - y\| = \|x - y\|\}$.

We say that an outcome $x \in X$ is weakly Pareto efficient (WPE) if there does not exist $y \in X$ such that $y \succ_i x$ for all $i \in N$. The outcome is Pareto efficient (PE) if there does not exist $y \in X$ such that $y \succsim_i x$ for all $i \in N$ and $y \succ_j x$ for at least one $j \in N$. We will say that a social choice rule φ satisfies *Pareto efficiency*

---

[3]Since $X = \mathbb{R}^d$, isometries are bijections that correspond to reflections, rotations, and translations.

if for any $Z \in X^N$, $\varphi(Z)$ is PE.

For an agent i with ideal point $z_i \in X$ we will often use the utility representation $u_i(y) = -(y - z_i) \cdot (y - z_i)$ for $y \in X$ to represent i's preference relation. For a profile of preferences $Z \in X^N$ and utility representations $(u_i(y))_{i \in N}$, consider the sets

$$comp(Z) = \{a = (a_1, \ldots, a_n) \mid \exists y \in X \text{ such that } a_i \leq u_i(y) \text{ for all } i \in N\}$$

and

$$con(Z) = \left\{ x \in X \mid x = \sum_{i=1}^{n} \lambda_i z_i \text{ such that } \lambda_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

The first set is said to be the comprehensive hull of the utility possibility set, where by comprehensive we mean that for any $a \in comp(Z)$ and $b \leq a$ coordinate-wise, then $b \in comp(Z)$. The latter set is simply the convex hull of the set of ideal points. For a set $S \subset \mathbb{R}^d$, we let $bd(S)$ denote its boundary and $int(S)$ denote its interior.

For a set of points $(a_1, \ldots, a_n)$ with each $a_i \in X$, we define a geometric median as a solution to the following minimization problem:

$$\min_{x \in X} \sum_{i=1}^{n} \|x - a_i\|. \tag{2.1}$$

That is, a geometric median minimizes the sum of distances between itself and all of the points.

The geometric median for a finite set of points always exists; further, it is unique if n is odd or if the points $(a_1, \ldots, a_n)$ are not collinear (Haldane (1948)). In the case of collinear points, there could be multiple geometric medians, in particular when the number of distinct points is even. In this case, the geometric median is a

set-valued concept; any selection from this set would result in a single-valued rule.

For simplicity, we assume throughout that n is odd (to ensure uniqueness). In the conclusion, we discuss how our results should be modified in the case of an even number of agents. For a preference profile $Z \in X^N$, we will let x_Z^* denote the geometric median of the ideal points.

Before presenting the axioms, we first state some facts about the geometric median for the case of $n = 3$ and $d = 2$ (see Deimling (2011) p. 325-326 and Coxeter (1989) p. 21-22).[4] Let $A = (a_1, a_2, a_3)$ be the points. First, suppose that the three points are not collinear and that $bd(con(A))$ forms a triangle (denoted as $\triangle_{a_1 a_2 a_3}$). Suppose all interior angles of the triangle are less than $120°$.[5] In this case the geometric median is the unique point in $int(con(A))$ such that the angle between any two line segments connecting the geometric median to the vertices of the triangle a_i and a_j with $i \neq j$ (denoted $\angle_{a_i x_A^* a_j}$) is $120°$. See Figure 2.1. For the special case in which $\triangle_{a_1 a_2 a_3}$ is equilateral, the geometric median will coincide with the intersection of the three medians of the triangle.[6] If the triangle is isosceles, then the geometric median lies on the axis of symmetry. If $\triangle_{a_1 a_2 a_3}$ has an angle that is at least $120°$, then the geometric median corresponds to the obtuse-angled vertex. If the points are collinear then the geometric median is the point lying between the other two or where multiple points are located if the points are not distinct.

We now briefly describe the axioms we will impose on a social choice rule. In the appendix, we show the axioms are independent.

Axiom 2.2.1. A social choice rule φ satisfies *anonymity* if for every bijection

---

[4]The geometric median has a rich history in this special case and is sometimes referred to as the Fermat-Torricelli point of a triangle.

[5]An interior angle is simply the angle formed by two adjacent sides of the triangle

[6]A median of a triangle is any of the line segments connecting a vertex to the midpoint of the opposite side of the triangle
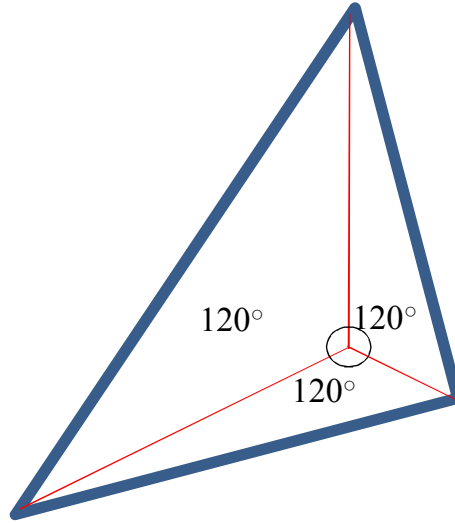
Figure 2.1. The geometric median

$\sigma : N \to N$ and for every $Z \in X^N$ we have $\varphi(z_{\sigma(1)}, \ldots, z_{\sigma(n)}) = \varphi(Z)$.

In words, anonymity states that the outcome from our social choice rule is invariant to changing the names of the agents.

Axiom 2.2.2. A social choice rule φ satisfies *neutrality* if for any isometry f and any $Z \in X^N$ we have $\varphi(f(z_1), \ldots, f(z_n)) = f(\varphi(z_1, \ldots, z_n))$.

Neutrality states that the social choice resulting from any reflection, rotation, or translation of agents' ideal points is the same as applying the reflection, rotation, or translation to the social choice from the untransformed ideal points.

The final axiom says that the social choice is preserved through monotonic transformations.

Axiom 2.2.3. A social choice rule φ satisfies *Maskin monotonicity* or *positive responsiveness* if for all $Z \in X^N$ and for any $Z' \in MT(Z, \varphi(Z))$ we have $\varphi(Z') = \varphi(Z)$.

2.3 Results

In this section, we propose a class of social choice rules that satisfy the axioms discussed in the previous section. We also show that Pareto Efficiency of a social choice rule is a consequence of imposing two of our axioms. Further, Nash-implementation is briefly discussed as we show that our proposed class meets conditions sufficient for such implementation. Finally, we show that this class of social choice rules is the only class satisfying our axioms for the case of $n = 3$.

Let us assume that n is odd. Consider the social choice rule φ such that for any $Z \in X^N$, $\varphi(Z) = x_Z^*$. That is, given any preference profile, the social choice is always the point in X that minimizes the total distance between itself and the agents' ideal points. This choice has a nice appeal in many settings. For example, the choice could be over the location of a supply distribution hub given the location of n factories. Then this rule selects the location that minimizes the total transportation cost between the hub and factories. Or, in a political science context, the choice for a socially optimal candidate would be the location that is minimally far from the policy relevant locations of the set of n voters.

Our first result shows that this rule satisfies our axioms.

Proposition 2.3.1. *Let φ be a social choice rule such that for any $Z \in X^N$, $\varphi(Z) = x_Z^*$. Then φ satisfies anonymity, neutrality, and Maskin monotonicity.*

Before presenting the proof, we first provide two lemmas, whose proofs can be found in Brady and Chambers (2015). The first characterizes monotonic transformations and the second establishes that the geometric median satisfies Maskin monotonicity. Figure 2.2 gives a visual depiction of the results for $d = 2$.

Lemma 2.3.1. *Suppose $z_i, z_i' \in X$ represent two preference relations. For $x \in X$, $z_i' \in MT(z_i, x)$ if and only if $z_i' \in \overline{z_i x}$.*

(a) $z_i' \in MT(z_i, x) \iff z_i' \in \overline{z_i x}$
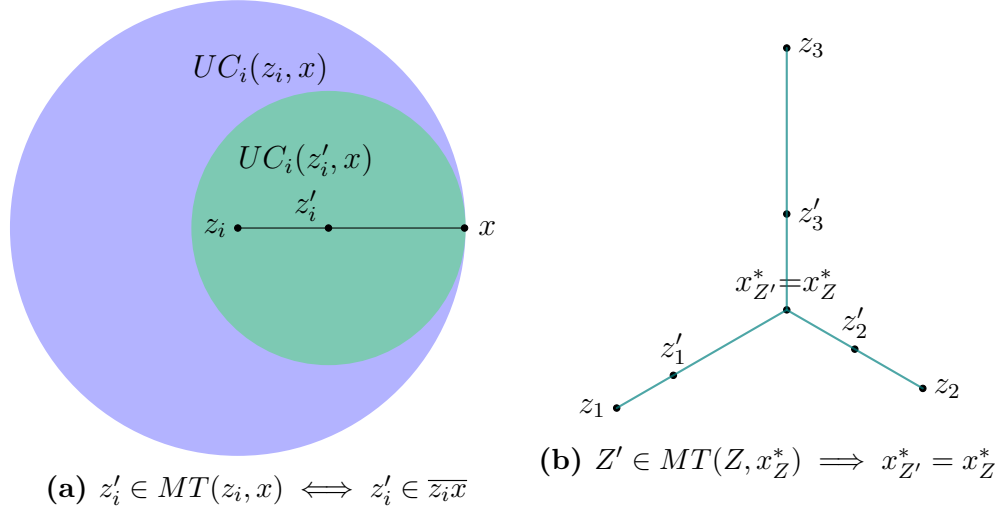
(b) $Z' \in MT(Z, x_Z^*) \implies x_{Z'}^* = x_Z^*$

Figure 2.2. Lemmas 2.3.1 and 2.3.2

The following result was originally established by Gini and Galvani (1929).

Lemma 2.3.2. *For $Z \in X^N$, if $Z' \in MT(Z, x_Z^*)$, then $x_{Z'}^* = x_Z^*$.*

We now prove Proposition 2.3.1.

Proof. It is trivial that φ satisfies anonymity.

Let f be any isometry. By uniqueness of the geometric median it follows that

$$\sum_{i=1}^n \|x_Z^* - z_i\| < \sum_{i=1}^n \|x - z_i\|$$

for any $x \in X \backslash \{x_Z^*\}$. Since f is an isometry we have

$$\sum_{i=1}^n \|x_Z^* - z_i\| = \sum_{i=1}^n \|f(x_Z^*) - f(z_i)\| < \sum_{i=1}^n \|f(x) - f(z_i)\|$$

for any $x \in X \backslash \{x_Z^*\}$. Since f is bijective it follows that $f(x_Z^*)$ is the geometric median for the set of points $(f(z_1), \ldots, f(z_n))$ and thus $f(\varphi(Z)) = f(x_Z^*) = \varphi(f(z_1), \ldots, f(z_n))$.

Finally, φ satisfying Maskin monotonicity immediately follows from Lemma

2.3.2. □

We did not impose any efficiency assumptions in our axiomatization. As we show next, PE of a social choice rule immediately follows from satisfying neutrality and Maskin monotonicity. Thus, the social choice rule such that $\varphi(Z) = x_Z^*$ satisfies PE by Proposition 2.3.1.

Proposition 2.3.2. *Let $Z \in X^N$ be a preference profile and $x \in X$ an outcome. Then, x is PE if and only if $x \in con(Z)$.*

Lemma 2.3.3. *For $Z \in X^N$, the set $comp(Z)$ is convex.*

Proof. Take any $a, b \in comp(Z)$ so that there exists $y_a, y_b \in X$ such that $u_i(y_a) \geq a_i$ and $u_i(y_b) \geq b_i$ for all $i \in N$. It follows by convexity of Euclidean distance that for any $t \in [0, 1]$ and for all $i \in N$ we have

$$u_i(ty_a + (1-t)y_b) \geq tu_i(y_a) + (1-t)u_i(y_b) \geq ta_i + (1-t)b_i.$$

Since X is convex and $y_a, y_b \in X$, it follows that $ty_a + (1-t)y_b \in X$ for any $t \in [0, 1]$. Thus, $ta + (1-t)b \in comp(Z)$ by definition, and so $comp(Z)$ is a convex set. □

Lemma 2.3.4. *For $Z \in X^N$ and an outcome $x \in X$, if x is PE then $u(x) = (u_1(x), \ldots, u_n(x)) \in bd(comp(Z))$.*

Proof. Fix $Z \in X^N$ and suppose $x \in X$ is PE but $u(x) \notin bd(comp(Z))$. Since $u(x)$ is not on the boundary, there exists $a = (a_1, \ldots, a_n) \in comp(Z)$ such that $a_i > u_i(x)$ for all i. By definition of $comp(Z)$ there exists $z \in X$ such that $u_i(z) \geq a_i > u_i(x)$ for all i, contradicting x is PE. □

We now prove Proposition 2.3.2.

Proof. First, note that $x = \sum_{i=1}^{n} \lambda_i x_i$ for $\lambda_i \geq 0$, $\sum_{i=1}^{n} \lambda_i = 1$, and $x_i \in X$ if and only if x solves

$$\max_{y \in X} - \sum_{i=1}^{n} \lambda_i (y - x_i) \cdot (y - x_i). \tag{2.2}$$

Suppose $x \in X$ is PE. By Lemma 2.3.4, $u(x) = (u_1(x), \ldots, u_n(x)) \in bd(comp(Z))$ and by Lemma 2.3.3 $comp(Z)$ is convex. Thus, it follows by the supporting hyperplane theorem (see Theorem 6.8 in Güler (2010)) that there exists $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ with $\alpha \neq 0$ such that $\alpha \cdot u(x) \geq \alpha \cdot a$ for all $a \in comp(Z)$. Further, $\alpha_i \geq 0$ for all i. If not, e.g. if $\alpha_i < 0$, by comprehensivity we could find $\tilde{a} \in comp(Z)$ with $\tilde{a}_i < 0$ small enough such that $\alpha \cdot u(x) < \alpha \cdot \tilde{a}$. Define $\lambda_i = \frac{\alpha_i}{\sum_{j=1}^{n} \alpha_j}$ so that $\lambda_i \geq 0$, $\sum_{i=1}^{n} \lambda_i = 1$, and $\lambda \cdot u(x) \geq \lambda \cdot a$ for all $a \in comp(Z)$. It follows that x solves

$$\max_{y \in X} \sum_{i=1}^{n} \lambda_i u_i(y), \tag{2.3}$$

which is of the same form as (2.2). Thus, $x = \sum_{i=1}^{n} \lambda_i z_i$ and we can conclude that $x \in con(Z)$.

