FACULTAD ECONOMÍA Y NEGOCIOS UNIVERSIDAD DE CHILE
DEPARTAMENTO DE ECONOMÍA

SERIE DE DOCUMENTOS DE TRABAJO

SDT 397

Monotonic Aggregation of Preferences and the Rationalization of Choice Functions

Autores:
Jerry R. Green
Daniel A. Hojman

Santiago, Marzo de 2015

sdt@econ.uchile.cl
econ.uchile.cl/publicaciones

Monotonic Aggregation of Preferences and the Rationalization of Choice Functions

Jerry R. Green
Harvard University

Daniel A. Hojman
University of Chile

This Version: July 2014

Abstract

We consider a decision maker that holds multiple preferences simultaneously, each with different strengths described by a probability distribution. Faced with a subset of available alternatives, the preferences held by the individual can be in conflict. Choice results from an aggregation of these preferences. We assume that the aggregation method is monotonic: improvements in the position of alternative x cannot displace x if it were originally the choice. We show that choices made in this manner can be represented by context-dependent utility functions that are monotonic with respect to a measure of the strength of each alternative among those available. Using this representation we show that any generic monotonic rule can generate an arbitrary choice function as we vary the distribution of preferences. Domain restrictions on the set of preferences (e.g. dual motivation models) or consistency restrictions on the aggregator across choice sets reduce the set of admissible behaviors. Applications to positive models of individual decision making with context effects and social choice are discussed.

JEL Classification: D01, D11, D60
Keywords: voting, social choice, scoring rules, plurality, Borda count, decision theory, context effects

1 Introduction

We study a model of individual choice in which the decision maker holds multiple preferences simultaneously, each with different strengths described by a probability distribution. Faced with a subset of available alternatives, the preferences held by the individual can be in conflict. In our model, the potential conflict between these preferences in each choice situation is resolved by an aggregation rule. The model captures the idea that individuals may hold conflicting motivations with different strengths and that the weight each of these motivations receives may depend on the set of available alternatives. This view is consistent with a wide class of modern psychological theories of decision making.[1]

Formally, an aggregation rule is a map that assigns to each distribution of preferences λ a choice correspondence $v(\lambda)$. For any set of available alternatives A, this choice correspondence describes the decision $v(\lambda)(A) \subseteq A$ associated to the distribution λ. We consider a family of aggregation rules that satisfy an intuitive monotonicity property. Consider two decision-makers DM_1 and DM_2 who have different distributions of preferences but aggregate these preferences with the same rule. Suppose that DM_2 chooses an alternative a. Suppose that DM_1 has the same distribution of preferences as DM_2 except for the fact that some of the weight on π, one of the preferences held by DM_2, has been transferred to another preference π' that ranks a higher but preserves the order of all other alternatives. Monotonicity requires that DM_1 must also choose a.

The paper has two main results. Theorem 1 provides a representation theorem of aggregation rules that satisfy this monotonicity axiom and two standard axioms, *neutrality* and *continuity*. We show that each monotonic rule is represented by a choice-set-dependent utility function, $g_A(., \lambda)$, that, for each subset A, is equal to some monotonic function H_A of the cumulative distributions that keep track of the proportion of preferences that rank each alternative first, second, and so on, within A. These distributions are induced by the distribution of preferences λ and can be explicitly computed. In sum, any monotonic rule v can be identified with a family $\{H_A\}_A$ of monotonic functions and, vice versa, any collection of such functions produces a monotonic rule.

Using this representation, Theorem 2 shows that in the absence of domain

[1]See our related paper Green and Hojman [2007, 2014] for further motivation.

restrictions on the set of allowable preferences or additional constraints on the aggregation procedure, any generic monotonic aggregation rule can explain any behavior as we vary the distribution of underlying preferences. There are exceptional rules that do not reach all possible behaviors, but these rules are rare in a sense that we describe.[2]

Domain restrictions are natural in many settings. For instance, recent theoretical and empirical research in behavioral economics has considered dual-motivation models of behavior for example behaviors explainable by the interaction of two preferences -materialistic and altruistic, for example. This is a form of domain restriction on the population. For models like these we show that a much smaller class of behaviors can be rationalized. For example, when there are three alternatives we show that a dual-preference restriction of our model can explain menu effects such as the "compromise effect" but not cyclic patterns of behavior. We also illustrate how adding consistency restrictions on the aggregator across different choice sets reduces the set of admissible behaviors.

This paper contributes to a growing theoretical literature that explains irrational choice as the result of the interaction between multiple preferences or rationales.[3] Within this literature we distinguish three classes of models. The first class is models such as that in this paper where a set of preferences determines the choice and no distinction is made among the preferences. They interact as voters would in a system with an anonymous voting procedure.[4]

The second class of models are "non-strategic, multiple-objective mod-

[2]As discussed in the sequel, the space of monotonic aggregation rules is infinite dimensional. The notion of genericity used in this paper is *relative prevalence* introduced by Anderson and Zame [2001], a measure-theoretic generalization of "almost everywhere" (Lebesgue) finite-dimensional Euclidean spaces. We adapt the techniques used by these authors and Shannon [2006].

[3]Rationales may be incomplete binary relations that provide justifications or interact with the preferences in various ways, in order to produce a decision different from what the unconstrained preferences would have chosen.

[4]In addition to our earlier working paper Green-Hojman [2007], Ambrus-Rosen [2013] and De Clippel-Eliaz [2012] are in this category. Ambrus-Rozen allow the cardinal representations of preferences to play a role in the aggregation. Thus their model is more inclusive, and also more informationally demanding than ours. On the other hand, De Clippel-Eliaz restrict the number of explanatory preferences to two. This limits the behaviors that can be generated, but it offers more tightly determined welfare calculations as a result.

els". These models use multiple objectives but do not apply the multiple objectives symmetrically or anonymously.[5] A seminal paper in this mold is Kalai, Rubinstein and Spiegler [2002]. More recent papers include Manzini and Mariotti [2007], Rubinstein and Salant [2007], Apesteguía and Ballester [2013]. The third type of model are "strategic, multiple-objective models" that assume that the multiple objectives are players in a game. The outcome of the game generates the observed choice of the decision maker. Models of irrational choice in this category recognize the existence of multiple conflicting preferences and impose a specific strategic structure within which these preferences interact. They take the nature of the multiple selves to be exogenous rather than determining them endogenously from the choice behavior.[6]

While we emphasize applications to individual decision-making, our paper draws from and contributes to the social choice literature. The representation of monotonic rules we characterize (Theorem 1) significantly generalize Young's representation of scoring rules (Young, 1978). Scoring rules correspond to the subset of monotonic rules characterized by a choice-set dependent utility function that is linear in the distribution of preferences. Our result showing that, in the absence of domain restrictions on the set distributions, a typical monotonic rule spans all choice functions (Theorem 2) relies on measure-theoretic techniques introduced by Anderson and Zame [2001] and Shannon [2006]. It greatly generalizes Saari [1989, 2001] who showed that the full-spanning property holds for almost-every scoring rule.

[5]There are three types of papers in this category. One uses sequential procedures, or protocols, to resolve the conflict among the preferences. For example, the alternatives may be described by a list of attributes which could be considered in a fixed order to eliminate or reorder the alternatives. In psychology, classic studies in this mode are Tversky [1972], Shafir [1993], and Shafir, Simonson and Tversky [1993]. A second type of multiple objective theory partitions the decision problems into groups, within each of which only one objective is operational as in Kalai, Rubinstein, and Spiegler [2002] and Rubinstein-Salant [2007]. The third type of model uses a single objective function but multiple, context-dependent constraints as in Sen [1993].

[6]Important papers in the strategic mold include Strotz [1956], Schelling [1984], Gul-Pessendorfer [2001], and Fudenberg-Levine [2006].

2 Monotonic Aggregation Rules

We are interested in characterizing the choice outcomes produced by a family of rules that aggregates the multiple preferences held by a decision maker. We start with the basic notation and assumptions. There is a finite set of outcomes X. A subset $A \subseteq X$ is called a choice situation and the domain of choice situations is denoted by $\mathcal{A}$. Unless noted otherwise, we assume that $\mathcal{A}$ is the set of all non-empty subsets of X. The set of choice correspondences is denoted by $\mathcal{C}^*=\{c : \mathcal{A} \to \mathcal{A} |\ c(A) \subseteq A\}$.

The decision maker we study is characterized by a distribution of preferences over X and and aggregation rule. The set of strict preferences over X is denoted by Π and the set of distributions over this set is the simplex Δ^Π. In what follows, π denotes a generic preference in Π while λ and μ designate typical distributions of preferences in Δ^Π. The scalar $\lambda_\pi \in [0, 1]$ is the weight that λ assigns to π. If we think of π as a motivation, λ_π can be interpreted as the strength that λ gives to motivation π. Clearly, $\lambda_\pi \geq 0$ and $\sum_{\pi\in\Pi} \lambda_\pi = 1$.

An *aggregation rule* $v : \Delta^\Pi \to \mathcal{C}^*$ assigns a choice correspondence $v(\lambda) \in \mathcal{C}^*$ to each distribution of preferences λ. A decision maker is characterized by a pair (λ, v), consisting of a distribution of preferences and an aggregation rule. We use $v(\lambda)(A)$ for the choice associated to a decision maker characterized by (λ, v) when of available alternatives A. Our purpose is to provide a complete characterization of aggregation rules satisfying three axioms: continuity, monotonicity, and neutrality.

2.1 Axioms

The key axiom explored in this paper is the monotonicity axiom, which we introduce by means of an example.

Example 1 *Suppose that $X = \{x, y, z\}$. For short, we use "abc" to designate the preference $a \succ b \succ c$. Consider two distributions of preferences λ and μ that use the same aggregation rule v. The distribution λ puts weight $2/3$ on preference $\pi = xyz$ and $1/3$ on preference $\pi' = zyx$. The distribution μ puts weight $2/3$ on π, $1/6$ on π' and $1/6$ on $\pi'' = yzx$. The distribution μ replaces some of the voters in λ who have the preference π' by voters with preference π''. Note that the relative order between y and z is the same for*

π' and π'' but alternative y is "promoted" from second to first place. Monotonicity says that if y is chosen by λ then y should also be chosen by μ. Observe also that the relationship between λ and μ can be expressed by a linear vector equation

$$\begin{pmatrix} 2/3 \\ 1/6 \\ 1/6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 2/3 \\ 1/3 \\ 0 \end{pmatrix}.$$

The elements of stochastic matrix relating both distributions can be interpreted as mass transfer from one preference to another.

We introduce notation to express these ideas more formally. Given $a \in X$ and $\pi \in \Pi$, let $m(a, \pi) \subseteq \Pi$ be the set of preferences that preserve the ranking of π for alternatives in $X \backslash \{a\}$ but rank a better than π does. That is, $m(a, \pi) = \{\rho \in \Pi | x\pi y \Rightarrow x\rho y \forall x \in X, y \in X\backslash\{a\}\}$. Fix an arbitrary population $\lambda \in \Delta^\Pi$ and suppose that preference $\pi \in \Pi$ has positive measure under λ

Consider any transformation μ of λ such that the weight $\lambda(\pi)$ is redistributed across preferences in $m(a, \pi)$. We define any such μ *as a monotonic transformation of λ with respect to a*. We next define $M(a, \lambda) \in \Delta^\Pi$ to be the set distributions that can be obtained from λ by a sequence of monotonic transformations with respect to a. As illustrated by the example above, the set can be characterized by means of stochastic matrices describing the "transitions" from λ. Indeed, let $\mathcal{W}(a)$ denote the set of $\Pi \times \Pi$ stochastic matrices such that for each $W \in \mathcal{W}(a)$ we have that $W(\pi, \pi') \in [0, 1]$, $W(\pi, \pi') = 0$ unless $\pi' \in m(a, \pi)$, and $\sum_{\pi' \in \Pi} W(\pi, \pi') = 1$. The set of monotonic transformations of λ with respect to a is then

$$M(a, \lambda) = \left\{\mu \in \Delta^\Pi |\ \mu = W^T \lambda,\ W \in \mathcal{W}(a)\right\}.$$

Axiom (A1) (**Monotonicity**) *If $a \in v(\lambda)(A)$ then $a \in v(\mu)(A)$ for all $\mu \in M(a, \lambda)$.*

Our second axiom is neutrality. It captures the idea that the labeling of the alternatives does not affect the outcome.[7]

[7] In practice we may want to consider social or individual decision procedures that discriminate between different alternatives. This might be natural if there exists a status quo outcome. In section 5 we sketch how to extend the results of the paper relaxing the neutrality assumption.

Axiom (A2) (**Neutrality**) *Let $\sigma : X \to X$ be any permutation of alternatives and $q_\sigma : \Pi \to \Pi$ be the permutation induced by σ on orderings of X ($\sigma(a) q_\sigma(\pi) \sigma(b) \Leftrightarrow a\pi b$). If λ and $\widetilde{\lambda}$ are two distribution of voters such that $\widetilde{\lambda}_{q_\sigma(\pi)} = \lambda_\pi$ then $v(\widetilde{\lambda})(A) = v(\lambda)(\sigma(A))$.*

Neutrality places an intuitive restriction on how the aggregation rule depends on the distribution λ. Specifically, it can only depend on the rank of an alternative or more precisely on the weights induced by λ on preferences that rank the alternative in given position (among those available). Formally, given a subset A, let $|A|$ denote the cardinality of the set and $N_{|A|} = \{1, 2, ..., |A|\}$ be the list of the possible ranks or positions of elements in A. Let $\Delta^{N_{|A|}}$ be the simplex on $N_{|A|}$. For each preference $\pi \in \Pi$, let $rank(a, A, \pi)$ be the rank of alternative a within A under preference π. For each alternative $a \in A$, given the distribution of motivations $\lambda \in \Delta^\Pi$ we can calculate the share of the population that rank that alternative at a given position $r \in N_{|A|}$ across alternatives in A. Thus, for each alternative a, the distribution λ induces a distribution over the set of ranks $N_{|A|}$:

$$q^r_{aA}(\lambda) = \sum_{\pi : rank(a,A,\pi)=r} \lambda_\pi \tag{1}$$

is the mass of preferences that rank alternative a at r in the set A.

The real-valued linear functions $q^r_{aA}(\lambda)$ can be expressed as vectorial linear map $q_{aA} : \Delta^\Pi \to \Delta^{N_{|A|}}$ given by $q_{aA}(\lambda) = (q^1_{aA}(\lambda), ..., q^{|A|}_{aA}(\lambda))$. Neutrality implies that the choice $v(A, \lambda)$ depends on λ only through the vectors $\{q_{aA}(\lambda)\}_{a \in A}$. For later reference, the (rank-ordered) cumulative distribution associated with $q_{aA}(\lambda) \in \Delta^{N_{|A|}}$ is defined by

$$Q^r_{aA}(\lambda) = \sum_{j=1}^{r} q^j_{aA}(\lambda), \tag{2}$$

or, in vector notation, $Q_{aA}(\lambda) = (Q^1_{aA}(\lambda), ..., Q^{|A|-1}_{aA}(\lambda)) \in [0, 1]^{|A|-1}$. Since $Q^{|A|}_{aA}(\lambda) = 1$ for any a, A, and λ, we omit this component.

Our final axiom is a standard continuity requirement.

Axiom (A3) (**Continuity**) *The decision correspondence v is upper-hemicontinuous.*

The rules we study satisfy all three axioms:

Definition 1 *(Monotonic Aggregation Rules) A monotonic aggregation rule satisfies axioms (A1)-(A3). The set of monotonic aggregation rules is denoted by V^m.*

Below we provide a representation that allows for an operational definition of these rules. We start with an example and overview of the results.

2.2 Scoring rules and Overview of the Results

Scoring rules are a canonical example of aggregation rules that satisfy axioms (A1)-(A3). A scoring rule v is characterized by a set of $|X|-1$ scoring vectors $\{\gamma_k\}_{k\in\{2,...,|X|\}}$, where γ_k is the scoring vector that applies when the available set has k alternatives. We write $v = (\gamma_2, \gamma_3, ..., \gamma_{|X|})$ and the k-alternative scoring vector

$$\gamma_k = (\gamma_k^1, \gamma_k^2, ..., \gamma_k^k)$$

has k components satisfying $\gamma_k^1 \geq \gamma_k^2 \geq ... \geq \gamma_k^k$, and at least one these inequalities is strict. Without loss of generality, $\gamma_k^1 = 1$ and $\gamma_k^k = 0$ for all $k \in \{2, ..., |X|\}$. The scoring vector gives the number of "points" γ_k^r assigned to the r^{th} ranked alternative among the k alternatives in a subset A. Given a choice situation $A \subseteq X$, the *score* of alternative a from A under the ordering π is $\gamma_{|A|}^{\rho(a,A,\pi)}$. The *total score* of alternative a in choice situation A given a population λ is then

$$g_A(a, \lambda) = \sum_{\pi\in\Pi} \gamma_{|A|}^{\rho(a,A,\pi)} \lambda_\pi = \sum_{\pi:\rho(a,A,\pi)=r} \gamma_k^r q_{aA}^r(\lambda) = \gamma_k^T q_{aA}(\lambda). \tag{3}$$

Note that for each scoring vector γ_k we can construct the k-component vector of score differences $_k = (\delta_k^1, \delta_k^2, ..., \delta_k^k)$ satisfying

$$\delta_k^k = \gamma_k^k \text{ and } \delta_k^j = \gamma_k^j - \gamma_k^{j+1}$$

for $j \in \{1, ..., k-1\}$. Clearly $\delta_k^j \geq 0$ as $\gamma_k^j \geq \gamma_k^{j+1}$, for all j. In fact, for each k-component positive vector δ_k can construct a scoring vector in this manner. Recalling that he (rank-ordered) cumulative distribution associated to $q_{aA}(\lambda) \in \Delta^{N_k}$ is Q_{aA}, after a discrete version of integration by parts, we can express (3) as

$$g_A(a, \lambda) = \delta_k^T Q_{aA}(\lambda). \tag{4}$$

This means that for a given set, the score of an alternative is a linear increasing function of the cumulative distribution vector $Q_{aA}(\lambda) = (Q_{aA}^1(\lambda), ..., Q_{aA}^k(\lambda)) \in [0,1]^k$. In particular, it follows that if the distribution $Q_{aA}(\lambda)$ first order stochastically dominates (FOSD) $Q_{bA}(\lambda)$ then $g_A(a,\lambda) \geq g_A(b,\lambda)$ for any scoring rule. In the section 3 we show that for any monotonic rule we can find a choice-set dependent utility function g_A that assigns a "generalized score" to each of the alternatives and such that the choice is the set of maximizers of g_A. It is shown that the tight connection between the stochastic dominance and the scores of alternatives is preserved by this function. Section 4 generalizes a result obtained by Saari for scoring rules: A typical monotonic rule, can induce any choice function as the population varies.

3 Representation Theorem

In this section we provide a general representation of monotonic aggregation rules. We start with some definitions:

Definition 2 *The choice-set dependent utility function $g_A : A \times \Delta^{\Pi} \to \mathbb{R}$ is said to represent a aggregation rule v at $A \in \mathcal{A}$ if for each $\lambda \in \Delta^{\Pi}$ we have that*

$$v(\lambda)(A) = \arg\max_{a \in A} \; g_A(a,\lambda)$$

If for every $A \in \mathcal{A}$ g_A represents v at A we write $v = \{g_A\}_{A \in \mathcal{A}}$.

In classical decision theory a representation associates a utility function to a preference and the form of the utility function relates characteristics that the preference displays. These characteristics may induce certain qualitative properties in the choices that are observed. In the present context the analog of "the preference" is a distribution over orderings. Choices are allowed to be inconsistent as the set of available alternatives varies. Therefore it is natural that our utility functions allow for choice set dependence. A representation in our context is a set of utility functions, one for each available set. We show that these utility functions, $g_A(a,\lambda)$, have a property that reflects the monotonicity assumption on the associated choice behavior.