Now suppose $x \in con(Z)$ and so it follows that x solves (2.3). However, suppose that x is not PE so that there exists $x' \in X$ such that $x' \succsim_i x$ for all i with $x' \succ_j x$ for at least one j. Since x solves (2.3) and x' Pareto dominates x, then it must be the case that $\lambda_j = 0$ for all j such that $x' \succ_j x$. Let $x'' = tx + (1-t)x'$ for some $t \in (0,1)$ so that $x'' \in X$ by convexity of the outcome space. By strict convexity of Euclidean preferences, $x'' \succ_i x' \succsim_i x$ for all i, contradicting x as a solution to (2.3). $\square$

Proposition 2.3.3. *Let φ be a social choice rule that satisfies Maskin monotonicity and neutrality. Then φ satisfies Pareto efficiency.*

Proof. Suppose φ is a social choice rule that satisfies Maskin monotonicity and

neutrality but not Pareto efficiency. Then, by Proposition 2.3.2, there exists $Z = (z_1, \ldots, z_n) \in X^N$ such that $\varphi(Z) \notin con(Z)$. By the separating hyperplane theorem (see Theorem 6.9 in Güler (2010)), there exists a hyperplane $H_{(\alpha,c)} = \{x \in X \mid \alpha \cdot x = c\}$ for some $0 \neq \alpha \in X$ and $c \in \mathbb{R}$ that separates $con(Z)$ and $\{\varphi(Z)\}$. For each $i \in N$ let $z_i' \in X$ satisfy $z_i' \in \overline{z_i \varphi(Z)}$ and $z_i' \in H_{(\alpha,c)}$, so z_i' is the intercept with the hyperplane of the line segment connecting agent i's ideal point to the social choice. It follows by Lemma 2.3.1 that $Z' = (z_1', \ldots, z_n') \in MT(Z, \varphi(Z))$ and thus $\varphi(Z') = \varphi(Z)$ by Maskin monotonicity. Let $f : X \to X$ be the reflection about the hyperplane $H_{(\alpha,c)}$ and define $z_i'' = f(z_i')$ for each $i \in N$. Since $z_i' \in H_{(\alpha,c)}$ it follows that $z_i'' = f(z_i') = z_i'$ and thus $\varphi(f(z_1'), \ldots, f(z_n')) = \varphi(Z')$. However, $\varphi(Z') \neq f(\varphi(Z'))$, contradicting neutrality $\qquad\square$

Maskin provides sufficient conditions under which a social choice function is Nash implementable by some game form (Maskin (1999)). The two conditions cited are monotonicity (which we have imposed as one of our axioms) and no veto power. In words, a social choice function satisfies no veto power if for any preference profile in which an outcome is most preferred by $n - 1$ of the agents, then this outcome is the social choice. To see that the function $\varphi(Z) = x_Z^*$ for $Z \in X^N$ satisfies no veto power note that the only instance in which the condition is not vacuously satisfied is when at least $n - 1$ of the agents have the same ideal point, in which case it is obvious that the geometric median also coincides with this point. Thus, the geometric median is Nash implementable as long as there are three agents.

2.3.1 The three agent case

We now show for $n = 3$ that the only social rule satisfying our axioms is the rule that selects the geometric median from the set of ideal points. In the conclusion we discuss the difficulties of extending the characterization to a fixed

$n > 3$.

We make note of a few preliminary results that will be used in the proof. Note that by Proposition 2.3.2 and Proposition 2.3.3 it must be the case that $\varphi(Z) \in con(Z)$. To this end, it is without loss to assume that $d = 2$, as the convex hull of any three points has at most two dimensions, and all monotonic transformations of those three points with respect to an element in the convex hull also lie in the convex hull. In what follows $B_{P,\varepsilon}$ will denote an ε-ball about a point $P \in X$ for some $\varepsilon > 0$.

Lemma 2.3.5. *Let $A' = (a'_1, a'_2, a'_3)$ be such that $\triangle_{a'_1 a'_2 a'_3}$ has all interior angles less than or equal to $120°$. Then $A' \in MT(A, x^*_A)$ for some A such that $\triangle_{a_1 a_2 a_3}$ is equilateral.*

Proof. Let $\triangle_{c_1 c_2 c_3}$ be some arbitrary equilateral triangle and x^*_C the associated geometric median. Since $x^*_C \in int(con(C))$, it follows that we can find an $\varepsilon > 0$ such that the associated $B_{x^*_C, \varepsilon} = \{c \in int(con(C)) \mid \|c - x^*_C\| < \varepsilon\} \subset int(con(C))$. Consider $\triangle_{c'_1 c'_2 c'_3} \cong \triangle_{a'_1 a'_2 a'_3}$ such that $con(C') \subset B_{x^*_C, \varepsilon}$, $c'_2 \in \overline{c_2 x^*_C}$, $c'_3 = x^*_C \in \overline{c_3 x^*_C}$, and $c'_1 \in con(\triangle_{c_1 x^*_C c_2})$.[7] Now consider a movement of $\triangle_{c'_1 c'_2 c'_3}$ such that $c'_2 = x^*_C \in \overline{c_2 x^*_C}$, $c'_3 \in \overline{c_3 x^*_C}$, and $c'_1 \in con(\triangle_{c_1 x^*_C c_3})$. Note that this movement is a continuous "shift" of c'_2 along $\overline{c_2 x^*_C}$ and c'_3 along $\overline{c_3 x^*_C}$ that sends c'_1 from $con(\triangle_{c_1 x^*_C c_2})$ to $con(\triangle_{c_1 x^*_C c_3})$. It thus follows that at some point along this movement we have $c'_i \in \overline{c_i x^*_C}$ for each i and thus $C' \in MT(C, x^*_C)$ by Lemma 2.3.1. See Figure 2.3 for an illustration of this procedure. Since $\triangle_{c'_1 c'_2 c'_3} \cong \triangle_{a'_1 a'_2 a'_3}$, the result follows by simply scaling both $\triangle_{c'_1 c'_2 c'_3}$ and $\triangle_{c_1 c_2 c_3}$ appropriately and then applying any needed isometries so that $\triangle_{c'_1 c'_2 c'_3} = \triangle_{a'_1 a'_2 a'_3}$ and then defining A to be the scaled and transformed C. $\square$

---

[7]Note that $c'_1 \in con(\triangle_{c_1 x^*_C c_2})$ will follow since it is assumed $\triangle_{c'_1 c'_2 c'_3}$ has all angles less than or equal to $120°$ while $\angle_{c_1 x^*_C c_2} = 120°$
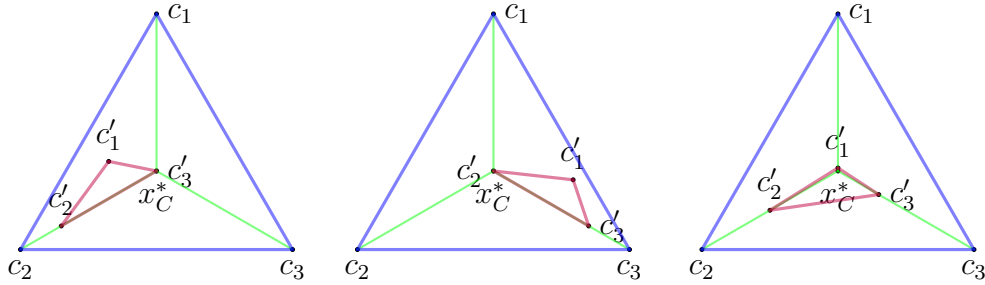
Figure 2.3. Finding $C' \in MT(C, x_C^*)$

Lemma 2.3.6. *Let $A' = (a_1', a_2', a_3')$ be such that $\triangle_{a_1' a_2' a_3'}$ is scalene[8] and has an angle greater than $120°$. Then $A' \in MT(A, x_A^*)$ for some A such that $\triangle_{a_1 a_2 a_3}$ is isosceles and the measure of the obtuse angles of the two triangles are the same.*

Proof. Without loss of generality let a_1' be the obtuse-angled vertex of $\triangle_{a_1' a_2' a_3'}$. Let $\triangle_{c_1 c_2 c_3}$ be an arbitrary isosceles triangle such that $\angle_{c_2 c_1 c_3} = \angle_{a_2' a_1' a_3'}$. Note that this implies $x_C^* = c_1$. Since $\triangle_{a_1' a_2' a_3'}$ is scalene, it follows that either $\angle_{a_1' a_2' a_3'} < \angle_{a_1' a_3' a_2'}$ or vice versa. Without loss of generality, assume the former holds. Since $\triangle_{c_1 c_2 c_3}$ is isosceles and $\angle_{c_2 c_1 c_3} = \angle_{a_2' a_1' a_3'}$ it then follows that $\angle_{a_1' a_2' a_3'} < \angle_{c_1 c_2 c_3} = \angle_{c_1 c_3 c_2} < \angle_{a_1' a_3' a_2'}$. Consider $c_3' \in \overline{c_3 x_C^*} = \overline{c_3 c_1}$ such that $\angle_{a_1' a_2' a_3'} = \angle_{c_1 c_2 c_3'}$ and $\angle_{c_1 c_3' c_2} = \angle_{a_1' a_3' a_2'}$.[9] Setting $c_2' = c_2$ and $c_1' = c_1$ gives us $C' \in MT(C, x_C^*)$ by Lemma 2.3.1 and $\triangle_{c_1' c_2' c_3'} \cong \triangle_{a_1' a_2' a_3'}$. The result follows by simply scaling both $\triangle_{c_1' c_2' c_3'}$ and $\triangle_{c_1 c_2 c_3}$ appropriately and then applying any needed isometries so that $\triangle_{c_1' c_2' c_3'} = \triangle_{a_1' a_2' a_3'}$ and then defining A to be the scaled and transformed C. $\qquad\square$

Now suppose A consists of distinct collinear points, and without loss of generality, assume $a_3 \in \overline{a_1 a_2}$. Further, assume $P \neq a_i$ for any i and that $P \in \overline{a_1 a_2}$. We now show how to construct $A' \in MT(A, P)$ such that a_3' is the midpoint of $\overline{a_1' a_2'}$, something we refer to as the collinear midpoint construction in the proof.

---

[8]A scalene triangle has all three interior angles of different measure.

[9]Note that as $c_3 \to c_1$ we have $\angle_{c_1 c_2 c_3} \to 0$ so finding such a c_3' is always possible by the Intermediate Value Theorem

First, assume $a_3 \in \overline{Pa_2}$. If $\|P - a_3\| < \frac{1}{2}\|P - a_2\|$ then choose $a_3' = a_3$ and $a_2' \in \overline{Pa_2}$ such that $\|P - a_3'\| = \frac{1}{2}\|P - a_2'\|$. If $\|P - a_3\| \geq \frac{1}{2}\|P - a_2\|$ then choose $a_2' = a_2$ and $a_3' \in \overline{Pa_3}$ such that $\|P - a_3'\| = \frac{1}{2}\|P - a_2'\|$. Choosing $a_1' = P$ gives us $A' \in MT(A, P)$ such that a_3' is the midpoint of $\overline{a_1'a_2'}$. Now, if $a_3 \in \overline{Pa_1}$, we can repeat the same procedure but reverse the roles of a_1 and a_2 so that a_1' and a_3' satisfy $\|P - a_3'\| = \frac{1}{2}\|P - a_1'\|$ and $a_2' = P$.

We now present our main result.

Theorem 2.3.1. *Suppose $n = 3$ and let φ be a social choice rule. Then φ satisfies anonymity, neutrality, and Maskin monotonicity if and only if for any $Z \in X^N$, $\varphi(Z) = x_Z^*$.*

Proof. Proposition 2.3.1 showed that the geometric median satisfies our three axioms. We prove the converse by cases.

Case 1 Let $Z \in X^N$ be a preference profile such that $\triangle_{z_1 z_2 z_3}$ is equilateral. Since the geometric median lies at the point of intersection of the three medians for the triangle, if we can show that $\varphi(Z)$ must lie on one of these medians chosen arbitrarily, then the claim will be true for Case 1. Let y be the midpoint between z_2 and z_3 and $\overline{z_1 y}$ the corresponding median. Suppose however that $\varphi(Z) \notin \overline{z_1 y}$. Let $f : X \to X$ be a reflection in $\overline{z_1 y}$ and define $z_i' = f(z_i)$ for each $i \in N$. It follows that f simply switches the vertices z_2 and z_3 in $\triangle_{z_1 z_2 z_3}$ so that $z_2' = z_3$ and $z_3' = z_2$. Note that f is equivalent to a bijection that switches agent 2's ideal point with agent 3's and, thus, by anonymity it must be the case that $\varphi(Z') = \varphi(Z)$. However, by neutrality $\varphi(f(z_1), f(z_2), f(z_3)) = \varphi(Z') = f(\varphi(Z))$, a contradiction. Thus, $\varphi(Z) \in \overline{z_1 y}$ and we must have $\varphi(Z) = x_Z^*$ when $\triangle_{z_1 z_2 z_3}$ is equilateral.

Case 2 Suppose now $\triangle_{z_1 z_2 z_3}$ is has all interior angles less than or equal to $120°$. By Lemma 2.3.5 there exists $\hat{Z}$ such that $Z \in MT(\hat{Z}, x_{\hat{Z}}^*)$ and $\triangle_{\hat{z}_1 \hat{z}_2 \hat{z}_3}$ is

equilateral. By Case 1, it follows that $\varphi(\hat{Z}) = x^*_{\hat{Z}}$. By Lemma 2.3.2 it follows that $x^*_Z = x^*_{\hat{Z}}$. The result then follows by Maskin montonicity.

Case 3 Consider now $\triangle_{z_1 z_2 z_3}$ that is isosceles with an interior angle greater than $120°$. Suppose, without loss of generality, that z_1 is the obtuse-angled vertex so that $x^*_Z = z_1$. By appealing to arguments similar to Case 1, it is easy to see that $\varphi(Z)$ must lie on the axis of symmetry. Suppose $\varphi(Z) \in int(con(Z))$ so it follows that $\angle_{z_2 \varphi(Z) z_3} > 120°$ and $\angle_{z_2 \varphi(Z) z_1} = \angle_{z_3 \varphi(Z) z_1}$. Note that we can find $z'_2 \in \overline{z_2 \varphi(Z)}$ and $z'_3 \in \overline{z_3 \varphi(Z)}$ such that $\angle_{z'_2 z_1 z'_3} \leq 120°$ and $\triangle_{z_1 z'_2 z'_3}$ is isosceles with an axis of symmetry through z_1. Choosing $z'_1 = z_1$ gives us $Z' \in MT(Z, \varphi(Z))$ and thus $\varphi(Z') = \varphi(Z)$ by Maskin monotonicity. By Case 2, it follows that $\varphi(Z') = x^*_{Z'}$ and thus $\angle_{z'_i \varphi(Z') z'_j} = 120°$ for all $i \neq j$. Thus, it follows by Lemma 2.3.1 that $\angle_{z_i \varphi(Z) z_j} = 120°$ for all $i \neq j$. But no such $\varphi(Z) \in int(con(Z))$ that lies on the axis of symmetry exists since $\angle_{z_2 \varphi(Z) z_3} > 120°$ by assumption. It then follows that the only choices for $\varphi(Z)$ are z_1 and the midpoint of $\overline{z_2 z_3}$. Suppose then that $\varphi(Z)$ is the midpoint of $\overline{z_2 z_3}$. Then, it is easy to see that we can find $Z' \in MT(Z, \varphi(Z))$ such that $\triangle_{z'_1 z'_2 z'_3}$ is equilateral[10] and $\varphi(Z')$ is the midpoint of $\overline{z'_2 z'_3}$, which contradicts Case 1. Thus we must have $\varphi(Z) = x^*_Z = z_1$.

Case 4 Suppose now $\triangle_{z_1 z_2 z_3}$ is scalene with an interior angle greater than $120°$. Without loss of generality, assume z_1 is the obtuse-angled vertex. By Lemma 2.3.6 there exists $\hat{Z}$ such that $Z \in MT(\hat{Z}, x^*_{\hat{Z}})$ and $\triangle_{\hat{z}_1 \hat{z}_2 \hat{z}_3}$ is isosceles with $\angle_{\hat{z}_2 \hat{z}_1 \hat{z}_3} = \angle_{z_2 z_1 z_3}$. By Case 3, it follows that $\varphi(\hat{Z}) = x^*_{\hat{Z}} = \hat{z}_1$. By Lemma 2.3.2 it follows that $x^*_Z = x^*_{\hat{Z}}$. The result then follows by Maskin montonicity.