Theorem 1 *Let $Q_{aA} : \Delta^{\Pi} \to [0,1]^{|A|-1}$ be the cumulative distribution vector map. An aggregation rule v satisfies axioms (A1)-(A3) if and only if there v for each $A \in \mathcal{A}$ there exists a increasing function $H_A : [0,1]^{|A|-1} \to \mathbb{R}$ such that $g_A(a,\lambda) = (H_A \circ Q_{aA})(\lambda)$ represents v at A.*

Theorem 1 is useful to derive important properties of the choice-set dependent scoring maps g_A. We later use these properties later to show that most aggregation rules satisfying our axioms can generate arbitrary choices as we vary the distribution of preferences. Observe that for scoring rules, H_A is linear: $H_A(Q) = \sum_{r=1}^{|A|} \delta_A^r Q^r$, where $\delta_A^r \geq 0$ with strict inequality for some r.

Before providing a proof, we state the following immediate corollary that summarizes intuitive properties of the representation.

Corollary 1 *If v satisfies axioms (A1)-(A3) then v can be represented by collection $\{g_A\}_{A \in \mathcal{A}}$ of choice-set dependent utility functions such that*
(i) $g_A(a, \cdot)$ is monotonic, Lipschitz-continuous, and differentiable almost-everywhere with respect to λ;
(ii) $g_A(a, \mu) \geq g_A(a, \lambda)$ if $Q_{aA}(\mu)$ FOSD $Q_{aA}(\lambda)$.

3.1 Proof of the Theorem

Theorem 1 is established with a lemmata that shows the tight connection between the Monotonicity axiom and FOSD. Specifically, as argued previously, from neutrality we know that the dependence of the choice-set-dependent utility function that represents a monotonic rule can be expressed as a function of the rank cumulative distributions $Q_{aA}(\lambda)$ defined earlier for each $a \in A$, $A \in \mathcal{A}$ and $\lambda \in \Delta^{\Pi}$. We show that one can identify the set choice-set-dependent utility functions that represent a monotonic aggregation rule with the functions that respect FOSD with respect to these distributions. To make this precise, we introduce the following definition.

Definition 3 *For $a \in X$ and $\lambda \in \Delta^{\Pi}$ the distribution $\mu \in \Delta^{\Pi}$ is said to a-FOSD λ if for any set of alternatives $A \subseteq X$ with $a \in A$ we have that*
(i) $Q_{aA}(\mu)$ FOSD $Q_{aA}(\lambda)$ and
(ii) $Q_{bA}(\lambda)$ FOSD $Q_{bA}(\mu)$ for $b \in A \setminus \{a\}$.
The set of distributions that a-FOSD λ is denoted by $FOSD(a, \lambda)$.

Condition (i) says that at any choice set $A \in A$ the mass of preferences that rank a better than any given rank $k \leq |A|$ is greater at μ than it is at λ. Condition (ii) says that all alternatives other than a have lower cumulative ranks. The idea of this definition is that while λ and μ may be quite different, the improvement of the status of a within A is unambiguous. Because no other alternative improves any of its rankings it is natural that once a is chosen at λ it remains chosen at μ.

The following Lemma establishes a connection between of the set of monotonic transformations of λ with respect to a, $M(a,\lambda)$, introduced to define the monotonicity axiom and the set of distributions that a-FOSD λ, $FOSD(a,\lambda)$. Let $Q_A : \Delta^\Pi \to CD_A$ be the map that assigns to each distribution of preferences $\mu \in \Delta^\Pi$ the "stack" of cumulative distribution vectors $Q_{bA}(\mu) \in [0,1]^{|A|-1}$ for a fixed A. Here, $CD_A = [0,1]^{(|A|-1)|A|}$. Consider $\{A_1, A_2, ..., A_S\}$, the collection of subsets that contain alternative a. Let $\widetilde{Q} : \Delta^\Pi \to \times_{s=1}^{S} CD_{A_s}$ be the map that stacks all the $Q_A(\mu)$'s for subsets in this collection.

Lemma 1 *Fix $a \in X$ and $\lambda \in \Delta^\Pi$. Then $\widetilde{Q}(FOSD(a,\lambda)) = \widetilde{Q}(M(a,\lambda))$.*

The proof is in the Appendix, where we show that $M(a,\lambda) \subset FOSD(a,\lambda)$.[8] However, it turns out that both sets span the same cumulative distribution maps.

Let Q and Q' be cummulative distribution vectors. That is, for some integer R we have that $Q, Q' \in [0,1]^R$ and each of these vectors has increasing components. Observe that Q FOSD Q' if and only if $Q > Q'$ where $>$ is a vector inequality indicating that all of the components of Q are greater than or equal to the respective component of Q'.

Definition 4 *Let $Q, Q' \in [0,1]^R$ be cummulative distribution vectors, for some $R \geq 1$. A real-valued function $H : [0,1]^R \to \mathbb{R}$ satisfies the FOSD property if Q FOSD $Q' \Rightarrow H(Q) \geq H(Q')$.*

We have the following Lemma:

[8] The transformations allowed by the monotonicity axiom and captured by $M(a,\lambda)$ are quite restrictive as they require a mass shift from from preferences that promote alternative a but do not alter the relative order of any other alternative. In contrast, the dominance conditions that define $FOSD(a,\lambda)$ allow for transformations that "on average" shift mass to preferences that rank a better without promoting any other alternative in sets that contain a.

Lemma 2 *Let $v \in V^m$. Any representation $\{g_A\}_{A\in\mathcal{A}}$ of v is such that:*

(i) For each $A \in \mathcal{A}$, there exists a continuous function $H_A : [0,1]^{|A|-1} \to \mathbb{R}$ such that $g_A(a, \lambda) = H_A(Q_{aA}(\lambda))$ for all $a \in A$ and $\lambda \in \Delta^\Pi$;

(ii) For each $A \in \mathcal{A}$ the function H_A satisfies the FOSD property.

Part (i) follows immediately from the neutrality and continuity axioms, the proof is omitted. Part (ii) follows directly from (i) and lemma 1. Indeed, from (i) the representation on the distribution of preferences is restricted to functions of the cumulative distributions $Q()$. Given this, from lemma 1, since the distributions that satisfy the monotonicity test span the same cumulative distributions than those satisfying the FOSD property, the representation must satisfy the FOSD property.

The previous Lemma identifies a representation with collections of functions H_{AA} that satisfy the FOSD property.

4 Almost Every Monotonic Rule Spans All Choice Functions

Recall that $\mathcal{C}^*$ denotes the set of choice correspondences. A choice function $c \in \mathcal{C}^*$ is such that for each $A \in \mathcal{A}$, the choice $c(A)$ is a singleton. The set of choice functions is denoted by $\mathcal{C}$. From above an aggregation rule $v = \{g_A\}_{A\in\mathcal{A}}$ and a population λ induce a choice correspondence $v(\lambda) \in \mathcal{C}^*$, where $v(\lambda)(A) = \arg\max_{a\in A} g_A(a, \lambda)$ for each $A \in \mathcal{A}$.

Definition 5 *(Explanation/Rationalization) The aggregation rule v is said to explain a choice function $c \in \mathcal{C}$ if there exists a full Lebesgue measure set $\Lambda \subseteq \Delta^\Pi$ such that for any $\lambda \in \Lambda$ we have that $c = v(\lambda)$, where $v : \Delta^\Pi \to \mathcal{C}$ is the decision correspondence induced by v. The aggregation rule is said to span all choice functions if $v(\Delta^\Pi) = \mathcal{C}$.*

Our purpose is to characterize the set of choice functions spanned by a generic monotonic rule. As shown in the previous section, any monotonic aggregation rule can be identified with a choice-set dependent utility function that determined by a collection monotonic transformations, one for each choice set. The space of monotonic transformations is an infinite-dimensional space. In this paper, we use a notion of "genericity" for infinite-dimensional metric spaces called *prevalence*, which is based on a measure theoretic notion. Prevalence extends the idea of "Lebesgue almost everywhere" used in finite dimensional spaces. It was introduced independently by Christensen [1974] and Hunt, Sauer and Yorke [1992]. In this paper we use the extension of this concept introduced by Anderson and Zame [2001] -*relative prevalence*.[9]

For exposition, we provide a brief intuitive definition and focus on the main intuitions that underlie our results. A detailed definition can be found in Appendix A2. A Borel subset N of the function space M is called *shy* if there exists a probability measure ν on M for which the measure of every translate of N is zero. That is, N is shy if $\nu(N+t)=0$ for every $t\in M$. A prevalent set is a set whose complement is shy. In particular, prevalent implies dense.[10] Like Lebesgue measure 0, (relative) shyness is translation invariant, preserved under countable unions, coincides with Lebesgue measure 0 in $\mathbb{R}^K$, and no relatively open set is relatively shy. In particular, every relatively prevalent set is dense. As used normally with the Lebesgue measure in a Euclidean space, we use "almost every" to identify a prevalent set in the infinite-dimensional space.

This is the main theorem of the paper:

Theorem 2 *For almost every aggregation rule $v\in V^m$ we have that $v(\Delta^{\Pi})=\mathcal{C}$. That is, almost every monotonic voting rule spans all choice functions.*

We differ a discussion of the behavioral interpretation of the theorem for the next section and start by focusing on the technical contribution. Our proof makes use of the representation of monotonic rules derived in Theorem 1 and, illustrates the precise conditions required for a family of aggregators to span all choice functions.[11] We introduce some notation to illustrate the

[9] Anderson and Zame's extension is substantial as in most economic models the primitives -preferences, endowments, technology- are naturally described as a subset of a convex topological space rather than a vector space as assumed in Hunt, Sauer, and Yorke [1992].

[10] There are two different type of "genericity" notions for infinite-dimensional metric spaces. Aside from the measure-theoretic approach considered herein, the other notion is topological and it is based on categories of sets. A set is said to be *residual* or generic if it contains a countable intersection of dense sets. The complement of such set is said to be of the *first-category, i.e.,* it is a countable union of nowhere dense sets.

[11] The result generalizes Saari [1989, 2001] who showed that almost any scoring rule spans all choice functions. Saari's proof exploits the linearity of scoring rules.

main argument.

Consider the representation $v = \{g_A\}_{A\in\mathcal{A}}$. For each $A \in \mathcal{A}$, $a, b \in A$, and $\lambda \in \Delta^\Pi$ let

$$\phi^v_A(a, b, \lambda) \equiv g_A(a, \lambda) - g_A(b, \lambda).$$

This is the difference between the score of alternative a and b at A. Consider an arbitrary enumeration of the choice sets in the domain $\mathcal{A}$ so that $\mathcal{A} = \{A_1, ..., A_K\}$, where $K = |\mathcal{A}|$. For each choice set A_j consider an arbitrary enumeration of its alternatives so that $A_j = \{a_{j1}, ..., a_{j|A_j|}\}$. For each $j \in \{1, ..., K\}$, let $\phi^v_j : \Delta^\Pi \to \mathbb{R}^{|A_j|-1}$ be the vector map with components $\phi^v_{jk}(\lambda) = \phi^v_{A_j}(a_{jk}, a_{jk+1}, \lambda)$ for $k \in \{1, ..., |A_j| - 1\}$. The map ϕ^v_j summarizes the differences in score between pairs of alternatives in A_j.

For short, let $n \equiv |X|$ be the total number of alternatives and $S_n \equiv n2^{n-1}$. Observe that $S_n = \sum_{j=1}^{K}(|A_j| - 1)$, that is, it is the total number of the coordinates of all of the maps ϕ^v_j. [12] For a fixed v, the *excess score* map $\phi^v : \Delta^\Pi \to \mathbb{R}^{S_n}$ is defined by $\phi^v(\lambda) = (\phi^v_1(\lambda), ..., \phi^v_K(\lambda))^T \in \mathbb{R}^{S_n}$. Using, theorem 1 it follows that each component of the excess score has the form $H^v_j \circ Q_{a_k,A_j} - H^v_j \circ Q_{a_{k+1},A_j}$, where H^v_j is an increasing function and $Q_{a,A}$ is a linear cumulative distribution map. We exploit this structure to derive the regularity properties of an excess score function for a typical rule $v \in V^m$.

Definition 6 *Let v be an aggregation rule with excess score map $\phi^v : \Delta^\Pi \to \mathbb{R}^{S_n}$. For a fixed $\lambda_0 \in \Delta^\Pi$ let $\varphi^v(\cdot, \lambda_0) : \Delta^\Pi \to \mathbb{R}^{S_n}$ be the map defined by $\varphi^v(\lambda) = \phi^v(\lambda) - \phi^v(\lambda_0)$. The rule v is said to be regular at $\lambda_0 \in \Delta^\Pi$ if there exists $\bar{\epsilon}$ such that for any $\epsilon < \bar{\epsilon}$ and any $\epsilon-$neighborhood R_0 of $0 \in \mathbb{R}^{S_n}$, there exists of open neighborhood $\Lambda_0 \subseteq \Delta^\Pi$ such that $\varphi^{-1}(R_0) \subseteq \Lambda_0$.*

In words, in a neighborhood of point of regularity λ_0, we can always find a point λ such that each component of $\phi^v(\lambda) - \phi^v(\lambda_0)$ achieves any predetermined sign. This means that ϕ^v is locally surjective.

Definition 7 *A distribution of preferences $\widehat{\lambda} \in \Delta^\Pi$ is neutral for aggregation rule v if $v(\widehat{\lambda})(A) = A$ for all $A \in \mathcal{A}$.*

[12] It's easy to check that S_n is strictly less than the dimensionality of the simplex Δ^Π, which is $n! - 1$.

Note that, given a representation $v = \{g_A\}_{A \in \mathcal{A}}$, $\widehat{\lambda}$ is said to be neutral if $g_A(a, \widehat{\lambda}) = g_A(b, \widehat{\lambda})$ for all $a, b \in A$, $A \in \mathcal{A}$. In terms of the excess score function, the latter is equivalent to $\phi^v(\widehat{\lambda}) = 0$.

The following Lemma shows how regularity at a point together with neutrality at that point implies full span.

Lemma 3 *An aggregation rule $v \in V^m$ spans all choice functions if there exists a $\widehat{\lambda} \in int(\Delta^{\Pi})$ such that (i) $\widehat{\lambda}$ is neutral for v and (ii) v is regular at $\widehat{\lambda}$.*

We provide the argument, a formal proof is omitted. The existence of an interior distribution $\widehat{\lambda}$ neutral for v implies that $\phi^v(\widehat{\lambda}) = 0$. If, in addition, v is regular at that point, we can always find λ such that the components of $\phi^v(\lambda) - \phi^v(\widehat{\lambda}) = \phi^v(\lambda)$ have any predetermined signs. That is, by regularity, we can always find a λ such that the ties in the scores in any choice set A are broken in any way we want. In particular, for any given choice function c, we can always find an open subset U such that if $\lambda \in U$ then $\phi_A^v(c(A), b, \lambda) > 0$ for all $b \neq c(A)$, or equivalently, $c(A)$ maximizes $g_A(a, \lambda)$ for each A.

From Lemma 3, to establish theorem 2 it is sufficient to show that for a typical v there exists a point $\widehat{\lambda}$ such that the $\widehat{\lambda}$ is neutral for v and the rule is regular at that point. Let $e_{\Pi} \in \Delta^{\Pi}$ be the distribution that puts equal weight on each ordering of Π.

Lemma 4 *If v satisfies the neutrality axiom (A3) then $\widehat{\lambda} = e_{\Pi}$ is neutral with respect to v.*

The result is an immediate consequence of the neutrality axiom, the proof is straightforward and omitted.

Lemma 5 *Almost every $v \in V^m$ is regular at $\widehat{\lambda} = e_{\Pi}$.*

The proof of Lemma 5 is in the Appendix. We provide the main argument. Let C^m denote the set of excess score functions associated with some monotonic rule $v \in V^m$. The result is established by showing that almost every $\phi \in C^m$ is regular at a point where it has a zero ($\phi(\widehat{\lambda}) = 0$). If $\phi \in C^m$ is differentiable, the result obtains from a version of Sard's Theorem. Indeed, for a smooth ϕ a sufficient condition for regularity at λ_0 is for the Jacobian $D\phi^v(\lambda_0)$ to have full rank at that point (or equivalently, $D\phi^v(\lambda_0)$ is surjective), which follows from Sard's theorem. Intuitively, the set of points such that one of the components, say ϕ^v_{jk}, of the excess score map is zero defines a surface $U_{jk} \subset \Delta^{\Pi}$. For example, in the case of linear scoring rules, this surface is an hyperplane. The gradient $\nabla\phi^v_{jk}(\lambda)$ at any point $\lambda \in U_{jk}$ defines a direction of increase of that particular component. A zero of ϕ^v is a point in the intersection of all of the surfaces defined by the zeros of each individual component. The full-rank condition implies that we can always find an orthogonal collection of vectors that have a positive projection on each of the component-gradients. That is, we can always find a point in the neighborhood of a zero such that an increase or decrease of a component of the scoring vector does not affect the value of another one.

However, our representation theorem does not imply smoothness and it is easy to come up with sensible monotonic rules that are not smooth. By Corollary 1, since the choice-set dependent utilities that represent a rule v are Lipschitz-continuous in λ, we have that so is the excess score ϕ^v . For this class of functions a generalized Jacobian is available. In effect our proof in the Appendix establishes a version of Sard's theorem for a subset of Lipschitz continuous maps. The argument has two main steps. We first show that the subset R^m of C^m corresponding to rules that are regular at the "zero" points of the associated excess function is a Borel set of the ambient space, which in our case is a set of Lipschitz continuous functions. Next we find a finite-dimensional subset of rules N_s that satisfies the property for the Lebesgue measure. We use the set of scoring rules, which is defined by a collection of scoring vectors (a subset of a Euclidean space) as seen in section 2.1. We know that for these rules the theorem holds [Saari, 2000]. The argument is established by showing that any projection of the non-regular scoring functions $C^m \backslash R^m$ onto N_s is a set of Lebsegue measure zero on the finite-dimensional space.

5 Domain Restrictions and Extensions

Theorem 2 says that any behavior, no matter how irrational, can be explained using a model of conflicting motivations aggregated with a typical monotonic rule. This means that the model of choice presented herein cannot be rejected using choice data alone. There are two important classes of comments that should be noted at this point. First, the range of experimentally observed choice patterns is indeed very wide. In the case of a three alternatives, there are four logically possible choice functions -modulo a relabeling of the alternatives. Indeed, choices from pairs of alternatives can respect transitivity or exhibit cycles. All choice functions with a cycle are the equivalent modulo a permutation of the alternatives. Alternatively, if choices from pairs respect transitivity, they will be consistent with unique order. For example, a choice function such that

$$c(\{x, y\}) = x,\ c(y, z\}) = y, \text{ and } c(\{x, z\}) = z$$

is consistent with the ranking $\pi_1 = xyz$. In this case, there are three possible choice functions depending on the choice from the triple. If $c(\{x, y, z\}) = x$ then the choice is *seemingly rational.* If $c(\{x, y, z\}) = y$ we say that c exhibits *second-place choice.* The compromise effect illustrated by our example on other-regarding behavior is an example of this type of behavior. Finally, if $c(\{x, y, z\}) = z$ we say that c exhibits *third-place choice.* Both experimental and field-based studies have displayed all four of the logically possible choice functions -seemingly rational, cyclic, second and third-place choice.[13] Few systematic studies with four or more alternatives have been made. Thus far, there is no reason to believe that any particular family of choice functions will never be seen in some data set. Therefore, the fact that our model can accommodate choice functions of arbitrary complexity may well be a strength rather than a weakness.

A second comment on this "full range" theorem relates to the support of the set of explanatory populations. Introspection and psychological research make "internal conflict" an appealing idea. The theorem above places no restriction on the extent of this conflict. However as we discuss below, specific limitations on the nature of a decision maker's internal conflict will place

[13] The three patterns inconsistent with rational behavior have been documented by the experimental psychology and decision-making literature that focuses on context effects in choice with multi-attribute alternatives. The classic paper by Tversky [1969] and more recent work by Roelofsma and Read [2000] show that cyclic choice can arise systematically. There is also robust evidence of Second Place Choice, as shown by Simonson [1989]. Third place choice seems to be more elusive but Redelmeier and Shafir [1995] finds this pattern. The prevailing psychology theories include sequential decision-making procedures such as *elimination by aspects* or theories based on context-dependent salience such as *asymmetric dominance.* A more comprehensive theory called *reason-based choice* is proposed by Shafir, Simonson, and Tversky [1993]. This theory, based on the idea that the context determines which among of many conflicting reasons prevails in a given choice situation, is close in spirit to the model presented in this paper.

restrictions on the family of choice functions that can be generated and are testable using choice data alone.