Case 5 The last case to consider is when Z is a set of collinear points. If $z_1 = z_2 = z_3$ then trivially we must have $\varphi(Z) = z_1 = x^*_Z$ since that is the only choice

---

[10]This is achieved by moving z_2 and z_3 in tandem towards $\varphi(Z)$ until each side of the triangle is equal in length

in $con(Z)$. Suppose the points in Z are distinct and that $z_3 \in \overline{z_1 z_2}$ so that $x_Z^* = z_3$. If $\varphi(Z) \neq z_3$ then we can find $Z' \in MT(Z, \varphi(Z))$ such that z_3' is the midpoint of $\overline{z_1' z_2'}$ and, without loss of generality, $z_1' = \varphi(Z') = \varphi(Z)$ by Maskin monotonicity[11]. Let $f : X \to X$ be a reflection in the line perpendicular to $\overline{z_1' z_2'}$ running through z_3' and let $z_i'' = f(z_i')$. By neutrality we must have $\varphi(Z'') = f(\varphi(Z)) = z_1''$ but by anonymity, we must have $\varphi(Z'') = z_2''$ a contradiction. Thus, we must have $\varphi(Z) = x_Z^*$ when the points are collinear and distinct. Note that if the points were not distinct e.g. $z_1 \neq z_2 = z_3$ so that $x_Z^* = z_3$ still, but $\varphi(Z) \neq z_3$, then the previous argument still goes through by choosing $z_2' = z_2$, $z_3' \in \overline{\varphi(Z) z_3}$ such that $\|\varphi(Z) - z_3'\| = \frac{1}{2}\|\varphi(Z) - z_2'\|$ and $z_1' = \varphi(Z)$. Thus, in all collinear cases we have $\varphi(Z) = x_Z^*$, which completes the proof. $\qquad \square$

2.4 Conclusion

This paper has considered a natural generalization of the classical notions of May to spatial environments. The main contribution includes a characterization of the geometric median for the case of $n = 3$. A difficulty in extending the result is that there is no explicit solution to (2.1) for $n > 4$, although a characterization to the dual problem does exist.[12]

For the general case with an even number of agents, we would need to weaken the implicit hypothesis that our rule is single-valued. Importantly, the natural generalization of Proposition 2.3.3 fails for multi-valued rules. Further, even if we were to assume as a hypothesis that our rule is Pareto efficient, a characterization of the geometric median would fail to materialize; it is easy to see that the rule which selects the Pareto efficient correspondence satisfies natural counterparts of our

---

[11]This follows by the midpoint collinear construction outlined previously

[12]See Brady and Chambers (2015) for a discussion and result using the dual solution.

axioms. However, it is quite easy to see that the rule selecting the set of geometric medians satisfies the natural counterparts of our three axioms in the multi-valued case. Our conjecture in this general environment is that, for a sufficient number of dimensions, the rule selecting the geometric median is the *smallest* rule satisfying our three axioms, in the sense that any other nonempty-valued rule satisfying our three axioms must contain the set of geometric medians for any profile of ideal points.

Chapter 2 will appear in a forthcoming issue of Social Choice and Welfare and was coauthored with Christopher P. Chambers. The copyright of this article is held by Springer Publishing.

2.5   Appendix: Independence of the axioms

We now show that the axioms are independent. First, consider the dictatorship social choice rule φ_d such that $\varphi_d(Z) = z_1$ for all $Z \in X^N$. It is clear that this rule satisfies neutrality and Maskin monotonicity but violates anonymity.

Next, consider the social choice rule φ_c such that, for each Z in X^N, $\varphi_c(Z)$ is the unique solution to

$$\min_{x \in X} \sum_{i=1}^{n} (z_i - x)^2 . \tag{2.4}$$

The solution to (2.4) is the mean (or centroid) of the ideal points. Clearly φ_c satisfies anonymity. It is also easy to see that φ_c satisfies neutrality.

However, φ_c does not satisfy Maskin monotonicity. Consider the profile of n agents $Z = (e_1, -e_1, 0, \dots, 0)$ with e_i being the ith standard basis vector in $\mathbb{R}^d$. It follows that $\varphi_c(Z) = 0$. Now $Z' = (\frac{1}{2}e_1, -e_1, 0, \dots, 0) \in MT(Z, \varphi_c(Z))$ by Lemma 2.3.1 but $\varphi_c(Z') \neq 0 = \varphi_c(Z)$.

Finally consider the social choice rule φ_m such that $\varphi_m(Z)$ solves

$$\min_{x \in X} \sum_{i=1}^{n} |z_i \cdot e_j - x \cdot e_j|$$

for each $j \in \{1, \dots, d\}$. That is, φ_m selects the coordinate-wise median of the ideal points. It is obvious φ_m satisfies anonymity. It is also fairly easy to see that φ_m satisfies Maskin monotonicity. To see this, simply apply the proof technique used in Lemma 2.3.2 (see Brady and Chambers (2015)) for each coordinate. However, it is a well known fact that the coordinate-wise median does not satisfy neutrality as it is not equivariant with isometries.

Chapter 3

Testability, Identification, and Estimation in Distance-Monotonic Noise Models with Application to Online Voting Platforms

Abstract: I work in an environment in which agents vote for the best alternative from a set of alternatives. The set of alternatives has a correct objective ranking as in the classic work of Condorcet (1785). A researcher samples votes from a large population of potential voters such as an online voting platform. I model voting behavior using the class of distance-monotonic noise models, a large class of voting models that contains the popular model of Mallows (1957). I present a number of theoretical results concerning testable implications, identification, and estimation based on observation of voting probabilities generated by these models. A large emphasis is placed on operationalizing the theoretical results by providing statistical methods based on sample data.

3.1 Introduction

Online crowdsourcing systems have become a common approach to obtaining contributions from a large group of people. For example, large-scale online voting platforms provide an efficient way to develop a ranking of a set of alternatives. Motivated by the prevalence of such platforms, I investigate the testability, identification, and estimation of a large class of voting models relevant for the crowdsourcing domain. My starting point is a finite set of alternatives X that has a true, objective ranking according to some criteria. A researcher samples from a large population of voters who are ignorant of the correct ranking. Each respondent is presented with a choice set $A \subseteq X$ and is asked to vote for the highest ranked (i.e. "best") alternative according to the specified criteria. Absent any strategic incentives, each respondent's vote can be viewed as a noisy estimate of which alternative is truly best. As a working example, consider a set of reading passages that have been assigned a difficulty level according to some accepted standards, such as Flesch-Kincaid grade level score (Kincaid, Fishburne Jr, Rogers, and Chissom, 1975). Each respondent is shown a subset of the passages and asked to vote for the one with the highest reading difficulty as in Collins-Thompson and Callan (2004) and Chen, Bennett, Collins-Thompson, and Horvitz (2013).

In some cases the researcher's goal is to approximate the correct ranking or the best alternative based on the sampled votes. For example, the EteRNA project (http://www.eternagame.org/web/) solicits votes on RNA folding designs to identify which designs are most stable. The designs receiving the most votes are then synthesized in a lab to verify the stability of each. Ideally, the researcher can approximate the correct ranking with a high degree of probabilistic certainty using as few samples as possible. Thus, knowing how many votes to sample and

how they should be used to approximate the correct ranking or best alternative are important considerations. When the correct ranking is known to the researcher, such as in the reading difficulty example mentioned previously, the goal usually involves developing/testing an algorithm for machine learning purposes. Developing and testing an algorithm requires modeling how votes are generated, in which case model refutability and identification are of paramount importance.

3.1.1 Contribution

I model voting behavior as if votes are generated by a *distance-monotonic noise model*. A noise model specifies a probability distribution over rankings of the alternatives. These models generate a vote as the top ranked alternative in the choice set according to a random draw from the distribution over rankings. A distance-monotonic noise model requires that the likelihood of a ranking is increasing in its "similarity" (as measured by a distance function) to the true ranking. The distance functions I employ satisfy two assumptions. First, they are *right-invariant* - the distance between two rankings is preserved by a relabeling of the alternatives. Secondly, they are *swap-increasing* - the distance between any two rankings that agree on the relative ranking of a pair of alternatives cannot decrease by swapping the ranks of those alternatives in one of the rankings.

In the first half of Section 3.3, I investigate voting behavior in terms of the population vote probabilities generated by a distance-monotonic noise model. The purpose of this section is to provide simple theoretical conditions by which the class of models can be refuted. The results can be interpreted as how difficult it would be for an individual drawn at random to identify the top ranked alternative in and across different choice sets. I show that behavior is quite plausible in this setting: higher ranked alternatives (according to the true ranking) are more likely than

lower ranked alternatives to be identified as the best (Proposition 3.3.2), increasing the size of the choice set makes the task more difficult (Proposition 3.3.1), and replacing an alternative that is not ranked best in a choice set with a lower ranked one makes it more likely the top ranked alternative is identified (Proposition 3.3.3). I also prove a strong invariance property for the class of models that are monotonic with respect to the Kendall tau distance, arguably the most ubiquitous distance function on rankings. The result states that the population vote probabilities are invariant across choice sets which have the same cardinality and are composed of alternatives that have the same relative spacing in the true ranking (Proposition 3.3.4). For example, identifying the best alternative from those ranked third, fourth, and eighth is identical in terms of difficulty as identifying the best from those ranked fifth, sixth, and tenth.

In the second half of Section 3.3, I introduce two statistical tests that can be used to refute the class of models when the true ranking is known. Both tests are shown to show control size (Type I error probability) even in finite samples (Theorem 3.3.1). This section is part of a recurring theme of the paper in which I bridge the gap between purely theoretical results, which involve empirically unobservable population vote probabilities, and data a researcher may actually be able to elicit, which are inherently finite.

Section 3.4 discusses identification of the true ranking and distribution over rankings. I show that observation of the population vote probabilities identifies the true ranking (Theorem 3.4.1). However, in the class of Kendall tau monotonic noise models, only set identification of the distribution over rankings is possible. That is, for every noise model in this class, there is a set of observationally equivalent models also in this class. I show how to construct such a model given observation of the population vote probabilities (Theorem 3.4.2). A linear program is presented

whose solutions constitute the identified set of noise models compatible with the population vote probabilities. In a similar fashion to the latter half of Section 3.3, I propose finite sample confidence sets for the true ranking and the distribution over rankings as a means of operationalizing the theoretical results.

Estimation and choice experiment design are discussed in Section 3.5. I introduce the class of *pairwise consistent* estimators for the true ranking. If the data can form a valid ranking based on pairwise comparisons of vote counts, then any estimator in this class must return this ranking.[1] In a sense, pairwise consistent estimators can be thought of as an extension to Condorcet consistent estimators for the best alternative. I show that with adequate data any Condorcet consistent estimator and any pairwise consistent estimator converge almost surely to the best alternative and the true ranking respectively (Theorems 3.5.1 and 3.5.2). Working with a particularly tractable class of models introduced by Mallows (1957), I investigate choice experiment design in the latter half of the section. Specifically, I find upper bounds on the number of samples needed to return the best alternative or the true ranking with high probability under a commonly used vote sampling scheme (Theorem 3.5.3).

3.1.2   Related work

The notion of votes being viewed as noisy estimates of an underlying ground truth is due to Condorcet (1785). In Condorcet's approach, each voter ranks every pair of alternatives. The voter correctly ranks any given pair with probability $q > \frac{1}{2}$. When the output of each voter's pairwise comparisons is required to be a consistent ranking, Condorcet's model is equivalent to one proposed by Mallows

---

[1] A similar class of estimators are discussed in Caragiannis, Procaccia, and Shah (2013) called *pairwise-majority consistent* voting rules. In their setting, votes are submitted as entire rankings of the entire set X rather than a single alternative from a presented choice set.

(1957). Mallows' model specifies that the likelihood of any ranking is exponentially decreasing in its distance from the true ranking. It is one of the first and arguably the most ubiquitous of all distance-monotonic noise models, and has inspired many generalizations (see e.g. Fligner and Verducci, 1986; Critchlow, Fligner, and Verducci, 1991).

These types of models have received considerable recent attention in the computer science and machine learning literatures since the assumption of a true ranking of alternatives is satisfied in many crowdsourcing and human computation platforms in which voting plays a crucial role.[2] Contrary to the current context, this literature usually assumes that voters submit an entire ranking of the set X. I choose to focus on voters submitting a single vote from a choice set for several reasons. First, it can be difficult to submit an entire ranking, particularly if the cardinality of the set X is large or if discerning differences among the alternatives is demanding. Thus, asking for a single vote for the best alternative shown greatly reduces the burden on voters. Secondly, by working with voting probabilities from different choice sets, the analysis is directly comparable to the stochastic choice literature as pioneered by the work of Luce (1959b) and Block and Marschak (1960). Finally, soliciting a vote for a single alternative is more common in practice than requiring the submission of an entire ranking.

The analysis in Section 3.3 is most closely related to the aforementioned stochastic choice literature, which seeks to explain the randomness observed in individual choice. Typically, a behavioral model is proposed and then characterized

---

[2]See e.g. Conitzer, Rognlie, and Xia (2009); Elkind, Faliszewski, and Slinko (2010); Lu and Boutilier (2011); Conitzer and Sandholm (2012); Procaccia, Reddi, and Shah (2012); Caragiannis, Procaccia, and Shah (2013); Mao, Procaccia, and Chen (2013); Soufiani, Chen, Parkes, and Xia (2013); Busa-Fekete, Hüllermeier, and Szörényi (2014); Caragiannis, Procaccia, and Shah (2014); Chierichetti and Kleinberg (2014); Jiang, Marcolino, Procaccia, Sandholm, Shah, and Tambe (2014); Soufiani, Parkes, and Xia (2014).

in terms of axioms which describe how choice probabilities differ across menus (see e.g. Echenique, Saito, and Tserenjigmid, 2013; Fudenberg, Iijima, and Strzalecki, 2013; Manzini and Mariotti, 2014; Aguiar, 2015a). These characterizations provide the testable content of the model. However, the axioms are usually complicated and no guidance is provided on how one might statistically test them in practice given sample choice data.

The class of random utility models (RUMs) proposed by Block and Marschak (1960) are particularly relevant. A variant of RUMs postulates choice as being generated from a probability distribution over preference orderings. Thus, the class of distance-monotonic noise models is a subset of this variant of RUMs. Section 3.4 is related to the small literature on identification of the distribution over preference orderings given observation of the RUM generated choice probabilities (Falmagne, 1978; Barbera and Pattanaik, 1986; Manski, 2007; Sher, Fox, Kim, and Bajari, 2011). This literature points to the general impossibility of unique identification, similar to Theorem 3.4.2. Sher, Fox, Kim, and Bajari (2011) give the most thorough treatment of the topic and investigate several variants of datasets which might reduce the degree of underidentification.

Finally, the work of Caragiannis, Procaccia, and Shah (2013) is of particular importance as some of the terminology used here was introduced in their paper. The authors ask which voting rules will return the true ranking with probability 1 given infinite samples of rankings drawn from distance-monotonic noise models, a concept they define as accuracy in the limit. They define two classes of voting rules and characterize the distance functions for which these rules are accurate in the limit. They also investigate the performance of common voting rules under sampling from Mallows' model in terms of the number of samples needed to return the true ranking with high probability. Thus, their analysis is most similar to that

presented in Section 3.5. However, in their setting votes consist of an entire ranking of the set of alternatives.

3.2 Preliminaries

Let X be a finite set of $m > 2$ alternatives. Denote the set of (strict) rankings over X as $\mathcal{L}(X)$. I represent a ranking by a bijection $\sigma : X \to \{1, \ldots, m\}$ where $\sigma(a)$ is the rank of alternative $a \in X$ according to σ. Thus, $\sigma(a) < \sigma(b)$ indicates a is ranked better (colloquially "higher") than b under σ, also denoted as $a \succ_\sigma b$. I will occasionally write a ranking as the ordering it induces with the alternatives ordered from best to worst such as $\sigma = abc \ldots$.

Any $A \subseteq X$ with $|A| > 1$ is called a *choice set* or *menu*. Let $\mathcal{O}$ be the set of *observations*, which are pairs of the form (A, a) where A is a choice set and $a \in A$. The interpretation of an observation is that a respondent voted a as the best alternative from A. A *dataset* is any collection of observations $\omega \in \mathcal{O}^n$ with $n \in \mathbb{N}$.

3.2.1 Noise models and distances

I assume there is a true ranking of the alternatives $\sigma^* \in \mathcal{L}(X)$. I will often label the alternatives in X according to their ranking in σ^*, so that the ith ranked alternative is a_i i.e. $\sigma^*(a_i) = i$.