5.1 Domain Restrictions

5.1.1 Dual-System Explanations

Recent research in behavioral and neuroeconomics has emphasized that departures from rationality could result from the conflict between two or more "systems" or motivations. For example, in several domains of decision-making (e.g. choice under uncertainty, intertemporal choice) non-rational choice is often attributed to the interaction between *affective* and *rational* brain systems.[14] The structure of internal conflict need not be based on biological systems, as illustrated by one of our motivating examples, in which choice results from the conflict between selfish and other-regarding motives. In our framework a "dual system" corresponds to a domain restriction on explanatory populations. Dual-system explanations can be identified with set $\Delta^{dual} = \{\lambda \in \Delta | \ \lambda_\pi + \lambda_{\pi'} = 1 \text{ for some } \pi, \pi' \in \Pi\}$. These are distributions that put weight on at most two preferences, one for each "system".

Proposition 1 *If $|X| = 3$ then dual-system explanations based on monotonic rules can only explain choice functions that are either seemingly rational or satisfy second-place choice (compromise effect). Cyclic choice and third-place choice cannot be explained by dual-system explanations based on monotonic rules.*

Thus, if the analyst incorporates specific assumptions about the nature of the conflict in a specific decision context, the theory yields sharper predictions. In particular, in the case of three alternatives, cyclic choice and third-place choice can only be generated by distributions that give positive weight to three or more motivations. There is a sense in which these behaviors are more "irrational" than seemingly rational and second-place choice as they can only be rationalized with "more conflicted" explanations. The following example is from our related paper Green and Hojman [2014].

Example 2 *Selfish and Other-regarding Motives in Conflict*

[14]See for example, Kanheman [2003] and Camerer-Loewenstein-Prelec [2005].

The clash between a "selfish motivation" that aims to maximize material well-being and motivations grounded on social norms of reciprocity is pervasive in a number of social dilemmas.[15] In this example there are three outcomes $x = (1, 0)$, $y = (\frac{3}{4}, \frac{1}{4})$ and $z = (\frac{1}{2}, \frac{1}{2})$ each representing the split of one dollar. Each alternative is $a = (a_1, a_2)$ where a_1 is what person 1 gets and a_2 is what person 2 gets, $a_1 + a_2 = 1$. Individual 1 chooses the outcome as in a dictator game. We focus on this individual's preferences and behavior.

A commonly observed choice behavior when outcomes have multiple attributes is the "compromise effect". In this example, consider the following choice pattern

$$\begin{aligned} c(\{x, y\}) &= x \\ c(\{x, z\}) &= x \\ c(\{y, z\}) &= y \\ c(\{x, y, z\}) &= y. \end{aligned}$$

Hence, when confronted with any pair of outcomes the dictator chooses the alternative that gives him or her the highest share. However, when all three alternatives are available, the decision maker chooses the "compromise" alternative. In line with Shafir, Simonsen, and Tversky [1993], this choice pattern can be explained as follows: when the two extreme outcomes x and z are available, the conflict between selfish and other-regarding motives becomes more salient and alternative y provides a "compromise" between these conflicting reasons. Without adhering to a specific "story", our model of choice is consistent with this view. If we restrict the domain of explanations to distributions of preferences over the "selfish motivation" $\pi_s = xyz$ and the "other-regarding motivation" $\pi_o = zyx$, there are two are four choice behaviors that can be explained using linear aggregation rules: seemingly rational choice consistent with π_s, "seemingly rational choice" consistent with π_o, and two instances of "second-place choice". Both second-place choice behaviors have the "compromise alternative" y chosen from the triple and in one case choices from the pairs are consistent with π_s while in the other they are consistent with π_o.

[15] The extensive literature on other-regarding behavior has focused on rationalizations based on "social preferences". See Sobel [2005] for a survey. The pattern of behavior in this example cannot be explained appealing to any of the models as it is inconsistent with the maximization of a preference.

5.1.2 "More is Better" and Natural Explanations

We have considered a general finite space of alternatives X. In many economic applications X has a structure that imposes natural restrictions on the set of explanatory preferences. For example, in choosing between different cars, the decision-maker may consider the miles per gallon of gas, safety indicators, and color of each option. If mileage per gallon and safety are represented as numbers, each of the attributes admits a natural order (color perhaps not). In the case of consumption bundles, the amount of each commodity is similarly ordered in the usual "more is better" fashion. In each of these examples, an alternative can be described by a $m-$component vector $x = (x_1, ..., x_m)$ of m attributes or commodities, so that $X = \times_{i=}^{m} X_i$, and there exists a (maximal) subset of attributes $O \subseteq \{1, ..., m\}$ such that for each $i \in O$, the space X_i has a linear order $>_i$. These orders induce a partial order $\succ$ on the space X: $x \succ y$ if $x_i >_i y_i$ for all $i \in O$ and $x_i = y_i$ for all $i \notin O$. The partial order $\succ$ has the usual interpretation of "more is better". Only preferences in the set $\Pi^O = \{\pi \in \Pi | x \succ y \Rightarrow x\pi y\}$ rank alternatives consistent with the order $\succ$, and it is natural to seek explanations that put weight exclusively on preferences in Π^O. If the domain of distributions is restricted to $\Delta^O = \{\lambda \in \Delta^\Pi | \lambda_\pi > 0 \Rightarrow \pi \in \Pi^O\}$, it can be shown that using a typical monotonic rule a choice function c can be explained only if it is consistent with the partial order $\succ$ on X. Thus, a choice function such that, for some A, $c(A)$ is dominated under $\succ$ by some other alternative in A cannot be explained.

5.1.3 Consistent Aggregators

Restricting the set of aggregators can also lead to sharper predictions. Property (M3) places no constraint on the function H_A across subsets. For illustration, suppose that $X = \{x, y, z, w\}$ and let $A = \{x, y, z\}$. Suppose that $g_A(a, \lambda) = Q^1_{aA}(\lambda)$, i.e., the score of an alternative is simply the weight of those preferences that rank it first. Instead, if X is available suppose that $g_X(a, \lambda) = Q^3_{aX}(\lambda)$, i.e., the scoring function gives equal importance to preferences that do not rank the alternative last out the four available. Unless we want to allow the aggregation of preferences to vary arbitrarily across choice situations, it might be reasonable to impose some consistency on the aggregator.

An example of "consistent aggregator" is the Borda rule, which is defined by scoring functions of the form $g_A(a,\lambda) = \frac{1}{|A|}\sum_j Q^j_{aA}(\lambda)$. It can be shown that the score of an alternative a at set A is the average of the score this alternative gets in each subset of A of size $|A|-1$ and thus, by recursion, it is an average of the score a obtains in pairwise contests against each of the other alternatives in the choice set. The tight connection between scores across different choice situations restricts the choice functions that can be explained using this type of aggregator. In particular, if $|X| = 3$ third-place choice cannot be explained by using Borda aggregators. More generally, consider the following definition:

Definition 8 *(Weak Condorcet Consistency) A choice function $c \in \mathcal{C}$ satisfies Weak Condorcet Consistency (WCC) there is no choice set A of three or more alternatives such that for some pair of alternatives $x_w, x_l \in A$ (i) x_l is never chosen from a pair that contains any other alternative in a ($x_l \notin c(\{x_l, a\})$, $a \in A\backslash\{x_l\}$), (ii) x_w is always chosen from a pair that contains any other alternative in a ($x_w = c(\{x_l, a\})$, $a \in A\backslash\{x_w\}$), and (iii) x_l is chosen from A ($x_l = c(A)$).*

For each set A, and each alternative $a \in A$, let $\widehat{n}_{aA} : \Delta^\Pi \to [0,1]^{|A|-1}$ be the map that assigns each distribution λ with the vector $\widehat{n}_{aA}(\lambda) = (Q^1_{a,\{a,b\}}(\lambda))_{b\in A\backslash\{a\}}$, i.e., the vector that collects the weight $Q^1_{a,\{a,b\}}(\lambda)$ on preferences that prefer a to b when the pair $\{a,b\}$ is available.

Proposition 2 *Suppose that $v = \{g_A\}_A$. If there exists an increasing and symmetric function $H_A : [0,1]^{|A|-1} \to \mathbb{R}$ such that $g_A(a,\lambda) = H_A(\widehat{n}_A(a,\lambda))$ for each $A \in \mathcal{A}$, $a \in A$ then then choice rule induced by v satisfies WCC. That is, $v(\lambda)$ satisfies WCC for any λ.*

The result follows immediately as, if H_A is symmetric and increasing, for any A and $x_w, x_l \in A$ that satisfy (i) and (ii) we have that $g_A(x_w,\lambda) > g_A(x_l,\lambda)$.

5.2 Non-Neutral Rules and Status Quo Bias

There are multiple examples in the psychology literature that emphasize the salience of certain alternatives over others. A canonical example is the existence of a status quo or default alternative. Extending our results to allow for non-neutral rules is non-trivial. However, it is possible to consider simple extensions of the framework that allow for certain types of non-neutrality. The following example provides an illustration. In this example, the decision maker is characterized by a distribution of preferences λ as before, and a choice-set dependent utility function with two components:

$$g_A^{\theta}(a,\lambda) = \widehat{g}_A(a,\lambda) + \theta h_A(a,\lambda).$$

The first part $\widehat{g}_A$ is a "neutral component" satisfying the same properties as the representations characterized earlier. The second part, is a "non-neutral component" scaled by a constant $\theta \in [0,\infty)$ and $h_A(a,\lambda) \geq 0$ if $a = a_0$ and zero for any other alternative. Alternative a_0 plays the role of a status quo receiving an additional "default" utility as in Rubinstein and Salant [2007]. This utility could depend on λ and A, indicating that the relevance of the status quo may depend on how conflicted the preferences are with respect to alternatives in A. For example, if λ is a degenerate distribution that puts weight on a single preference, the status quo could be irrelevant. In contrast, the status quo could become more salient if λ has a larger support.

From a purely formal perspective, note that the example allows us to illustrate the role of the neutrality axiom in Theorem 2. For simplicity suppose that $h_A(a,\lambda)$ is independent of λ and A, $h_A(a_0,\lambda) = h_0 > 0$. Note that for $\theta = 0$, the rule is neutral and the distribution that gives equal weight to all preferences, e_Π, is neutral for the rule. By continuity, for small values of θ -i.e. a small departure from neutrality, there exists λ_θ close to e_Π that is neutral for the rule defined by g_A^{θ}. Thus, Theorem 2 will remain valid. In contrast, for large θ this is no longer true. In particular, for θ large enough the status quo a_0 is always chosen whenever available. This holds for any λ. Thus, an arbitrary choice function is no longer rationalizable by varying the distribution of preferences.

6 Conclusion

We provide a choice-set dependent utility representation for a decision maker that aggregates multiple preferences with a monotonic rule. Using this representation we show that, in the absence of domain restrictions on the set of allowable preferences or additional restrictions on the aggregation procedure, a typical monotonic aggregation rule can explain any behavior as we vary the distribution of underlying preferences. That is, given any choice observations, for almost every aggregation rule one can find a distribution of preferences that rationalizes that pattern of choice. Domain restrictions on the set of preferences (e.g. dual motivation models) or consistency restrictions on the aggregator reduce the set of admissible behaviors. Applications to positive models of individual decision making with context effects and social choice are discussed.

References

[1] Anderson, Robert M., Zame, William R. 2001. “Genericity with Infinitely Many Parameters. ” *Advances in Theoretical Economics* 1(1).

[2] Ambrus, Attila, and Kareen Rozen. 2013. "Rationalising Choice with Multi-Self Models", The Economic Journal.

[3] Apesteguia, J., Ballester, M. A. (2013). “Choice by sequential procedures.”*Games and Economic Behavior* 77(1), 90-99.

[4] Camerer, Colin, George Loewenstein, and Drazen Prelec.2005. “Neuroeconomics: How neuroscience can inform economics”, *Journal of Economic Literature*, 43: 5-60.

[5] Chambers, Christopher and Takashi Hayashi. 2012. “Choice and IndividualWwelfare” *Journal of Economic Theory* 147(5), 1818-1849.

[6] hristensen, J. P. 1974. *Topology and Borel Structure,* vol. 10 in North Holland Mathematics Studies.

[7] De Clippel, G., Eliaz, Kfir, 2012. Reason-based choice: “A bargaining rationale for the attraction and compromise effects, ” *Theoretical Economics* 7(1): 125-162.

[8] Fudenberg, Drew and David Levine. 2006. "A Dual Self Model of Impulse Control", *American Economic Review* 96:1449-1476.

[9] Green, Jerry R. and Daniel A. Hojman. 2007. "Choice, Rationality and Welfare Measurement," *Harvard Institute of Economic Research Discussion Paper* # 2144.

[10] Green, Jerry R. and Daniel A. Hojman. 2008. “Monotonic Voting Rules,” mimeo.

[11] Gul, Faruk and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control" *Econometrica* 69: 1403–1435.

[12] Hunt, B. R., Sauer, T., Yorke, J. A. (1992). “Prevalence: a translation-invariant ’almost every’ on infinite-dimensional spaces.” *Bulletin of the American Mathematical Society* 27(2): 217-238.

[13] Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review* 93(5): 1449-1475.

[14] Kalai, Gil, Ariel Rubinstein, and Ran Spiegler. 2002. "Rationalizing Choice Functions by Multiple Rationales," *Econometrica* 70(6): 2481-2488.

[15] Manzini, Paola, and Marco Mariotti. 2007. "Rationalizing Boundedly Rational Choice," *American Economic Review* 97: 1824-1839.

[16] Redelmeier, D. and Eldar Shafir. 1995. "Medical decision making in situations that offer multiple alternatives," *Journal of the American Medical Association* 273: 302-305.

[17] Roelofsma, Peter. H.M.P. and Daniel Read. 2000. "Intransitive Intertemporal Choice," *Journal of Behavioral Decision Making* 13: 161-177.

[18] Salant, Yves and Ariel Rubinstein. 2007. "(A,f): Choices with Frames," *Review of Economic Studies*, forthcoming.

[19] Saari, Donald G. 1989. "A dictionary for voting paradoxes." *Journal of Economic Theory* 48(2): 443-475.

[20] Saari, Donald G. 2000. "Mathematical structure of aggregation paradoxes II: Positional aggregation," *Economic Theory* 15: 55-102.

[21] Schelling, Thomas. C. 1984. "Choice and Consequence," Harvard University Press.

[22] Sen, Amartya. 1977. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy and Public Affairs* 6(4): 317-344.

[23] Shafir, Eldar. 1993. "Choosing versus rejecting: Why some options are both better and worse than others", *Memory and Cognition*, 21(4): 546-556.

[24] Shafir, Eldar, Itamar Simonson and Amos Tversky. 1993. "Reason-based Choice," *Cognition* 49: 11-36.

[25] Shannon, Chris. 2006. “A prevalent transversality theorem for Lipschitz functions.” *Proceedings of the American Mathematical Society* 134(9): 2755-2765.

[26] Simonoson, Itamar. 1989. "Choice based on reasons: the case of attraction and compromise effects," *Journal of Consumer Research* 16: 158-174.

[27] Sobel, Joel. 2005. “Interdependent Preferences and Reciprocity,” *Journal of Economic Literature* 43(2): 392-436.

[28] Strotz, R.H. 1956. “Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies* 23(3): 165-180.

[29] Tversky, Amos. 1969. "Intransitivity of Preference," *Psychological Review* 76: 31-48.

[30] Young, H. Peyton. 1975. "Social choice scoring functions." *SIAM Journal on Applied Mathematics* 28(4): 824-838.

A Proofs

A.1 Representation

To establish lemma 1 we introduce some notation. Recall that $n \equiv X$. Fix $a \in X$. Let Π^{-a} denote the set of strict preferences on $X \setminus \{a\}$. For each $\rho \in \Pi^{-a}$ let $\Pi^{\rho} \equiv \{\pi \in \Pi |\ x\rho y \Rightarrow x\pi y,\ x, y \in X \backslash \{a\}\}$, i.e., the set of preferences on X that are consistent with ρ on $X \setminus \{a\}$. Observe that $\{\Pi^{\rho}\}_{\rho \in \Pi^{-a}}$ is a partition of Π. Note also that, since preferences in Π^{ρ} share the same relative ranking of all alternatives other than a, each preference in the set is defined precisely by the ranking of alternative a. In particular, we can write $\Pi^{\rho} = \{\pi_1^{\rho}, ..., \pi_n^{\rho}\}$ where π_k^{ρ} is the preference in Π^{ρ} that ranks alternative a in position $k \in \{1, ..., n\}$.

For any $\lambda \in \Delta^{\Pi}$ and $K \subseteq \Pi$ let $\lambda(K) \equiv \sum_{\pi \in K} \lambda_{\pi}$. Fix $\rho \in \Pi^{-a}$ and $\lambda \in \Delta^{\Pi}$ and let $\lambda^{\rho} \in \Delta^{\Pi}$ denote the distribution on Π^{ρ} induced by λ. Formally, for ρ such that $\lambda(\Pi^{\rho}) > 0$, let $\lambda_{\pi}^{\rho} = 0$ if $\pi \notin \Pi^{\rho}$ and $\lambda_{\pi}^{\rho} = \frac{\lambda_{\pi}}{\lambda(\Pi^{\rho})}$ if $\pi \in \Pi^{\rho}$; for ρ such that $\lambda(\Pi^{\rho}) = 0$ we just set λ^{ρ} equal to an arbitrary distribution independent of λ with support in Π^{ρ}. By construction, the vectors $\{\lambda^\rho\}_{\rho\in\Pi^{-a}}$ are mutually orthogonal to each other, $supp(\lambda^\rho) \subseteq \Pi^\rho$, and λ can be decomposed as

$$\lambda = \sum_{\rho:\lambda(\Pi^\rho)>0} \lambda(\Pi^\rho)\lambda^\rho. \tag{5}$$

The next lemma shows that a-FOSD for distributions in Π^ρ reduces to a much simpler condition.

Lemma 6 *Fix $a \in X$. For any $\rho \in \Pi^\rho$ we have that μ^ρ $a-$FOSD $\lambda^\rho \Leftrightarrow Q_{aX}(\mu^\rho)$ FOSD $Q_{aX}(\lambda^\rho)$.*

Proof. By definition, showing μ^ρ $a-$FOSD λ^ρ requires proving that, for all A that contains a, (i) $Q_{aA}(\mu^\rho)$ FOSD $Q_{aA}(\lambda^\rho)$ and (ii) $Q_{a'A}(\lambda^\rho)$ FOSD $Q_{a'A}(\mu^\rho)$ for all $a' \in A\backslash\{a\}$. Clearly, $Q_{aX}(\mu^\rho)$ FOSD $Q_{aX}(\lambda^\rho)$ is a necessary condition for (i) to hold. We need to show that it is sufficient for both (i) and (ii). To show this, we show that the cumulative distributions Q_{aA} can be expressed in terms of Q_{aX}.

Fix $\rho \in \Pi^{-a}$. Since all preferences in Π^ρ have the same relative ranking of all alternatives other that a, we can relabel the alternatives so that

$$X = \{x_1, x_2, ..., x_{n-1}, a\}$$

and $x_j \pi x_{j+1}$ for all $\pi \in \Pi^\rho$ and $j \in \{1, 2, ..., n-1\}$. With this convention, any choice set $A \subset X$ containing alternative a can be expressed as

$$A = \{x_{j_1}, x_{j_2}, ..., x_{j_{|A|-1}}, a\},$$

so that $x_{j_k} \pi x_{j_{k+1}}$ for all $\pi \in \Pi^\rho$.

Recall that $rank(a, A, \pi) \in \{1, ..., |A|\}$ denotes the rank of $a \in A$ among alternatives in A under order π and π_k^ρ is the preference in Π^ρ that ranks a at position k. Using the above notation it is easy to verify that

$$rank(a, A, \pi_k^\rho) = \begin{cases} 1 & \text{if } k \leq j_1 \\ r & \text{if } j_{r-1} < k \leq j_r,\ r \in \{2, ..., |A|-1\} \\ |A| & \text{if } k > j_{|A|-1}. \end{cases} \tag{6}$$

Now, by definition, $Q_{aA}^r(\lambda^\rho) = \sum_{k:rank(a,A,\pi_k^\rho)\leq r} \lambda_{\pi_k^\rho}^\rho$, which combined with (6) yields

$$Q_{aA}^r(\lambda^\rho) = \sum_{k=1}^{j_r} \lambda_{\pi_k^\rho}^\rho = Q_{aX}^{j_r}(\lambda^\rho).$$

Similarly, $Q^r_{aA}(\mu^\rho) = Q^{j_r}_{aX}(\mu^\rho)$. Thus, a sufficient condition for $Q_{aA}(\mu^\rho)$ FOSD $Q_{aA}(\lambda^\rho)$ to hold for any set A containing a is that $Q_{aX}(\mu^\rho)$ FOSD $Q_{aX}(\lambda^\rho)$.