A function $d : \mathcal{L}(X) \times \mathcal{L}(X) \to \mathbb{R}_+$ is called a *distance function* or *metric* if for all $\sigma, \sigma', \sigma'' \in \mathcal{L}(X)$ it satisfies: i) $d(\sigma, \sigma') \geq 0$ with equality if and only if $\sigma = \sigma'$, ii) $d(\sigma, \sigma') = d(\sigma', \sigma)$, and iii) $d(\sigma, \sigma') \leq d(\sigma, \sigma'') + d(\sigma'', \sigma')$. The following are considered some of the most widely used distance functions on rankings (Diaconis,

1988):

$$d_K(\sigma, \sigma') = \sum_{i<j} \mathbb{1}\left\{[\sigma(a_i) - \sigma(a_j)] \times [\sigma'(a_i) - \sigma'(a_j)] < 0\right\}$$

is the *Kendall tau* distance,

$$d_F(\sigma, \sigma') = \sum_i |\sigma(a_i) - \sigma'(a_i)|$$

is the *footrule* distance, and

$$d_H(\sigma, \sigma') = \sum_i \mathbb{1}\left\{\sigma(a_i) \neq \sigma(a_i')\right\}$$

is the *Hamming* distance.

Each distance function above provides a degree of dissimilarity between two rankings: d_K counts the number of discordant pairs, d_F is the total difference in the alternatives' rankings, and d_H is the number of alternatives ranked differently in the two rankings. The Kendall tau distance d_K is especially prominent: it has a natural axiomatic foundation (Kemeny, 1959) and is often used in the social choice, statistics, and computer science literatures.[3]

A distance function d is said to be *right-invariant* if the distance between any two rankings does not change if alternatives are relabeled. Note that a distance function that is right-invariant is fully specified by the distances of all rankings from a single base ranking. For a ranking σ, let $\sigma_{a \leftrightarrow b}$ denote the ranking in which the ranks of a and b are swapped in σ i.e. $\sigma(c) = \sigma_{a \leftrightarrow b}(c)$ for all $c \neq a, b$ and $\sigma(a) = \sigma_{a \leftrightarrow b}(b)$ and $\sigma(b) = \sigma_{a \leftrightarrow b}(a)$. A distance function d is said to be *swap-*

---

[3]See e.g. Young and Levenglick (1978); Fligner and Verducci (1986); Young (1988); Critchlow, Fligner, and Verducci (1991); Bossert and Storcken (1992); Klamler (2008); Lu and Boutilier (2011); Baldiga and Green (2013); Caragiannis, Procaccia, and Shah (2013); Mao, Procaccia, and Chen (2013); Busa-Fekete, Hüllermeier, and Szörényi (2014); Coffman (2015); Procaccia and Shah (2015).

increasing if for any σ, σ' and any alternatives a, b such that $a \succ_\sigma b$ and $a \succ_{\sigma'} b$ it follows that $d(\sigma, \sigma') \leq d(\sigma_{a \leftrightarrow b}, \sigma')$ and is said to be *strictly swap-increasing* if the inequality is strict. Most, if not all, commonly used distance functions are right-invariant, including the three discussed above (Diaconis, 1988). Further, it has been shown that d_F and d_H are swap-increasing (Critchlow, Fligner, and Verducci, 1991) while d_K is strictly swap-increasing (Caragiannis, Procaccia, and Shah, 2013). I assume all distance functions discussed in this paper are right-invariant and swap-increasing.

Given a true ranking $\sigma^* \in \mathcal{L}(X)$, a *noise model* defines the probability of observing any ranking σ, denoted as $\Pr(\sigma|\sigma^*)$ for all $\sigma \in \mathcal{L}(X)$. Let d be a distance function on rankings. A noise model is said to be *monotonic with respect to d* if for any rankings σ, σ' and true ranking σ^* it follows that $\Pr(\sigma|\sigma^*) \geq \Pr(\sigma'|\sigma^*) \iff d(\sigma, \sigma^*) \leq d(\sigma', \sigma^*).$[4] Thus, the further a ranking is from the true ranking according to d, the less likely it is realized. Let $\mathcal{G}$ be the set of all noise models that are monotonic with respect to some distance function d. When referring to a specific distance function d I will denote this subset of $\mathcal{G}$ as $\mathcal{G}_d$, e.g. $\mathcal{G}_{d_K}$.

One of the most common distance-monotonic noise models was proposed by Mallows (1957) in which the probability of observing any particular ranking is exponentially decreasing in its Kendall tau distance from the true ranking. For Mallows' model, the probability of observing a ranking σ given true ranking σ^* is given by

$$\Pr(\sigma|\sigma^*) = \frac{\phi^{d_K(\sigma, \sigma^*)}}{Z_\phi^m} \tag{3.1}$$

where $\phi \in (0, 1)$ and Z_ϕ^m is a normalization factor independent of the true ranking.

---

[4]Models in this class have been used as early as Mallows (1957). This particular naming of the class is due to Caragiannis, Procaccia, and Shah (2013).

Mallows' model is a simple model and provides a basis for much of the recent computational voting literature.[5] The scale parameter ϕ has the feature that as $\phi \to 0$ the distribution over rankings approaches degeneracy at the true ranking and as $\phi \to 1$ the distribution approaches uniformity.

Due to the prominence of the Kendall tau distance, I pay particular attention to the general class $\mathcal{G}_{d_K}$, of which Mallows' model is a special case. For any noise model $G \in \mathcal{G}$, I maintain the assumption of a full support distribution on $\mathcal{L}(X)$.

3.2.2 Data collection

I assume a data collection process of a researcher soliciting the votes of a large number of people such as in a crowdsourcing platform. Each respondent is presented with a choice set from which they are asked to submit a vote for the top ranked alternative. The result is an observation e.g. (A, a). I assume that the correct ranking of alternatives is not known to each respondent. The researcher controls the experiment design by choosing which choice sets to use and how many observations to collect on each choice set. The researcher may collect as many observations as desired to achieve her goal. Given a choice set A and alternative $a \in A$, let $\mathcal{L}(a|A) = \{\sigma \in \mathcal{L}(X) : a \succ_\sigma b \text{ for all } b \in A \setminus \{a\}\}$ be the set of *rationalizing rankings* and let $p(a|A) = \sum_{\sigma \in \mathcal{L}(a|A)} \Pr(\sigma|\sigma^*)$ be the *population probability* of a vote for a from A. When confusion may arise with the alternative numbering scheme according to σ^*, I will refer to observation i in a dataset $\omega \in \mathcal{O}^n$ as (B_i, b_i). Votes are sampled independently, so the probability of observing the

---

[5]See e.g. Lu and Boutilier (2011); Meila, Phadnis, Patterson, and Bilmes (2012); Procaccia, Reddi, and Shah (2012); Aledo, Gámez, and Molina (2013); Caragiannis, Procaccia, and Shah (2013); Busa-Fekete, Hüllermeier, and Szörényi (2014); Aledo, Gámez, and Molina (2016).

dataset $\omega = ((B_1, b_1) \dots, (B_n, b_n))$ is given by

$$\Pr(\omega | \sigma^*) = \prod_{i=1}^{n} p(b_i | B_i).$$

3.3 Voting behavior and testability

In this section, I examine some testable implications of the population vote probabilities generated by a d-monotonic noise model. Rather than provide a complete characterization, I investigate simple conditions which have a clear interpretation. As a means to operationalize some of the conditions, I propose some statistical tests by which the model can be refuted with adequate data. I end the section with a brief discussion on how more complicated testing procedures might be employed.

3.3.1 Testable implications

I first look at probabilistic voting behavior generated by the general class of models $\mathcal{G}$. A reasonable interpretation for the results is in terms of the relative difficulty an individual would have in correctly identifying the best alternative within and across choice sets. The results suggest behavior in line with intuition. For example, correctly identifying the passage with highest reading difficulty out of two should be more likely than identifying that same passage as most difficult when more options have been added to the choice set, which is the content of Proposition 3.3.1.

Proposition 3.3.1. *Let $G \in \mathcal{G}$ be a d-monotonic noise model. For any choice sets A, C such that $A \subset C$ and alternative $a \in A$ it follows that $p(a|A) > p(a|C)$.*

Proof. Trivial by the full support assumption on $\mathcal{L}(X)$ and noting that $\mathcal{L}(a|A) \subset \mathcal{L}(a|C)$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

The result in Proposition 3.3.1 is not new and is referred to as *regularity* in the choice-theoretic literature, although it is usually presented with a weak inequality. It is known to hold for a large class of models known as random utility models (Block and Marschak, 1960) which subsumes the class $\mathcal{G}$.

Proposition 3.3.2 and its corollary compare the relative likelihood of alternatives being voted for across different choice sets. Intuitively, higher ranked alternatives would be more likely to be identified as the best from a choice set than lower ranked alternatives. For instance, it is reasonable to expect a reading passage of grade level 10 to receive at least as many votes for being the most difficult in a choice set as one of grade level 3 regardless of the choice set's composition. The proof relies on creating a bijection between rationalizing sets of rankings using the mapping $\sigma_{a \leftrightarrow b}$ for any two alternatives a, b in a choice set. If a is truly better than b, then such a bijection always transforms a ranking yielding a vote for a into a ranking yielding a vote for b which is at least as far away from the true ranking. A result using Mallows specification is also presented, which will be helpful in the analysis on choice experiment designs in Section 3.5. Recall that the numbering of alternatives is given by their ranking according to σ^*.

Proposition 3.3.2. *Let $G \in \mathcal{G}$ be a d-monotonic noise model. Then for any choice set A and $a_i, a_j \in A$ such that $i < j$ it follows that $p(a_i|A) \geq p(a_j|A)$ where the inequality is strict if d is strictly swap-increasing. In Mallows' model, if $j = i + 1$ then $p(a_j|A) = \phi\, p(a_i|A)$.*

Proof of Proposition 3.3.2. Take any $\sigma \in \mathcal{L}(a_i|A)$. Note that $\sigma_{a_i \leftrightarrow a_j}$ is a bijection from $\mathcal{L}(a_i|A)$ to $\mathcal{L}(a_j|A)$. Further, $d(\sigma, \sigma^*) \leq d(\sigma_{a_i \leftrightarrow a_j}, \sigma^*)$ if $i < j$ by definition of

swap-increasing. Using this, the result then follows by definition of d-monotonic noise model. Using the definition for strictly swap-increasing d proves the strict inequality claim. To prove the case of Mallows' model where $d = d_K$, I use the following lemma.

Lemma 3.3.1. *If $j = i + 1$ then*

$$d_K(\sigma, \sigma^*) + 1 = d_K(\sigma_{a_i \leftrightarrow a_j}, \sigma^*).$$

Proof of Lemma 3.3.1. Since $j = i + 1$ the following equalities hold:

$$|\{k : a_k \succ_{\sigma^*} a_i \text{ and } a_i \succ_\sigma a_k\}| + 1 = |\{k : a_k \succ_{\sigma^*} a_j \text{ and } a_j \succ_{\sigma_{a_i \leftrightarrow a_j}} a_k\}|,$$

$$|\{k : a_i \succ_{\sigma^*} a_k \text{ and } a_k \succ_\sigma a_i\}| = |\{k : a_j \succ_{\sigma^*} a_k \text{ and } a_k \succ_{\sigma_{a_i \leftrightarrow a_j}} a_j\}|,$$

$$|\{k : a_k \succ_{\sigma^*} a_j \text{ and } a_j \succ_\sigma a_k\}| = |\{k : a_k \succ_{\sigma^*} a_i \text{ and } a_i \succ_{\sigma_{a_i \leftrightarrow a_j}} a_k\}|, \text{ and}$$

$$|\{k : a_j \succ_{\sigma^*} a_k \text{ and } a_k \succ_\sigma a_j\}| = |\{k : a_i \succ_{\sigma^*} a_k \text{ and } a_k \succ_{\sigma_{a_i \leftrightarrow a_j}} a_i\}|$$

from which the result follows. □

To prove the case of Mallows' model, note that

$$\begin{aligned}
p(a_j|A) &= \sum_{\sigma' \in \mathcal{L}(a_j|A)} \frac{\phi^{d_K(\sigma', \sigma^*)}}{Z_\phi^m} \\
&= \sum_{\sigma \in \mathcal{L}(a_i|A)} \frac{\phi^{d_K(\sigma_{a_i \leftrightarrow a_j}, \sigma^*)}}{Z_\phi^m} \\
&= \sum_{\sigma \in \mathcal{L}(a_i|A)} \frac{\phi^{d_K(\sigma, \sigma^*)+1}}{Z_\phi^m} \\
&= \phi \left(\sum_{\sigma \in \mathcal{L}(a_i|A)} \frac{\phi^{d_K(\sigma, \sigma^*)}}{Z_\phi^m} \right) = \phi p(a_i|A)
\end{aligned}$$

where the second transition follows since $\sigma_{a_i \leftrightarrow a_j}$ is a bijection from $\mathcal{L}(a_i|A)$ to $\mathcal{L}(a_j|A)$ and the third transition follows by Lemma 3.3.1. □

The corollary to Proposition 3.3.2 states that the best alternative in a choice set has the highest likelihood of receiving a vote.

Corollary 3.3.1. *If $a \in A$ satisfies $\sigma^*(a) < \sigma^*(b)$ for all $b \in A \setminus \{a\}$ then*

$$p(a|A) > p(b|A)$$

for all $b \in A \setminus \{a\}$.

Proof. Since $a \succ_{\sigma^*} b$ for all $b \neq a$ it follows $\sigma^* \in \mathcal{L}(a|A)$. Thus, $d(\sigma^*, \sigma^*) < d(\sigma^*_{a \leftrightarrow b}, \sigma^*)$ and the result follows by Proposition 3.3.2 and definition of d-monotonic noise model. □

Suppose a respondent was shown passages of grade difficulty levels 10 and 9 and asked to vote for the one of higher difficulty. This should be a more challenging task than if the respondent were comparing the grade level 10 document to one of a grade level lower than 9 e.g. grade level 5. Proposition 3.3.3 formalizes this intuition. The result follows by similar arguments as those used to prove Proposition 3.3.2.

Proposition 3.3.3. *Let $G \in \mathcal{G}$ be a d-monotonic noise model. Then for any two choice sets C, D such that $C = D \setminus \{a_j\} \cup \{a_i\}$ for some $a_j \in D$, $a_i \notin D$ and $i < j$ it follows that $p(a|D) \geq p(a|C)$ for all $a \in C \cap D$ where the inequality is strict if d is strictly swap-increasing.*

Proof of Proposition 3.3.3. Take any $\sigma \in \mathcal{L}(a|D)$. If $\sigma(a) < \sigma(a_i)$ then $\sigma \in \mathcal{L}(a|C)$. Suppose $\sigma(a) > \sigma(a_i)$ so that $a_i \succ_\sigma a \succ_\sigma a_j$. Note that such a ranking σ always exists since $a_i \notin D$. Since $a_j \notin C$ it follows that $\sigma_{a_i \leftrightarrow a_j} \in \mathcal{L}(a|C)$. Further,
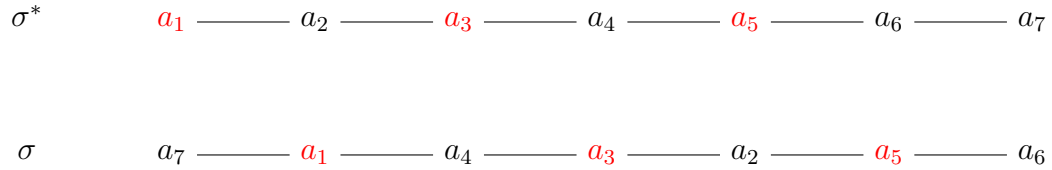
$$\sigma^* \qquad a_1 \text{------} a_2 \text{------} a_3 \text{------} a_4 \text{------} a_5 \text{------} a_6 \text{------} a_7$$

$$\sigma \qquad a_7 \text{------} a_1 \text{------} a_4 \text{------} a_3 \text{------} a_2 \text{------} a_5 \text{------} a_6$$

Figure 3.1. Orderings induced by σ^* and σ

$d(\sigma^*, \sigma) \leq d(\sigma^*, \sigma_{a_i \leftrightarrow a_j})$ by definition of swap-increasing. Thus, for every $\sigma \in \mathcal{L}(a|D)$, either $\sigma \in \mathcal{L}(a|C)$ or there is a corresponding $\sigma_{a_i \leftrightarrow a_j} \in \mathcal{L}(a|C)$ that is at least as far away from σ^* as σ is. The result then follows by definition of d-monotonic noise model. Proving the strict inequality case follows by using the definition of strictly swap-increasing. $\qquad\square$

The final result in this section establishes a strong invariance property for the class $\mathcal{G}_{d_K}$. It states that choice sets of the same cardinality that are composed of alternatives with the same relative rank differences in the true ranking can be viewed as equivalent according to the vote probabilities they induce. For example, choosing the most difficult among reading passages of grade levels 1, 3, and 6 would be considered equally as challenging as choosing the most difficult among passages of grade levels 4, 6, and 9. The result is a significant generalization of one first shown by Mallows (1957). The proof is rather technical. Fix any choice set and fix the relative rank differences of the alternatives in that choice set according to σ^*. Then, take any ranking σ in which the alternatives in the choice set have the same relative rank differences according to σ as they do according to σ^* (see Figure 3.1). It turns out that the vote probabilities induced on the choice set are equal to the vote probabilities that would be induced if σ were the true ranking rather than σ^*. This can then be used to establish the equivalence *across* choice sets satisfying the conditions of the Proposition.