On there hand, if $A = \{x_{j_1}, x_{j_2}, ..., x_{j_{|A|-1}}, a\}$ as above we have that

$$rank(x_{j_r}, A, \pi^\rho_k) = \begin{cases} r & \text{if } k > j_r \\ r+1 & \text{if } k \leq j_r. \end{cases} \tag{7}$$

By definition, $Q^r_{x_{j_r}A}(\mu^\rho) = \sum_{k:rank(x_{j_r},A,\pi^\rho_k)\leq r} \mu^\rho_{\pi^\rho_k}$ which using (7) gives

$$Q^r_{x_{j_r}A}(\mu^\rho) = \sum_{k>j_r} \mu^\rho_{\pi^\rho_k} = 1 - \sum_{k\leq j_r} \mu^\rho_{\pi^\rho_k} = 1 - Q^{j_r}_{aX}(\mu^\rho).$$

Similarly, $Q^r_{x_{j_r}A}(\lambda^\rho) = 1 - Q^{j_r}_{aX}(\lambda^\rho)$.

We conclude that $Q_{a'A}(\lambda^\rho)$ FOSD $Q_{a'A}(\mu^\rho)$ for all $a' \in A\backslash\{a\}$ and any A containing a translates into $Q_{aX}(\mu^\rho)$ FOSD $Q_{aX}(\lambda^\rho)$. The proof is complete. ■

Recall that $m(a, \pi)$ is the set of preferences that preserve the ranking of π for alternatives in $X \setminus \{a\}$ but rank a better than π does. On the other hand, $M(a, \lambda) = \{\mu \in \Delta^\Pi | \ \mu = W^T\lambda, \ W \in \mathcal{W}(a)\}$ is the set distributions that can be obtained from λ by a sequence of monotonic transformations with respect to a. Here $\mathcal{W}(a)$ is the set of $\Pi \times \Pi$ stochastic matrices such that for each $W \in \mathcal{W}(a)$ we have that $W(\pi, \pi') \in [0, 1]$, $W(\pi, \pi') = 0$ unless $\pi' \in m(a, \pi)$, and $\sum_{\pi'\in\Pi} W(\pi, \pi') = 1$.

Lemma 7 *Fix a. For any $W \in \mathcal{W}(a)$, $W(\pi, \pi') > 0$ only if for some $\rho \in \Pi^{-a}$, we have that $\pi, \pi' \in \Pi^\rho$ and $\pi' \in m(a, \pi)$. Thats is, $\pi = \pi^\rho_k$ and $\pi' = \pi^{\rho'}_j$ with $\rho' = \rho$ and $j \leq k$.*

Proof. The result follows from two facts. First, $\{\Pi^\rho\}_{\rho\in\Pi^{-a}}$ is a partition of Π, where each element of the partition is defined precisely by the fact that the relative order of all alternatives other than a is fixed. Second, as observed earlier, $\Pi^\rho = \{\pi^\rho_1, ..., \pi^\rho_n\}$ where π^ρ_k is the preference in Π^ρ that ranks alternative a in position $k \in \{1, ..., n\}$.

From the previous it follows that $\pi = \pi^\rho_k$ and $\pi = \pi^{\rho'}_j$ for some ρ, ρ', k and j. It also follows that $m(a, \pi^\rho_k) = \{\pi^\rho_1, ..., \pi^\rho_k\}$, so that $\pi' \in m(a, \pi)$ only if $\rho' = \rho$ and $j \leq k$. The conclusion follows. ■

Lemma 8 *If $\mu \in M(a, \lambda)$ then for all $\rho \in \Pi^{-a}$ (i) $\mu(\Pi^\rho) = \lambda(\Pi^\rho)$ and (ii) μ^ρ a-FOSD λ^ρ.*

Proof. Let $\mu \in M(a, \lambda)$. By definition, this means that $\mu = W^T \lambda$ for some $\Pi \times \Pi$ matrix $W \in \mathcal{W}(a)$. Part (i) is immediate from lemma 8 as any stochastic matrix in $W \in \mathcal{W}(a)$ implies probability mass transfers within each set Π^ρ. That is, it preserves the mass of of each Π^ρ. To establish (ii), using (5), by linearity we have

$$\mu = \sum_{\rho:\lambda(\Pi^\rho)>0} \lambda(\Pi^\rho) W^T \lambda^\rho = \sum_{\rho:\mu(\Pi^\rho)>0} \mu(\Pi^\rho) W^T \lambda^\rho$$

where the last step uses part (i). Thus, using (5) for μ, we conclude that $W^T \lambda^\rho$ is the distribution on Π^ρ induced by μ, i.e., $\mu^\rho = W^T \lambda^\rho$.

From lemma 6, we just need to show that $Q_{aX}(\mu^\rho)$ FOSD $Q_{aX}(\lambda^\rho)$. This is straightforward as any matrix $W \in \mathcal{W}(a)$ is such that, the vector $\mu^\rho = W^T \lambda^\rho$ "shifts" mass in λ^ρ to preferences in Π^ρ that rank alternative a better. Indeed, some algebra shows that, for any $r \in \{1, ..., n-1\}$,

$$Q^r_{aX}(\mu^\rho) = Q^r_{aX}(\lambda^\rho) + \sum_{j>r} \sum_{k=1}^{r} W(\pi^\rho_j, \pi^\rho_k) \lambda^\rho_{\pi^\rho_j} \geq Q^r_{aX}(\lambda^\rho).$$

■

Lemma 9 *Fix $a \in X$ and $\lambda \in \Delta^\Pi$. Then $M(a, \lambda) \subseteq FOSD(a, \lambda)$.*

Proof.

Let A be any subset that contains alternative a and a' be generic alternative in that set. For any $\lambda \in \Delta^\Pi$ we have that $\lambda = \sum_{\rho:\lambda(\Pi^\rho)>0} \lambda(\Pi^\rho)\lambda^\rho$ and, by the linearity of the cumulative distribution map $Q_{a'A}$, we have

$$Q_{a'A}(\lambda) = \sum_{\rho:\lambda(\Pi^\rho)>0} \lambda(\Pi^\rho) Q_{a'A}(\lambda^\rho).$$

Let $\mu \in M(a, \lambda)$. We show that $\mu \in FOSD(a, \lambda)$. We start by showing that $Q_{aA}(\mu)$ FOSD $Q_{aA}(\lambda)$. Indeed, from Lemma 8, for each $\rho \in \Pi^{-a}$ we have that for any such μ, $\mu(\Pi^\rho) = \lambda(\Pi^\rho)$ and μ^ρ $a-$FOSD λ^ρ. The latter means that $Q_{aA}(\mu^\rho) \geq Q_{aA}(\lambda^\rho)$. Hence,

$$\begin{aligned} Q_{aA}(\mu) - Q_{aA}(\lambda) &= \sum_\rho \mu(\Pi^\rho) Q_{aA}(\mu^\rho) - \lambda(\Pi^\rho) Q_{aA}(\lambda^\rho) \\ &= \sum_\rho \lambda(\Pi^\rho) \left[Q_{aA}(\mu^\rho) - Q_{aA}(\lambda^\rho) \right] \geq 0. \end{aligned}$$

We conclude that $Q_{aA}(\mu)$ FOSD $Q_{aA}(\lambda)$.
The same argument shows that if $Q_{bA}(\lambda^\rho)$ FOSD $Q_{bA}(\mu^\rho)$ for each ρ and $b \neq a$ then $Q_{bA}(\lambda)$ FOSD $Q_{bA}(\mu)$. It follows that μ $a-$FOSDλ, i.e., $\mu \in FOSD(a, \lambda)$. ■

A.1.1 Proof of lemma 1

We introduce some notation. Let $\mathcal{S} \subset \mathcal{A}$ be the collection of subsets of X with two or more elements that contain $a \in X$. The cardinality of $\mathcal{S}$ is $S_n = \sum_{k=2}^{n} \binom{n-1}{k-1} = 2^{n-1} - 1$. Consider an arbitrary labelling of these subsets so that $\mathcal{S} = \{A_1, ..., A_{S_n}\}$. Similarly, for each $A_j \in \mathcal{S}$, we label alternatives so that $A_j = \{b_{j_1}, ..., b_{j_{|A_j|}}\}$. Given $A \in \mathcal{S}$ and $a' \in A$ each cumulative distribution vector $Q_{a'A}(.) \in [0,1]^{|A|-1}$ write $CD_A = [0,1]^{(|A|-1)\times|A|}$ and let $Q_A : \Delta^\Pi \to CD_A$ the map that assigns to each distribution of preferences $\mu \in \Delta^\Pi$ the "stack" of all cumulative distribution vectors $Q_{b_{j_k}A_j}(\mu) \in [0,1]^{|A_j|-1}$. Let $\widetilde{Q} : \Delta^\Pi \to \times_{s=1}^{S_n} CD_{A_s}$ be the map that stacks all the $Q_A(\mu)$'s together with the convention that components that do not involve a have a negative sign. With this convention, $\mu \in FOSD(a, \lambda) \iff \widetilde{Q}(\mu) \geq \widetilde{Q}(\lambda)$, where the inequality holds for each component of the vector.
We need to show that $\widetilde{Q}(FOSD(a, \lambda)) = \widetilde{Q}(M(a, \lambda))$. From lemma 9 it follows that $\widetilde{Q}(M(a, \lambda)) \subset \widetilde{Q}(FOSD(a, \lambda))$. It remains to show that $\widetilde{Q}(FOSD(a, \lambda)) \subset \widetilde{Q}(M(a, \lambda))$. We start by noting that since the $Q_{bA}(\cdot)$ maps are linear, so is $\widetilde{Q}(\mu)$. We establish the result by exploiting the linearity of $\widetilde{Q}(\cdot)$ and a dimensionality argument.