Proposition 3.3.4. *Let $G \in \mathcal{G}_{d_K}$ be a d_K-monotonic noise model and let C and D be two choice sets with $|C| = |D| = s$. Label the alternatives in C according to their ranking in σ^* i.e. $\{c_1, \ldots, c_s\}$ with $i < j \iff c_i \succ_{\sigma^*} c_j$ and label the alternatives in D accordingly. If $\sigma^*(c_{i+1}) - \sigma^*(c_i) = \sigma^*(d_{i+1}) - \sigma^*(d_i)$ for each $i \in \{1, \ldots, s-1\}$ then $p(c_i|C) = p(d_i|D)$ for each i.*

The proof relies on the following lemma. The result says that any ranking in which the relative spacing of the alternatives in C is the same as that in σ^* induces the same distribution of distances on rankings that solicit a vote for c_i for every i.

Lemma 3.3.2. *Take any $\sigma' \in \mathcal{L}(X)$ such that $\sigma'(c_{i+1}) - \sigma'(c_i) = \sigma^*(c_{i+1}) - \sigma^*(c_i)$ for each $i \in \{1, \ldots, s-1\}$. Then*

$$|\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma^*, \sigma) = \delta\}| = |\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma', \sigma) = \delta\}|$$

for each possible distance δ and for every i.

Proof of Lemma 3.3.2. If $\sigma' = \sigma^*$ then the result is trivially true, so assume $\sigma' \neq \sigma^*$. The proof proceeds in cases. For any i, pick a δ such that $\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma^*, \sigma) = \delta\}$ is nonempty and let σ be an arbitrary ranking in this set.

Case 1: $\sigma'(c_i) = \sigma^*(c_i)$ for every i.

Let $\tau : X \to X$ be the unique bijection acting on the *ordering* induced by σ^* that yields the ordering induced by σ' written as $\tau(\sigma^*) = \sigma'$. Note that this operation constitutes a simple relabeling of the alternatives in $X \setminus C$. Further $\tau(\sigma) \in \mathcal{L}(c_i|C)$ by the assumption in place for Case 1. By right-invariance of d_K, it follows that $d_K(\sigma^*, \sigma) = d_K(\tau(\sigma^*), \tau(\sigma)) = d_K(\sigma', \tau(\sigma)) = \delta$. Thus, for every ranking in $\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma^*, \sigma) = \delta\}$ there is a unique ranking in $\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma', \sigma) = \delta\}$, proving the result for Case 1.

Case 2: $a_m \notin C$ and $\sigma' = a_m a_1 \ldots a_{m-1}$.

Suppose $\sigma(a_m) = r$. Construct the ranking σ'' such that the ordering of the alternatives in $X \setminus \{a_m\}$ agrees with the ordering according to σ but $\sigma''(a_m) = m - (r-1)$. It is clear based on the assumptions of Case 2 that $\sigma'' \in \mathcal{L}(c_i|C)$. For any pair of alternatives in $X \setminus \{a_m\}$ it follows that σ^* and σ disagree on their ordering if and only if σ' and σ'' do also by construction. Further, the ranking of a_m in σ is responsible for an additional $m - r$ disagreements with σ^* and the ranking of a_m in σ'' is responsible for an additional $m - r$ disagreements with σ'. Thus, for every ranking in $\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma^*, \sigma) = \delta\}$ there is a unique ranking in $\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma', \sigma) = \delta\}$, proving the result for Case 2.

Case 3: $a_m \notin C$ and $\sigma'(c_i) = \tilde{\sigma}(c_i)$ for every i where $\tilde{\sigma} = a_m a_1 \ldots a_{m-1}$.

Let $\tau : X \to X$ be the unique bijection acting on the *ordering* induced by $\tilde{\sigma}$ that yields the ordering induced by σ' written as $\tau(\tilde{\sigma}) = \sigma'$. By arguments similar to Case 1 it is clear that

$$|\{\sigma \in \mathcal{L}(c_i|C) : d_K(\tilde{\sigma}, \sigma) = \delta\}| = |\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma', \sigma) = \delta\}|$$

for every possible distance δ and every i and thus by Case 2

$$|\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma^*, \sigma) = \delta\}| = |\{\sigma \in \mathcal{L}(c_i|C) : d_K(\sigma', \sigma) = \delta\}|$$

which proves the result for Case 3.

Case 4: all other cases.

Note that if $a_{m-1} \notin C$ then using $\tilde{\sigma} = a_{m-1} a_m a_1 \ldots a_{m-2}$ and repeating the arguments from Case 2 and Case 3 would establish the result for all σ' such that $\sigma'(c_i) = \tilde{\sigma}(c_i)$ for every i. It is straightforward to see how continuing to cycle

in this manner will stop once the result has been established for all σ' such that $\sigma'(c_s) = m$.

Additionally, if $a_1 \notin C$, then considering $\tilde{\sigma} = a_2 \ldots a_m a_1$ and arguments as in Cases 2 and 3 will establish the result for all σ' such that $\sigma'(c_i) = \tilde{\sigma}(c_i)$ for every i. Continuing to cycle in this manner will stop once the result has been established for all σ' such that $\sigma'(c_1) = 1$. This concludes the proof of Case 4 and Lemma 3.3.2. $\qquad \square$

Proof of Proposition 3.3.4. If $C = D$ then the result is trivially true, so suppose $C \neq D$. The proof proceeds in cases.

Case 1: $c_1 = a_1$ and $d_1 = a_2$.

Consider the ranking $\sigma' = a_2 \ldots a_m a_1$. I claim that

$$|\{\sigma \in \mathcal{L}(c_i | C) : d_K(\sigma^*, \sigma) = \delta\}| = |\{\sigma \in \mathcal{L}(d_i | D) : d_K(\sigma', \sigma) = \delta\}| \qquad (3.2)$$

for each possible distance δ and for every i. To prove this, take any i and possible δ such that the former set is nonempty, and let σ be any ranking in the former set. Let $\tau : X \to X$ be the unique bijection acting on the *ordering* induced by σ^* that yields the ordering induced by σ' written as $\tau(\sigma^*) = \sigma'$. Note that $\tau(\sigma) \in \mathcal{L}(d_i | D)$ since c_i maps to d_i for every i under τ. Further, by right-invariance $d_K(\sigma^*, \sigma) = d_K(\sigma', \tau(\sigma)) = \delta$, which proves (3.2) holds. Further, applying Lemma 3.3.2 to choice set D and using (3.2), it is straightforward to see that

$$|\{\sigma \in \mathcal{L}(d_i | D) : d_K(\sigma^*, \sigma) = \delta\}| = |\{\sigma \in \mathcal{L}(d_i | D) : d_K(\sigma', \sigma) = \delta\}|$$
$$= |\{\sigma \in \mathcal{L}(c_i | C) : d_K(\sigma^*, \sigma) = \delta\}|$$

for every possible distance δ and every i. Thus, $p(c_i | C) = p(d_i | D)$ since G is a

d_K-monotonic noise model, proving the result for Case 1.

Case 2: $c_1 = a_1$ and $d_1 \neq a_2$.

If $d_1 = a_3$ then a similar argument as in Case 1 using $\sigma' = a_3 \ldots a_m a_1 a_2$ shows that (3.2) holds. The result then holds by a simple application of Lemma 3.3.2 to choice set D and using the definition of d_K-monotonic noise model. It is easy to see how this process can continue for $d_1 = a_j$ for any $j \neq 1, 2, 3$ by using the ranking $\sigma' = a_j \ldots a_{j-2} a_{j-1}$ and proceeding as in Case 1. This establishes the result for Case 2.

Case 3: all other cases.

Now $c_1 \neq a_1$ and $d_1 \neq a_1$. Note that there exists a choice set E such that $|E| = s$, $e_1 = a_1$ (where the labeling is done according to ranking by σ^*), and $\sigma^*(e_{i+1}) - \sigma^*(e_i) = \sigma^*(c_{i+1}) - \sigma^*(c_i) = \sigma^*(d_{i+1}) - \sigma^*(d_i)$ for each $i \in \{1, \ldots, s-1\}$. Using Case 1 or 2 will establish the equalities $p(e_i|E) = p(c_i|C)$ and $p(e_i|E) = p(d_i|D)$ for every i, which yields the result for Case 3. This concludes the proof of Proposition 3.3.4. $\qquad\qquad\square$

3.3.2 Specification tests when σ^* is known - votes from the set X

Using the previous results, I now propose some simple statistical testing procedures that can be used to refute models in the class $\mathcal{G}$. I assume that the true ranking σ^* is known and that the available dataset ω consists of n votes from the choice set X. I will let Y_i be the random variable of total votes for a_i and let $p_i = p(a_i|X)$. Note that the random vector $Y = (Y_1, \ldots, Y_m)$ has a multinomial distribution with parameters $p = (p_1, \ldots, p_m)$ and n. Further, by Proposition 3.3.2, if votes are generated according to some $G \in \mathcal{G}$, then it must be the case that

$p_1 \geq p_2 \geq \cdots \geq p_m$. Thus, a rejection of the hypothesis

$$H_0 : p_1 \geq p_2 \geq \cdots \geq p_m > 0 \tag{3.3}$$

is sufficient for rejecting the hypothesis of a d-monotonic noise model.[6] At the risk of stating the obvious, failure to reject such a hypothesis does not imply consistency with some $G \in \mathcal{G}$ since a d-monotonic noise model will in general be incapable of generating all vote probability vectors consistent with (3.3).

Chafai and Concordet test

The first test I present adapts a procedure from Chafai and Concordet (2009) to the current context, which I refer to as the CC test. Chafai and Concordet propose confidence regions for a multinomial probability vector that control coverage probability and have small volume even in small samples. Let

$$\Lambda_m = \left\{ (q_1, \ldots, q_m) \in [0,1]^m : \sum_{i=1}^{m} q_i = 1 \right\}$$

be the probability simplex and let

$$O_{m,n} = \left\{ (y_1, \ldots, y_m) \in \{0, \ldots, n\}^m : \sum_{i=1}^{m} y_i = n \right\}$$

be the set of potential realizations of an m-dimensional multinomial random vector with n trials. For $y \in O_{m,n}$ and $q \in \Lambda_m$, let $\mu_q(y)$ be the multinomial probability

---

[6]Testing a strict inequality version for a strictly swap-increasing d will generally be identical to the procedures for the weak inequality version.

mass function evaluated at y. Given $\alpha \in (0,1)$ let

$$u(q,\alpha) = \sup \left\{ u \in [0,1] : \sum_{\substack{y \in O_{m,n} \\ \mu_q(y) \geq u}} \mu_q(y) \geq 1 - \alpha \right\}.$$

The set of realizations $y \in O_{m,n}$ such that $\mu_q(y) < u(q,\alpha)$ are the most "extreme" (i.e. least likely) possible given α and q. For a realization of Y, Chafai and Concordet propose the confidence region for p given by

$$CR_\alpha(Y) = \{q \in \Lambda_m : \mu_q(Y) \geq u(q,\alpha)\}. \tag{3.4}$$

The set in (3.4) is all probability vectors for which the realization of Y is not one of the most extreme possible.

Adapting the above method to test (3.3) is straightforward. Consider the subset of the probability simplex consistent with H_0:

$$\Lambda_{m,H_0} = \{q \in \Lambda_m : q_1 \geq \cdots \geq q_m > 0\}.$$

Given a sample size $n > 0$ and $\alpha \in (0,1)$, the CC test for H_0 has a rejection region given by

$$R_{CC}(n,\alpha) = \{y \in O_{m,n} : CR_\alpha(y) \cap \Lambda_{m,H_0} = \emptyset\}.$$

This is the set of realizations for which the confidence region proposed by Chafai and Concordet contains no probability vectors consistent with (3.3). The CC test is given by

$$\psi_{CC}(Y) = \mathbb{1}\{Y \in R_{CC}(n,\alpha)\}.$$

One disadvantage of the CC test is that it can be computationally demanding,

particularly if m or n or both are large. However, it should be used to construct exact confidence sets for the true ranking σ^* when it is unknown and votes are generated by some $G \in \mathcal{G}$ (see Section 3.4.3).

Conditional binomial test

I now propose a test that relies on a series of binomial tests by conditioning on the vote counts of the alternatives. A well known fact about multinomial distributions is that the conditional distribution of (Y_1, Y_2) given $(Y_3, \ldots, Y_m) = (y_3, \ldots, y_m)$ is equivalent to a binomial distribution with parameters $n - \sum_{i=3}^{m} y_i$ and $\frac{p_1}{p_1+p_2}$.[7] According to (3.3) the probability of "success" (a vote for a_1) from this distribution should be at least $\frac{1}{2}$. Thus, a rejection of $\frac{p_1}{p_1+p_2} \geq \frac{1}{2}$ is sufficient for rejecting the hypothesis of a d-monotonic noise model. Intuitively, it should not be the case that a_2 is voted for significantly more often than a_1 is when considering votes for only that pair.

Rewrite (3.3) as

$$H_0 : \frac{p_i}{p_i + p_{i+1}} \geq \frac{1}{2} \text{ for all } i \in \{1, \ldots, m-1\}. \tag{3.5}$$

Given $n > 0$ and $\alpha \in (0, 1)$ the CB test specifies the rejection region

$$R_{CB}(n, \alpha) =$$
$$\left\{ y \in O_{m,n} : \left(\frac{1}{2}\right)^{y_i+y_{i+1}} \sum_{\ell=0}^{y_i} \binom{y_i + y_{i+1}}{\ell} \leq \frac{\alpha}{m-1} \text{ for any } i \leq m-1 \right\}.$$

The rejection region specifies rejecting in any case in which a_i receives significantly less than half of the votes out of the pair $\{a_i, a_{i+1}\}$ assuming equal probability of

---

[7]Note that this type of conditioning will work for any pair of the random vector Y.

voting for either alternative. The CB test is given by

$$\psi_{CB}(Y) = \mathbb{1}\{Y \in R_{CB}(n, \alpha)\}.$$

Both the CC and CB tests control Type I error probability even in finite samples.

Theorem 3.3.1. *Given $n > 0$ and $\alpha \in (0, 1)$ both the CC and CB statistical tests are level α tests.*

Proof of Theorem 3.3.1. I first prove the result for the CC test. Consider the sub hypothesis $H_{0,q} : p = q$ for some $q \in \Lambda_{m,H_0}$. Specify a rejection region for a test of this hypothesis by

$$R_q(n, \alpha) = \{y \in O_{m,n} : q \notin CR_\alpha(y)\}.$$

The rejection region for the CC test can now be rewritten as

$$R_{CC}(n, \alpha) = \bigcap_{q \in \Lambda_{m,H_0}} R_q(n, \alpha).$$

By construction

$$\Pr(Y \in R_q(n, \alpha)) \leq \alpha$$

for each $q \in \Lambda_{m,H_0}$. If H_0 is true, it follows that $p = q'$ for some $q' \in \Lambda_{m,H_0}$ and therefore

$$\Pr(Y \in R_{CC}(n, \alpha)) \leq \Pr(Y \in R_{q'}(n, \alpha)) \leq \alpha.$$

I now prove the result for the CB test. For each $i \in \{1, \ldots, m - 1\}$ let $R_{CB,i}(n, \alpha) \subset R_{CB}(n, \alpha)$ be the realizations in which $\left(\frac{1}{2}\right)^{y_i + y_{i+1}} \sum_{\ell=0}^{y_i} \binom{y_i + y_{i+1}}{\ell} \leq \frac{\alpha}{m-1}$. Let Y_{-i} be the random vector Y without Y_i and Y_{i+1}. The set of potential

realizations of this vector for all i is $O_{m-2,n}$. By the law of total probability it follows that

$$\Pr\left(Y \in R_{CB,i}(n,\alpha)\right) = \sum_{y_{-i} \in O_{m-2,n}} \Pr\left(Y \in R_{CB,i}(n,\alpha)|Y_{-i} = y_{-i}\right)\Pr\left(Y_{-i} = y_{-i}\right).$$

$$(3.6)$$

Let $(y_i, y_{i+1}, y_{-i}) = (y_1, \ldots, y_i, y_{i+1}, \ldots, y_m)$. For each $y_{-i} \in O_{m-2,n}$, define $y_i^*(y_{-i}) = \sup\{y_i : (y_i, y_{i+1}, y_{-i}) \in R_{CB,i}(n,\alpha)\}$.[8] Given $Y_{-i} = y_{-i}$, it follows that $Y \in R_{CB,i}(n,\alpha)$ if and only if $Y_i \leq y_i^*(y_{-i})$. Thus, (3.6) becomes

$$\Pr\left(Y \in R_{CB,i}(n,\alpha)\right) = \sum_{y_{-i} \in O_{m-2,n}} \Pr\left(Y_i \leq y_i^*(y_{-i})|Y_{-i} = y_{-i}\right)\Pr\left(Y_{-i} = y_{-i}\right).$$

$$(3.7)$$

Note that $\Pr\left(Y_i \leq y_i^*(y_{-i})|Y_{-i} = y_{-i}\right)$ is the cumulative distribution function of a binomial random variable, which is decreasing in its success probability $\frac{p_i}{p_i + p_{i+1}}$. Thus, when (3.5) is true, it follows by definition of $y_i^*(y_{-i})$ that

$$\Pr\left(Y \in R_{CB,i}(n,\alpha)\right) = \sum_{y_{-i} \in O_{m-2,n}} \Pr\left(Y_i \leq y_i^*(y_{-i})|Y_{-i} = y_{-i}\right)\Pr\left(Y_{-i} = y_{-i}\right)$$

$$\leq \frac{\alpha}{m-1} \sum_{y_{-i} \in O_{m-2,n}} \Pr\left(Y_{-i} = y_{-i}\right) \leq \frac{\alpha}{m-1}.$$

---

[8] By convention $\sup(\emptyset) = -\infty$.

Using this to conclude the proof, note that when (3.5) holds

$$\Pr\left(Y \in R_{CB}(n,\alpha)\right) \leq \Pr\left(\bigcup_{i \in \{1,\ldots,m-1\}} Y \in R_{CB,i}(n,\alpha)\right)$$

$$\leq \sum_{i \in \{1,\ldots,m-1\}} \Pr\left(Y \in R_{CB,i}(n,\alpha)\right)$$

$$\leq (m-1)\left(\frac{\alpha}{m-1}\right) = \alpha.$$

$\square$

For large m, n it is clear to see that the CB test is more convenient for implementation purposes, and so there is reason to prefer it in practice. However, as mentioned earlier, the CC method is convenient for constructing exact confidence sets for σ^*.

I chose to focus on testing for violations of Proposition 3.3.2 since it is the easiest condition to test of those presented. It is straightforward to test the condition *across* choice sets using the independence of votes across choice sets. For example, suppose the data consists of votes from two binary menus. Testing for a violation of Proposition 3.3.2 would then involve a binomial test at significance level $\frac{\alpha}{2}$ in each menu to achieve an overall significance level α. If the dataset is rich enough, more complicated testing procedures involving the conditions from other Propositions is possible, although developing reasonable tests that control size in finite samples is difficult. An asymptotic approach to testing various inequality restrictions on multiple multinomial probability parameters has been developed. I refer the interested reader to Davis-Stober (2009).

3.4   Identification and confidence sets

The first half of this section concerns identifying parameters of the model under the assumption that a researcher observes the population vote probabilities $p(a|A)$ for all $(A, a) \in \mathcal{O}$ generated by a noise model $G \in \mathcal{G}$. I say that a ranking σ is *identified as the true ranking from vote probabilities* if the vote probabilities are inconsistent with any other ranking σ' being the true ranking. I say the model $G \in \mathcal{G}$ is *identified from vote probabilities* if $G \neq G'$ implies that $p^G(a|A) \neq p^{G'}(a|A)$ for some observation (A, a) where I have added superscripts to distinguish between vote probabilities generated by G and G'.

3.4.1   Identification of the true ranking

The first result shows that the population vote probabilities always identify a unique true ranking. The intuition follows directly from Proposition 3.3.2 and Corollary 3.3.1: the true ranking will always reveal itself by placing the highest vote probability on the highest ranked alternative in any choice set, yielding asymmetric and transitive vote probabilities.

Theorem 3.4.1. *Define the binary relation τ as: for any $a, b \in X$, $a \succ_\tau b$ if and only if there exists a choice set $A \supseteq \{a, b\}$ such that $p(a|A) > p(b|A)$. Then the ranking according to τ is identified as the true ranking from vote probabilities.*

Proof of Theorem 3.4.1. Since voting probabilities are generated by some $G \in \mathcal{G}$, it is easy to see that $\tau \in \mathcal{L}(X)$ by appealing to Propositions 3.3.2 and 3.3.3 and Corollary 3.3.1. Take any ranking $\sigma \neq \tau$ so that there is a pair a, b such that $a \succ_\tau b$ but $b \succ_\sigma a$. If $\sigma = \sigma^*$ were true (i.e. the true ranking were σ), then it follows by Corollary 3.3.1 that $p(b|\{a, b\}) > p(a|\{a, b\})$, a contradiction to the definition of τ and thus inconsistent with the vote probabilities. Since σ was arbitrary, it follows

that τ is identified as the true ranking from vote probabilities. $\qquad\square$

3.4.2 Identification of the distribution over rankings

I now turn to the task of identifying the distribution over rankings as specified by G. Several papers have discussed the general impossibility of point identification of a probability distribution over rankings from observing choice probabilities (Falmagne, 1978; Barbera and Pattanaik, 1986; Sher, Fox, Kim, and Bajari, 2011). In particular, Sher, Fox, Kim, and Bajari quantify the extent of underidentification, which can be severe for large m.[9]

To keep the analysis tractable, I focus on the case $G \in \mathcal{G}_{d_K}$ i.e. Kendall tau monotonic noise models. Restricting the analysis to this class of models greatly reduces the number of parameters that need to be identified in order to point identify the entire distribution. Specifically, identifying the likelihood of any ranking of distance d from σ^* for each $d \in \left\{0, \ldots, \frac{m(m-1)}{2}\right\}$ will suffice. This seemingly manageable task unfortunately is not possible when $m > 3$.

Theorem 3.4.2. *If $m > 3$, then for every $G \in \mathcal{G}_{d_K}$ there exists a $G' \in \mathcal{G}_{d_K}$ with $G' \neq G$ such that $p^G(a|A) = p^{G'}(a|A)$ for every $(A, a) \in \mathcal{O}$.*

The proof relies on the following lemma.

Lemma 3.4.1. *If $m > 3$ then for every $(A, a) \in \mathcal{O}$ it follows that*

$$\left|\{\sigma \in \mathcal{L}(a|A) : d_K(\sigma, \sigma^*) \text{ is even}\}\right| = \left|\{\sigma \in \mathcal{L}(a|A) : d_K(\sigma, \sigma^*) \text{ is odd}\}\right|.$$

Proof of Lemma 3.4.1. The proof uses concepts and notation introduced in Appendix A. Take any choice set A with $|A| = s$ for some $s \leq m$. For $m > 3$ and any

---

[9]Specifically, the number of potential mass points of the distribution is the number of rankings $m!$, whereas the number of observable population moments is $\sum_{i=2}^{m} i\binom{m}{i} = \left(2^{m-1} - 1\right) m$.

$a \in A$ it is obvious that $|\mathcal{L}(a|A)| = \frac{m!}{s}$ is an even number and thus divisible by 2, so it is always possible for the conjecture of the lemma to hold.

Let $\tilde{a} \in A$ be the best alternative in choice set A according to σ^*. Take any $\sigma \in \mathcal{L}(\tilde{a}|A)$ and consider $\sigma_{\tilde{a} \leftrightarrow b} \in \mathcal{L}(b|A)$ for some $b \in A \setminus \{\tilde{a}\}$. Note that the ordering induced by $\sigma_{\tilde{a} \leftrightarrow b}$ can be expressed as the composition of a permutation (the unique $\pi \in S_m$ associated with σ) and a transposition, the latter simply switching the ranks of $\tilde{a}$ and b. It follows that the permutations associated with σ and $\sigma_{\tilde{a} \leftrightarrow b}$ are of different parity. Therefore, if $d_K(\sigma, \sigma^*)$ is even (odd) then $d_K(\sigma_{\tilde{a} \leftrightarrow b}, \sigma^*)$ is odd (even). It is straightforward to verify that there are both even and odd permutations associated with the rankings in $\mathcal{L}(\tilde{a}|A)$. Thus, since σ was arbitrary it follows that for every ranking in $\mathcal{L}(\tilde{a}|A)$ of an even (odd) distance from σ^* there is a unique ranking in $\mathcal{L}(b|A)$ of an odd (even) distance from σ^*. Since b was arbitrary, the previous conclusion holds for every $a \in A \setminus \{\tilde{a}\}$. The conjecture of the lemma will follow by establishing an equal number of odd-distanced and even-distanced rankings in $\mathcal{L}(\tilde{a}|A)$. For $s \geq 3$ this obviously holds - otherwise there would be an unequal number of even and odd permutations in S_m.

To establish the result for $s = 2$, I provide a constructive proof. To this end, let $|A| = 2$ and take any $\sigma \in \mathcal{L}(\tilde{a}|A)$ such that $\sigma(\tilde{a}) = r$ for some $r \leq m - 1$. Construct $\sigma' \in \mathcal{L}(\tilde{a}|A)$ by fixing the $r - 1$ alternatives that are ranked better than $\tilde{a}$ in σ and permuting only the $m - r$ alternatives ranked worse than $\tilde{a}$ in σ. Note that such a permutation π is part of the symmetric group of a set with $m - r$ elements which is a subset of S_m i.e. $\pi \in S_{m-r} \subset S_m$. There are an equal number of even and odd permutations in S_{m-r} (including the identity permutation, which yields σ) so applying each of these yields an equal number of even-distanced and odd-distanced rankings contributed to $\mathcal{L}(\tilde{a}|A)$. Since the $r - 1$ alternatives ranked better than $\tilde{a}$ in σ were arbitrary, the preceding holds for every distinct instance of

possible $r-1$ alternatives being ranked ahead of $\tilde{a}$ in the set $\mathcal{L}(\tilde{a}|A)$. Conclude that there are an equal number of even-distanced and odd-distanced rankings in $\mathcal{L}(\tilde{a}|A)$ with $\tilde{a}$ in rank $r \leq m-1$. Since r was arbitrary, conclude that there are an equal number of even-distanced and odd-distanced rankings in $\mathcal{L}(\tilde{a}|A)$. This concludes the proof. □

Proof of Theorem 3.4.2. Let $m > 3$ and $G \in \mathcal{G}_{d_K}$. I will construct a $G' \in \mathcal{G}_{d_K}$ with $G' \neq G$ such that $p^G(a|A) = p^{G'}(a|A)$ for every $(A,a) \in \mathcal{O}$. Let σ' be the reverse ranking of σ^* i.e. the unique least likely ranking according to G. Define

$$\delta^* = \min_{\substack{\sigma, \tau \in \mathcal{L}(X) \\ |d_K(\sigma,\sigma^*) - d_K(\tau,\sigma^*)|=1}} \left| \mathrm{Pr}^G(\sigma|\sigma^*) - \mathrm{Pr}^G(\tau|\sigma^*) \right|$$

so that $\delta^* > 0$ is the smallest difference in likelihood between any two rankings of different distances from σ^*. Let $\varepsilon > 0$ satisfy

$$\varepsilon < \min\left\{ \frac{\delta^*}{2}, \mathrm{Pr}^G(\sigma'|\sigma^*) \right\}.$$

Note that such an ε exists due to the full support assumption on G.

Define G' by

$$\mathrm{Pr}^{G'}(\sigma|\sigma^*) = \mathrm{Pr}^G(\sigma|\sigma^*) + (-1)^{d_K(\sigma,\sigma^*)} \times \varepsilon$$

for every $\sigma \in \mathcal{L}(X)$ so that every even-distanced ranking from σ^* gets ε mass added to its likelihood in G and every odd-distanced ranking from σ^* gets ε mass taken away from its likelihood in G. There are an equal number of even-distanced and odd-distanced rankings in $\mathcal{L}(X)$, so the net effect of this mass transfer is zero. Further, ε was constructed in such a way so that $\mathrm{Pr}^{G'}(\sigma|\sigma^*) > 0$ for every $\sigma \in \mathcal{L}(X)$.

Thus G' is a valid probability distribution on $\mathcal{L}(X)$. Additionally, $G' \in \mathcal{G}_{d_K}$ since it was constructed from G and an ε that preserves the monotonicity with respect to d_K.

By Lemma 3.4.1, for any $(A, a) \in \mathcal{O}$ there are an equal number of even-distanced and odd-distanced rankings from σ^* in $\mathcal{L}(a|A)$. Thus, the net effect on the voting probabilities of the mass transfer in the construction of G' is zero, so that $p^G(a|A) = p^{G'}(a|A)$ for all $(A, a) \in \mathcal{O}$. $\hspace{2em}\square$

The proof of Theorem 3.4.2 relies on some results from group theory (see Appendix A). In the proof, I construct a new distribution on $\mathcal{L}(X)$ from the distribution specified by G. The proof hinges on the fact that there an equal number of even-distanced and odd-distanced rankings from σ^* in $\mathcal{L}(X)$ and the set $\mathcal{L}(a|A)$ for every $(A, a) \in \mathcal{O}$ when $m > 3$. Thus I can shift mass from odd-distanced rankings to even-distanced rankings systematically while preserving the monotonicity with respect to d_K and the population vote probabilities. The property does not hold for $m = 3$ since the cardinality of the set of rationalizing rankings on binary choice sets is $\frac{3!}{2} = 3$. From a linear algebra perspective, the number of unknowns is $\frac{m(m-1)}{2} + 1$ i.e. the number of possible distances, but the number of linearly independent equations is at most $\frac{m(m-1)}{2}$ when $m > 3$. For $m = 3$, point identification of any distribution on rankings is possible from observation of the population vote probabilities.[10]

A linear programming approach

In light of Theorem 3.4.2, I present a linear program whose solutions constitute the identified set $\{G\} \subset \mathcal{G}_{d_K}$ given the population vote probabilities. A similar approach has been provided by Sher, Fox, Kim, and Bajari (2011) and Manski (2007).

---

[10]See Sher, Fox, Kim, and Bajari (2011) for a discussion.

Define $\sigma^* \in \mathcal{L}(X)$ as $a \succ_{\sigma^*} b \iff p(a|\{a,b\}) > p(b|\{a,b\})$ for any $a, b \in X$. The program is given by the following linear system:

$$p(a|A) = \sum_{\sigma \in \mathcal{L}(a|A)} \Pr^G(\sigma|\sigma^*), \ (A, a) \in \mathcal{O}, \tag{3.8}$$

$$\sum_{\sigma \in \mathcal{L}(X)} \Pr^G(\sigma|\sigma^*) = 1; \ \Pr^G(\sigma|\sigma^*) > 0, \ \sigma \in \mathcal{L}(X), \tag{3.9}$$

$$G \in \mathcal{G}_{d_K}. \tag{3.10}$$

The noise models solving (3.8), (3.9), and (3.10) constitute the identified set in this framework. It is easy to adapt the program to find probabilistic bounds on any set of rankings $S \subset \mathcal{L}(X)$. For example, a researcher could be interested in knowing an upper bound on the probability of realizing a ranking of at most distance δ away from σ^*. This is given by

$$\sup_{G \in \mathcal{G}_{d_K}} \sum_{\substack{\sigma \in \mathcal{L}(X) \\ d_K(\sigma, \sigma^*) \leq \delta}} \Pr^G(\sigma|\sigma^*) \quad \text{subject to (3.8), (3.9), and (3.10).}$$

3.4.3 Finite sample confidence sets

I now bring the results of the previous section into the context of a researcher with a finite dataset $\omega \in \mathcal{O}^n$ generated by some $G \in \mathcal{G}_{d_K}$. That is, the researcher does not observe the population vote probabilities. My concern here is with the construction of confidence sets for the underlying distribution on rankings that have a desired coverage probability. The results here are derived in an analogous fashion to those in Manski (2007).

Votes from the set X

I start with the simplest case in which the dataset consists of n votes from the set X. Given a significance level $\alpha \in (0, 1)$, the researcher forms a confidence set for the probability vector $(p(a|X), a \in X)$ using the method proposed in Chafai and Concordet (2009) which is detailed in Section 3.3.2. For simplicity, assume that all of the vote probability vectors in this confidence set are consistent with only one ranking in $\mathcal{L}(X)$. Denote this ranking as σ'. By definition, $\{\sigma'\}$ is a confidence set (in this case, a singleton) for the true ranking σ^* with coverage probability at least $1 - \alpha$. Let $Y = (y_1, \ldots, y_m)$ be the realized vector of votes for each alternative ordered sequentially from highest to lowest count so that the ordering is consistent with σ'. Let $CR_\alpha(Y)$ be the confidence set for the vote probability vector. Take any $q = (q_1, \ldots, q_m) \in CR_\alpha(Y)$ and let G_q be the set of noise models satisfying:

$$q_i = \sum_{\sigma \in \mathcal{L}(i|X)} \mathrm{Pr}^G\left(\sigma|\sigma'\right), \ 1 \le i \le m,$$

$$\sum_{\sigma \in \mathcal{L}(X)} \mathrm{Pr}^G\left(\sigma|\sigma'\right) = 1; \ \mathrm{Pr}^G\left(\sigma|\sigma'\right) > 0, \ \sigma \in \mathcal{L}(X),$$

$$G \in \mathcal{G}_{d_K}.$$

Then $G_{CR_\alpha(Y)} = \cup_{q \in CR_\alpha(Y)} G_q$ is a confidence set for the distribution over rankings with coverage probability at least $1 - \alpha$.

Votes from all binary choice sets

I now consider the case of votes from all binary choice sets. For simplicity assume $\omega \in \mathcal{O}^n$ contains an equal number of votes from each binary choice set. The method proceeds by first constructing a confidence interval for either vote

probability in each binary menu, which can be done again using the method of Chafai and Concordet (2009). For example, take an arbitrary binary choice set $\{a, b\}$ and suppose the researcher chooses to construct a confidence interval for $p(a|\{a, b\})$. Let $y_{a,b}$ be the realized vote count for a from this choice set and $\alpha_{a,b}$ a specified significance level, where I will explain how to choose $\alpha_{a,b}$ shortly. Let $CR_{\alpha_{a,b}}(y_{a,b})$ be the realized confidence interval constructed for $p(a|\{a, b\})$ that has coverage probability at least $1 - \alpha_{a,b}$. For each choice set, the researcher chooses $\alpha_{a,b}$ such that $\prod_{\{a,b\} \subset X}(1 - \alpha_{a,b}) \geq 1 - \alpha$ where $\alpha \in (0, 1)$ is an overall desired significance level. Since the vote counts across menus are independent, the confidence rectangle defined by the Cartesian product

$$CR_\alpha = \underset{\{a,b\} \subset X}{\times} CR_{\alpha_{a,b}}(y_{a,b})$$

has coverage probability at least $1 - \alpha$ for the vector $(p(a|\{a, b\}), \{a, b\} \subset X)$. Assume that this construction is consistent with only one ranking $\sigma' \in \mathcal{L}(X)$. Take any $q = \left(q_{(a|\{a,b\})}, \{a, b\} \subset X\right) \in CR_\alpha$ and let G_q be the set of noise models satisfying:

$$q_{(a|\{a,b\})} = \sum_{\sigma \in \mathcal{L}(a|\{a,b\})} \Pr^G(\sigma|\sigma'), \ \{a, b\} \subset X,$$

$$\sum_{\sigma \in \mathcal{L}(X)} \Pr^G(\sigma|\sigma') = 1; \ \Pr^G(\sigma|\sigma') > 0, \ \sigma \in \mathcal{L}(X),$$

$$G \in \mathcal{G}_{d_K}.$$

Then $G_{CR_\alpha} = \cup_{q \in CR_\alpha} G_q$ is a confidence set for the distribution over rankings with coverage probability at least $1 - \alpha$.

It is straightforward to see how the previous constructions can be generalized to accommodate other datasets including those whose realized vote counts may generate a non-singleton confidence set for the true ranking σ^*.

3.5 Estimators and choice designs

The first portion of this section is devoted to estimation of the best alternative or true ranking. The theory shows that for certain datasets a large class of estimators will almost surely return the correct best alternative or ranking as the sample size tends to infinity. I then investigate choice experiment designs using the tractable model of Mallows. Using a ubiquitous sampling scheme, I derive upper bounds on the number of samples needed to estimate either the best alternative or the true ranking with a high degree of probabilistic confidence.

3.5.1 Estimators

In cases where the true ranking is unknown, I assume the goal of the researcher is to estimate either the best alternative a_1 or the ranking σ^* based on the sampled votes. Past literature has explored this topic and by finding conditions under which common voting rules can be viewed as a maximum likelihood estimator (MLE) of the true ranking when observations consist of full rankings in $\mathcal{L}(X)$ (see e.g. Young, 1988; Conitzer and Sandholm, 2012; Procaccia, Reddi, and Shah, 2012). With the general class of noise models I am considering here, requiring an estimator to be an MLE is too stringent a task. Rather, I ask that an estimator return the true best alternative or true ranking with probability one given an infinite number of samples from any noise model $G \in \mathcal{G}$. Any such estimator will be "correct" with high probability given a large number of sampled votes, an appealing feature in line with the convenience of crowdsourcing systems.

I define two types of estimators. An *estimator for the top-ranked alternative* is any mapping $\hat{a}_n : \mathcal{O}^n \to 2^X \setminus \emptyset$. An *estimator for the true ranking* is any mapping $\hat{\sigma}_n : \mathcal{O}^n \to 2^{\mathcal{L}(X)} \setminus \emptyset$. Note that both types of estimators may be set-valued. Let Y_{ac} denote the random variable that counts the number of times alternative a was voted for over c in a dataset i.e. $Y_{ac} = |\{i \leq n : b_i = a \text{ and } c \in B_i\}|$. The estimator $\hat{a}_n$ is said to be *Condorcet consistent* if for any dataset such that there exists $a \in X$ with $Y_{ac} > Y_{ca}$ for all $c \neq a$ then $\hat{a}_n = a$. The estimator $\hat{\sigma}_n$ is said to be *pairwise consistent* if for any dataset such that there exists a ranking $\tilde{\sigma} \in \mathcal{L}(X)$ defined by $a \succ_{\tilde{\sigma}} c \iff Y_{ac} > Y_{ca}$ then $\hat{\sigma}_n = \tilde{\sigma}$.

The next results show that any estimators in these classes will almost surely converge to the correct best alternative and ranking respectively. The proofs rely on Proposition 3.3.2 and its Corollary, which provide the intuition for the results: as the number of votes solicited from any binary choice set tends to infinity, the true better alternative will emerge by almost surely receiving more votes.

Theorem 3.5.1. *Let $(B_1, b_1), (B_2, b_2), \ldots, (B_n, b_n)$ be a sequence of observations from some $G \in \mathcal{G}$ and $\omega_n = ((B_i, b_i))_{i \leq n}$ the corresponding datasets satisfying $|\{i \leq n : a, c \in B_i\}| \to \infty$ as $n \to \infty$ for every distinct $a, c \in X$. Then for any sequence of Condorcet consistent estimators $\{\hat{a}_n\}$ it follows that $\hat{a}_n \xrightarrow{a.s.} a_1$.*[11]

Proof of Theorem 3.5.1. For a choice set A and alternative $a \in A$ let $Y_{A,a} = |\{i \leq n : b_i = a \text{ and } B_i = A\}|$ be the number of times a was voted for from choice set A and let $n_A = |\{i \leq n : B_i = A\}|$ be the number of times choice set A appears in the dataset. By assumption, for a_1 and each other alternative c there is a choice set $D \supseteq \{a_1, c\}$ such that $n_D \to \infty$ as $n \to \infty$. Define

$\lambda_{D,a_1c} = \mathbb{E}\left[(Y_{D,a_1} - Y_{D,c})/n_D\right] = p(a_1|D) - p(c|D)$ which is strictly positive by Corollary 3.3.1. Thus,

$$\Pr\left(\frac{Y_{D,a_1} - Y_{D,c}}{n_D} \leq 0\right) \leq \Pr\left(\left|\frac{Y_{D,a_1} - Y_{D,c}}{n_D} - \mathbb{E}\left(\frac{Y_{D,a_1} - Y_{D,c}}{n_D}\right)\right| \geq \lambda_{D,a_1c}\right)$$

$$\leq 2e^{-2n_D\lambda_{D,a_1c}^2} \tag{3.11}$$

where the second transition follows from Hoeffding's inequality (Hoeffding, 1963). Since $\sum_{k=1}^{\infty} e^{-k} < \infty$ it follows from the Borel-Cantelli Lemma (Durrett, 2010) that almost surely $Y_{D,a_1} - Y_{D,c} > 0$ for all but finitely many values of n_D. Since D was arbitrary and since $Y_{a_1c} - Y_{ca_1} = \sum_{A \supseteq \{a_1,c\}} Y_{A,a_1} - Y_{A,c}$, almost surely $Y_{a_1c} - Y_{ca_1} > 0$ for all but finitely many observations. Since c was arbitrary, it follows that

$$\Pr\left(\bigcup_{c \neq a_1} Y_{a_1c} - Y_{ca_1} \leq 0 \text{ for infinitely many observations}\right) = 0$$

so that for all but finitely many observations, almost surely $Y_{a_1c} - Y_{ca_1} > 0$ for every $c \neq a$. By Condorcet consistency, almost surely $\hat{a}_n = a_1$ for all but finitely many n, which proves the result. $\square$

Note that the result stated in Theorem 3.5.1 holds with less strict requirement $|\{i \leq n : a_1, c \in B_i\}| \to \infty$ as $n \to \infty$ for every $c \neq a_1$. However, without a priori knowledge of a_1, the sufficient condition given in the theorem is better suited for data collection.

Theorem 3.5.2. *Let $(B_1, b_1), (B_2, b_2), \ldots, (B_n, b_n)$ be a sequence of observations from some $G \in \mathcal{G}$ and $\omega_n = ((B_i, b_i))_{i \leq n}$ the corresponding datasets satisfying $|\{i \leq n : B_i = \{a, c\}\}| \to \infty$ as $n \to \infty$ for every distinct $a, c \in X$. Then for any sequence of pairwise consistent estimators $\{\hat{\sigma}_n\}$ it follows that $\hat{\sigma}_n \xrightarrow{a.s.} \sigma^*$.*

Proof of Theorem 3.5.2. For a choice set A and alternative $a \in A$ let $Y_{A,a} = |\{i \leq n : b_i = a$ and $B_i = A\}|$ be the number of times a was voted for from choice set A and let $n_A = |\{i \leq n : B_i = A\}|$ be the number of times choice set A appears in the dataset. Let a, c be two distinct alternatives. By assumption $n_{\{a,c\}} \to \infty$ as $n \to \infty$. Suppose WLOG that $a \succ_{\sigma^*} c$. Using Corollary 3.3.1 and arguments similar to those in the proof of Theorem 3.5.1, almost surely $Y_{\{a,c\},a} - Y_{\{a,c\},c} > 0$ for all but finitely many values of $n_{\{a,c\}}$. For any other $D \supset \{a, c\}$ satisfying $n_D \to \infty$ as $n \to \infty$, it follows by Proposition 3.3.2 and a simple Strong Law of Large Numbers argument (Durrett, 2010) that almost surely $Y_{D,a} - Y_{D,c} \geq 0$ as $n \to \infty$. Thus, it follows that almost surely $Y_{ac} - Y_{ca} > 0$ as $n \to \infty$. Since a, c were arbitrary the preceding analysis applies to every pair of alternatives. Thus, by pairwise consistency it follows that almost surely $\hat{\sigma}_n = \sigma^*$ as $n \to \infty$. $\qquad \square$

3.5.2 Choice experiment design

As a natural counterpart to the previous section, I now investigate the efficiency of a commonly used choice experiment design in achieving a researcher's goal of estimating the best alternative or true ranking. Specifically, I analyze the number of samples needed to guarantee that the classes of estimators in the previous section return the best alternative or true ranking with probability at least $1 - \varepsilon$ for some $\varepsilon \in (0, 1)$. For tractability, I work with Mallows' model (introduced in Section 3.2).

I make use of a *binary choice design* in which votes are solicited from numerous binary choice sets. When approximating the best alternative, I assume the researcher first pairs two alternatives at random and keeps the winning alternative. The winner is then to be paired with another randomly chosen alternative. Thus, $m - 1$ such comparisons will need to be made, with the winner in each pair being

determined with high enough probability so that the overall achieved probabilistic confidence is $1 - \varepsilon$. When approximating the ranking σ^* I assume all $\binom{m}{2}$ pairwise comparisons are made.

The sufficient number of samples is presented as a function of $\phi, \varepsilon,$ and m.

Theorem 3.5.3. *Let $\varepsilon \in (0, 1)$. Then with Mallows' model, any Condorcet consistent estimator will determine the top ranked alternative with probability at least $1 - \varepsilon$ given $\frac{m-1}{2} \left\lceil \left(\frac{1+\phi}{1-\phi}\right)^2 \log\left(\frac{2m}{\varepsilon}\right)\right\rceil$ samples from the binary choice design. Further, any pairwise consistent estimator will determine the true ranking with probability at least $1 - \varepsilon$ given $\frac{m(m-1)}{4} \left\lceil \left(\frac{1+\phi}{1-\phi}\right)^2 \log\left(\frac{m(m-1)}{2\varepsilon}\right)\right\rceil$ samples from the binary choice design.*

Proof of Theorem 3.5.3. Let $\omega \in \mathcal{O}^n$ be a dataset of n observations generated from Mallows' model where each observation is a vote from some binary choice set. For any two alternatives a, c let $n_{\{a,c\}}$ be defined as in Theorem 3.5.1. Note that by the independence of votes across choice sets, identifying the better alternative in any of the $m - 1$ sampled binary choice sets with probability at least $1 - \frac{\varepsilon}{m}$ is sufficient for guaranteeing that a_1 is identified as the best with overall probability at least $1 - \varepsilon$. Using Propositions 3.3.3 and 3.3.4 it is easy to see that determining the better alternative from a binary choice set is most difficult for choice sets $\{a_i, a_{i+1}\}$ where $i \in \{1, \ldots, m - 1\}$. Further, by Proposition 3.3.2 it follows that $p(a_i | \{a_i, a_{i+1}\}) - p(a_{i+1} | \{a_i, a_{i+1}\}) = \frac{1-\phi}{1+\phi}$.

Let $Y_{A,a}$ be defined as in the proof of Theorem 3.5.1. By Condorcet consistency of the estimator, it is sufficient that $Y_{\{a,c\},a} - Y_{\{a,c\},c} > 0$ for any two paired alternatives a, c with $a \succ_{\sigma^*} c$. Using Hoeffding's inequality as in (3.11) it follows that

$$\Pr\left(\frac{Y_{\{a,c\},a} - Y_{\{a,c\},c}}{n} \leq 0\right) \leq 2e^{-2n_{\{a,c\}}\lambda^2_{\{a,c\},ac}} \leq 2e^{-2n_{\{a,c\}}\lambda^2_{\min}}$$

where $\lambda_{\min} = p(a_i|\{a_i, a_{i+1}\}) - p(a_{i+1}|\{a_i, a_{i+1}\}) = \frac{1-\phi}{1+\phi}$ and λ is defined as in the proof of Theorem 3.5.1. Thus, having $n_{\{a,c\}} \geq \frac{(1+\phi)^2}{2(1-\phi)^2} \log\left(\frac{2m}{\varepsilon}\right)$ is sufficient to guarantee that the probability of not identifying the better alternative from $\{a, c\}$ is at most $\frac{\varepsilon}{m}$. Collecting the sufficient number of samples in each of $m - 1$ binary choice sets gives the result for the binary choice design.

The results for estimating the correct ranking with pairwise consistent estimators are easily derived from the above results by collecting the sufficient number of samples in each of the $\binom{m}{2}$ binary menus where in each menu the better alternative is identified with probability at least $1 - \frac{\varepsilon}{\binom{m}{2}}$. $\qquad\square$

Although the sufficient bounds are complicated expressions, they provide useful guidance as to experiment designs, particularly if the researcher has some knowledge as to the value of ϕ. Future work should focus on deriving similar upper bounds for other popular sampling schemes such as soliciting votes from the entire set X. Additionally, it would be interesting to see if the results using Mallows' model can be used to bound sample sizes for more general models in the class $\mathcal{G}_{d_K}$.

Chapter 3 is solo authored and is being prepared for submission for publication.

3.6 Appendix: Preliminaries for Theorem 3.4.2

The following results from group theory are used in the proof of Theorem 3.4.2 (see e.g. Jacobson, 2012). Labeling the alternatives as $X = \{1, \ldots, m\}$, an alternative way to identify rankings in $\mathcal{L}(X)$ is to view them as permutations. A *permutation* is any bijective mapping $\pi : X \to X$ and the set of permutations on X is known as the *symmetric group* on X and denoted S_m. I associate each ranking $\sigma \in \mathcal{L}(X)$ with a unique permutation $\pi \in S_m$ by the relation $\sigma(i) = \pi^{-1}(i)$ for each i. Thus, $\pi(j)$ is the item with rank j. I label the identity permutation as π^* such that $\sigma^*(i) = \pi^{*-1}(i) = i$.

A *transposition* is a permutation which consists of permuting only two elements. It is known that every permutation $\pi \in S_m$ can be expressed as the composition of transpositions. Such a composition is generally not unique; however, every composition of a permutation will involve only an even number of transpositions or an odd number of transpositions. A permutation can be classified by its *parity* which is either even or odd depending on whether every composition involves an even or odd number of transpositions. There are an equal number of even and odd permutations in S_m i.e. $\frac{m!}{2}$. Since every transposition is an odd permutation, composing a permutation with a transposition results in a permutation of the opposite parity as the original.

The number of inversions of a permutation π relative to the identity permutation is defined as

$$\text{Inv}(\pi) = \left| \left\{ (i,j) : j > i \text{ and } \pi^{-1}(j) < \pi^{-1}(i) \right\} \right|.$$

If $\sigma \in \mathcal{L}(X)$ is the ranking associated with $\pi \in S_m$ then it is clear that $\text{Inv}(\pi) =$

$d_K(\sigma, \sigma^*)$. Further, the parity of a permutation can also be determined by the evenness or oddness of the number of inversions since each inversion can be considered a transposition of adjacent alternatives. Since there are an equal number of even and odd permutations in S_m, it follows that there are an equal number of even-distanced and odd-distanced rankings from σ^* in $\mathcal{L}(X)$. This fact is crucial to the proof of Theorem 3.4.2.

Bibliography

AGUIAR, V. (2015a): "Stochastic Choice and Attention Capacities: Inferring Preferences from Psychological Biases," *Available at SSRN 2607602*.

AGUIAR, V. H. (2015b): "Stochastic choice and categorization: Inferring preferences from psychological biases," Working Paper, Brown University.

ALEDO, J. A., J. A. GÁMEZ, AND D. MOLINA (2013): "Computing the Consensus Permutation in Mallows Distribution by Using Genetic Algorithms," in *Recent Trends in Applied Artificial Intelligence*, pp. 102–111. Springer.

——— (2016): "Using metaheuristic algorithms for parameter estimation in generalized Mallows models," *Applied Soft Computing*, 38, 308–320.

BALDIGA, K. A., AND J. R. GREEN (2013): "Assent-maximizing social choice," *Social Choice and Welfare*, 40(2), 439–460.

BARANCHUK, N., AND P. H. DYBVIG (2009): "Consensus in diverse corporate boards," *Review of Financial Studies*, 22(2), 715–747.

BARBERÀ, S., AND B. GRODAL (2011): "Preference for flexibility and the opportunities of choice," *Journal of Mathematical Economics*, 47(3), 272–278.

BARBERA, S., AND P. K. PATTANAIK (1986): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica: Journal of the Econometric Society*, pp. 707–715.

BARBERA, S., AND B. PELEG (1990): "Strategy-proof voting schemes with continuous preferences," *Social choice and welfare*, 7(1), 31–38.

BERGA, D., AND B. MORENO (2009): "Strategic requirements with indifference: single-peaked versus single-plateaued preferences," *Social Choice and Welfare*, 32(2), 275–298.

BLACK, D. (1948): "On the rationale of group decision-making," *The Journal of Political Economy*, 56(1), 23–34.

BLOCK, H. D., AND J. MARSCHAK (1960): "Random orderings and stochastic theories of responses," *Contributions to probability and statistics*, pp. 97–132.

BOSSERT, W., AND T. STORCKEN (1992): "Strategy-proofness of social welfare functions: the use of the Kemeny distance between preference orderings," *Social Choice and Welfare*, 9(4), 345–360.

BRADY, R. L., AND C. P. CHAMBERS (2015): "Spatial implementation," *Games and Economic Behavior*, 94, 200–205.

BUSA-FEKETE, R., E. HÜLLERMEIER, AND B. SZÖRÉNYI (2014): "Preference-based rank elicitation using statistical models: The case of Mallows," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1071–1079.

CARAGIANNIS, I., A. D. PROCACCIA, AND N. SHAH (2013): "When do noisy votes reveal the truth?," in *Proceedings of the fourteenth ACM conference on Electronic commerce*, pp. 143–160. ACM.

——— (2014): "Modal ranking: A uniquely robust voting rule," AAAI.

CERVONE, D. P., R. DAI, D. GNOUTCHEFF, G. LANTERMAN, A. MACKENZIE, A. MORSE, N. SRIVASTAVA, AND W. S. ZWICKER (2012): "Voting with rubber bands, weights, and strings," *Mathematical Social Sciences*, 64(1), 11–27.

CHAFAI, D., AND D. CONCORDET (2009): "Confidence regions for the multinomial parameter with small sample size," *Journal of the American Statistical Association*, 104(487), 1071–1079.

CHEN, X., P. N. BENNETT, K. COLLINS-THOMPSON, AND E. HORVITZ (2013): "Pairwise ranking aggregation in a crowdsourced setting," in *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 193–202. ACM.

CHIERICHETTI, F., AND J. KLEINBERG (2014): "Voting with limited information and many alternatives," *SIAM Journal on Computing*, 43(5), 1615–1653.

CHUNG, H., AND J. DUGGAN (2014): "Directional Equilibria," working paper.

COFFMAN, K. B. (2015): "Representative democracy and the implementation of majority-preferred alternatives," *Social Choice and Welfare*, pp. 1–18.

COLLINS-THOMPSON, K., AND J. P. CALLAN (2004): "A Language Modeling Approach to Predicting Reading Difficulty.," in *HLT-NAACL*, pp. 193–200.

CONDORCET, M. D. (1785): *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: De l'imprimerie royale.

CONITZER, V., M. ROGNLIE, AND L. XIA (2009): "Preference Functions that Score Rankings and Maximum Likelihood Estimation.," in *IJCAI*, vol. 9, pp. 109–115.

CONITZER, V., AND T. SANDHOLM (2012): "Common voting rules as maximum likelihood estimators," *arXiv preprint arXiv:1207.1368*.

COXETER, H. (1989): *Introduction to Geometry*. Wiley.

CRITCHLOW, D. E., M. A. FLIGNER, AND J. S. VERDUCCI (1991): "Probability models on rankings," *Journal of mathematical psychology*, 35(3), 294–318.

DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979): "The implementation of social choice rules: Some general results on incentive compatibility," *The Review of Economic Studies*, pp. 185–216.

DAVIS-STOBER, C. P. (2009): "Analysis of multinomial models under inequality constraints: Applications to measurement theory," *Journal of Mathematical Psychology*, 53(1), 1–13.

DEIMLING, K. (2011): *Nonlinear functional analysis*. Courier Dover Publications.

DIACONIS, P. (1988): "Group representations in probability and statistics," *Lecture Notes-Monograph Series*, 11, i–192.

DUGGAN, J. (2015): "May's Theorem in One Dimension," working paper.

DURRETT, R. (2010): *Probability: theory and examples*. Cambridge university press.

ECHENIQUE, F., K. SAITO, AND G. TSERENJIGMID (2013): "The perception-adjusted Luce model," Discussion paper, mimeo, California Institute of Techonology.

——— (2014): "The Perception-Adjusted Luce Model," Caltech Discussion paper.

ELKIND, E., P. FALISZEWSKI, AND A. M. SLINKO (2010): "Good Rationalizations of Voting Rules.," in *AAAI*, vol. 10, pp. 774–779.

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

FIORINI, S. (2004): "A short proof of a theorem of Falmagne," *Journal of mathematical psychology*, 48(1), 80–82.

FLIGNER, M. A., AND J. S. VERDUCCI (1986): "Distance based ranking models," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 359–369.

FUDENBERG, D., R. IIJIMA, AND T. STRZALECKI (2013): "Stochastic Choice and Revealed Perturbed Utility," Discussion paper, Mimeo Harvard University.

——— (2015): "Stochastic Choice and Revealed Perturbed Utility," *Econometrica*, 83(6), 2371–2409.

GINI, C., AND L. GALVANI (1929): "Di talune estensioni dei concetti di media a caratteri qualitivi," *Metron*, 8, 3–209.

GUL, F., P. NATENZON, AND W. PESENDORFER (2014): "Random Choice as Behavioral Optimization," *Econometrica*, 82(5), 1873–1912.

GÜLER, O. (2010): *Foundations of optimization*, vol. 258. Springer.

HALDANE, J. (1948): "Note on the median of a multivariate distribution," *Biometrika*, 35(3-4), 414–417.

HOEFFDING, W. (1963): "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, 58(301), 13–30.

HORAN, S. (2014): "Random Consideration and Choice," Working Paper, Université de Montréal.

HOTELLING, H. (1929): "Stability in competition," *The economic journal*, 39(153), 41–57.

JACOBSON, N. (2012): *Basic algebra I*. Courier Corporation.

Jiang, A., L. S. Marcolino, A. D. Procaccia, T. Sandholm, N. Shah, and M. Tambe (2014): "Diverse randomized agents vote to win," in *Advances in Neural Information Processing Systems*, pp. 2573–2581.

Kemeny, J. G. (1959): "Mathematics without numbers," *Daedalus*, 88(4), 577–591.

Kincaid, J. P., R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom (1975): "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," Discussion paper, DTIC Document.

Klamler, C. (2008): "A distance measure for choice functions," *Social Choice and Welfare*, 30(3), 419–425.

Lu, T., and C. Boutilier (2011): "Learning Mallows models with pairwise preferences," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 145–152.

Luce, R. D. (1959a): *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York.

Luce, R. D. (1959b): *Individual choice behavior: A theoretical analysis*. John Wiley and sons.

Machina, M. J. (1985): "Stochastic choice functions generated from deterministic preferences over lotteries," *The economic journal*, pp. 575–594.

Mallows, C. L. (1957): "Non-null ranking models," *Biometrika*, 44(1/2), 114–130.

Manski, C. F. (2007): "Partial identification of counterfactual choice probabilities," *International Economic Review*, 48(4), 1393–1410.

Manzini, P., and M. Mariotti (2014): "Stochastic choice and consideration sets," *Econometrica*, pp. 1153 – 1176.

Manzini, P., M. Mariotti, and L. Ülkü (2015): "Stochastic Complementarity," Working Paper, University of St. Andrews.

Mao, A., A. D. Procaccia, and Y. Chen (2013): "Better Human Computation Through Principled Voting.," in *AAAI*. Citeseer.

Maskin, E. (1999): "Nash equilibrium and welfare optimality," *The Review of Economic Studies*, 66(1), 23–38.

Mattsson, L.-G., and J. W. Weibull (2002): "Probabilistic choice and procedurally bounded rationality," *Games and Economic Behavior*, 41(1), 61–78.

May, K. O. (1952): "A set of independent necessary and sufficient conditions for simple majority decision," *Econometrica*, 20(4), 680–684.

Meila, M., K. Phadnis, A. Patterson, and J. A. Bilmes (2012): "Consensus ranking under the exponential model," *arXiv preprint arXiv:1206.5265*.

Muller, E., and M. A. Satterthwaite (1977): "The equivalence of strong positive association and strategy-proofness," *Journal of Economic Theory*, 14(2), 412–418.

Plott, C. R. (1967): "A notion of equilibrium and its possibility under majority rule," *American Economic Review*, 57(4), 787–806.

Procaccia, A. D., S. J. Reddi, and N. Shah (2012): "A maximum likelihood approach for selecting sets of alternatives," *arXiv preprint arXiv:1210.4882*.

Procaccia, A. D., and N. Shah (2015): "Optimal Aggregation of Uncertain Preferences," .

Rieskamp, J., J. R. Busemeyer, and B. A. Mellers (2006): "Extending the bounds of rationality: evidence and theories of preferential choice," *Journal of Economic Literature*, pp. 631–661.

Shafer, G. (1976): *A mathematical theory of evidence*. Princeton university press.

Shapley, L. S. (1953): "A value for n-person games," in *Contributions to the theory of games*, ed. by H. Kuhn, and A. Tucker, vol. II, pp. 307–317. Princeton University Press.

Sher, I., J. T. Fox, K. Kim, and P. Bajari (2011): "Partial identification of heterogeneity in preference orderings over discrete choices," Discussion paper, National Bureau of Economic Research.

Soufiani, H. A., W. Chen, D. C. Parkes, and L. Xia (2013): "Generalized method-of-moments for rank aggregation," in *Advances in Neural Information Processing Systems*, pp. 2706–2714.

Soufiani, H. A., D. C. Parkes, and L. Xia (2014): "Computing parametric ranking models via rank-breaking," in *Proceedings of The 31st International Conference on Machine Learning.* International Conference on Machine Learning.

Young, H. P. (1988): "Condorcet's theory of voting," *American Political Science Review*, 82(04), 1231–1244.

Young, H. P., and A. Levenglick (1978): "A consistent extension of Condorcet's election principle," *SIAM Journal on Applied Mathematics*, 35(2), 285–300.