Let $\Delta(a, \lambda) = \{\mu \in \Delta^\Pi | \mu(\Pi^\rho) = \lambda(\Pi^\rho),\ \rho \in \Pi^{-a}\}$. Observe that $M(a, \lambda) \subset \Delta(a, \lambda)$. Our first step is to show that $\widetilde{Q}(\Delta^\Pi) = \widetilde{Q}(\Delta(a, \lambda))$ for any $\lambda \in \Delta^\Pi$. This implies that for any $\mu \in FOSD(a, \lambda)$, there exists $\mu' \in \Delta(a, \lambda)$ such that $Q_{bA}(\mu) = Q_{bA}(\mu')$ for all $b \in A, A \in \mathcal{A}$. (Clearly, μ' is also in $FOSD(a, \lambda)$ as it spans the same cumulative distributions as μ). Using this the result, the second step shows that for any $\mu \in FOSD(a, \lambda) \cap \Delta(a, \lambda)$ there exists $\mu \in M(a, \lambda)$ such that $\widetilde{Q}(\mu') = \widetilde{Q}(\mu)$.
Step 1: $\widetilde{Q}(\Delta^\Pi) = \widetilde{Q}(\Delta(a, \lambda))$

We show the result by assessing the dimensionality of $\widetilde{Q}(\Delta^\Pi)$. Note that for a fixed b and A, $Q_{bA}(\mu)$ has $|A|-1$ components and the component $Q^r_{bA}(\mu)$, $r \in \{1, ..., |A|-1\}$ is the mass that μ assigns to preferences that rank b in position r or better in set A. Now, for any r, since the mass on preferences that rank some alternative in position j is one, $\sum_{b\in A} Q^r_{bA}(\mu) = r$ for all μ. This implies $|A|-1$ constraints on the components of the stack vector Q_A for each set A. In addition, the Borda score of b in A is simply $e^T Q_{bA}(\mu) = \sum_{r=1}^{|A|-1} Q^r_{bA}(\mu)$. If $|A| \geq 3$ this score is entirely determined by scores from pairs that contain b and some other alternative $b' \in A$. Indeed, $e^T Q_{bA}(\mu)$ is proportional to $\sum_{B=\{b,b'\},b'\in A\setminus\{b\}} Q^1_{bB}(\mu)$ (see for example, Saari [2000]). Thus, if A has three or more elements, given the cumulative vectors for pairs, there are $|A|$ additional "Borda" constraints, one for each alternative in b. It follows that, excluding the $\binom{n-1}{1} = n-1$ pairs that contain a, the number of free parameters in Q_A is at most $(|A|-1)|A| - (|A|-1) - |A| = (|A|-2)(|A|-1)$. For each pair $B = \{a,b\}$, $b \in X\setminus\{a\}$, there is a single free parameter $Q^1_{bB}(\mu)$. Thus, adding across subsets $A \in \mathcal{S}$, we get the number of free parameters of $\widetilde{Q}$ is at most

$$d_n \equiv \sum_{k=3}^{n} (k-2)(k-1)\binom{n-1}{k-1} + n - 1 = (n-1)(n-2)2^{n-3} + n - 1,$$

In sum, the dimension of the image of $\widetilde{Q}$ satisfies $\dim(\widetilde{Q}(\Delta^\Pi)) \leq d_n = (n-1)(n-2)2^{n-3} + n - 1$.

We now calculate $\dim(\Delta(a,\lambda))$. Any $\mu \in \Delta(a,\lambda)$ satisfies $\mu(\Pi^\rho) = \lambda(\Pi^\rho)$. These are $|\Pi^{-a}| = (n-1)!$ constraints, but one of them is implied by all of the others as $\sum_\rho \mu(\Pi^\rho) = \sum_\rho \lambda(\Pi^\rho) = 1$. Since the dimensionality of Δ^Π is $|\Pi| - 1 = n! - 1$, we conclude that

$$\dim(\Delta(a,\lambda)) = n! - 1 - ((n-1)! - 1) = (n-1)! \times (n-1).$$

Observe that for $n = 3$, we have that $\dim(\Delta(a,\lambda)) = 4$ and, from above, $d_3(\widetilde{Q}) = 4$. By induction on n, it is straightforward to show that $d_n \leq \dim(\Delta(a,\lambda))$ for all $n \geq 3$. (Intuitively, $(n-1)! \times (n-1) = O((n/e)^n)$ grows faster than $d_n = O(n^2 2^n)$.) Thus, $\dim(\widetilde{Q}(\Delta^\Pi)) \leq d_n \leq \dim(\Delta(a,\lambda))$, so that

$$\dim(\widetilde{Q}(\Delta^\Pi)) \leq \dim(\Delta(a,\lambda)).$$

On the other hand, $\Delta(a,\lambda) \subset \Delta^\Pi$. Combining these two, we conclude that $\widetilde{Q}(\Delta^\Pi) = \widetilde{Q}(\Delta(a,\lambda))$.

Step 2: For any $\mu \in \Delta(a,\lambda)$ such that $\mu \in FOSD(a,\lambda)$ there exists $\mu' \in M(a,\lambda)$ such that $\widetilde{Q}(\mu') = \widetilde{Q}(\mu)$.

Let $\mu \in \Delta(a,\lambda) \cap FOSD(a,\lambda)$. We provide a procedure o find $\mu' \in M(a,\lambda)$ such that $\widetilde{Q}(\mu') = \widetilde{Q}(\mu)$. Note that $\mu' \in M(a,\lambda)$ if there exists a "transfer" matrix $W \in \mathcal{W}(a)$ such that $\mu' = W^T\lambda$.

We start by observing that for any λ and μ there exists a stochastic matrix $\hat{W}$ -possibly multiple matrices, such that $\mu = \hat{W}^T\lambda$. This matrix describes how to relocate the mass described by the distribution λ to obtain μ, each element $\hat{W}(\pi',\pi)$ is the fraction of $\lambda_{\pi'}$ transferred to μ_π. If we consider the restriction that $\mu \in \Delta(a,\lambda)$, which implies that the mass of preferences in Π^ρ remains fixed, we can further restrict $\hat{W}$ so that $\hat{W}(\pi',\pi) > 0 \Rightarrow \pi,\pi' \in \Pi^\rho$ for some ρ. This means $\hat{W}$ describes mass transfers for preferences within the same set Π^ρ. Of course, if $\hat{W}(\pi',\pi) = 0$ for all $\pi \notin m(a,\pi')$ we are done as it would mean that $\hat{W} \in \mathcal{W}(a)$ or, equivalently, that $\mu \in M(a,\lambda)$.

Suppose instead that $\hat{W}(\pi',\pi) > 0$ for some π' and $\pi \notin m(a,\pi')$. We refer to this type of transfer from preferences that rank a higher to preferences that rank a lower preserving the relative ordering of the other alternatives as a "negative transfer". Of course, it must also be the case that $\hat{W}(\pi'',\pi) > 0$ for some π'' and $\pi \in m(a,\pi'')$, otherwise we would have that $\lambda \in FOSD(a,\mu)$ rather than $\mu \in FOSD(a,\mu)$. These transfers are referred as "positive transfers". We can thus decompose the vector μ into three components μ^-, μ_0^+ and μ_1^+ such that

$$\mu = \mu^- + \mu_0^+ + \mu_1^+.$$

The first component μ_- captures all the "negative transfers" described in W_μ, the second and third components μ_0^+ and μ_1^+ correspond to "positive transfers". The vector μ_0^+ is chosen to neutralize the negative transfers captured by μ_-. More precisely, $\mu^- + \mu_0^+$ is in the Kernel of the transformation $\Delta\widetilde{Q}(v) = \widetilde{Q}(v) - \widetilde{Q}(\lambda)$. It can be shown that $\mu' \equiv \mu_1^+ + \lambda \in M(a,\lambda)$, the details are omitted.

A.2 Prevalence: Almost Every Monotonic Rule is Regular

The following definitions are from Anderson and Zame [2001].

Definition 9 *[Shyness/Prevalence] Let Z be a topological vector space and let $K \subset Z$ be a convex Borel subset of Z which is completely metrizable in the relative topology. Let $\phi \in K$. A universally measurable subset $S \subset Z$ is* *shy in K at ϕ if for each $\delta > 0$ and each neighborhood W of 0 in Z, there is a regular Borel probability measure ν on Z with compact support such that supp $\rho \subset (\delta(K - K) + K) \cap (W + K)$ and $\nu(S + z) = 0$ for every $z \in Z$. The set E is shy in K if it is shy at each point $K \in K$. A (not necessarily universally measurable) subset $F \subset K$ is shy in K if it is contained in a shy universally measurable set. A subset $P \subset K$ is prevalent in K if its complement $K \backslash P$ is shy in K.*

We use L_F to denote the Lebesgue measure on finite-dimensional vector space F. A straightforward example of a shy set is a set such that it and all of its translates have Lebesgue measure 0 in some finite-dimensional subspace. Formally:

Definition 10 *[Finitely Shy/Prevalent] Let Z be a topological vector space and let $K \subset Z$ be a convex Borel subset of Z which is completely metrizable in the relative topology. A universally measurable subset $S \subset K$ is finitely shy in K if there is a finite-dimensional subspace $N \subset Z$ such that $L_N(K + a) > 0$ for some $a \in Z$ and $L_N(S + z) = 0$ for every $z \in Z$. A (not necessarily universally measurable) subset $T \subset K$ is finitely shy in K if it is contained in a finitely shy universally measurable set. A subset $P \subset K$ is finitely prevalent in K if its complement $K \backslash P$ is finitely shy in K.*

All finitely prevalent sets in K are also prevalent in K (Anderson and Zame [2001]). If a set P is prevalent in K we say that "almost every"element of K satisfies the property that defines the elements in P.

A.2.1 Proof of Theorem 3

Throughout the proof, we use the same notation introduce in the section 4.

In our case the ambient space Z is the space of locally Lipschitz continuous functions from Δ^{Π} to $\mathbb{R}^{S_n}$. Observe that Z endowed with the sup norm is a topological vector space. From Corollary 1 the excess score function $\phi_v : \Delta^{\Pi} \rightarrow \mathbb{R}^{S_n}$ associated to a monotonic rule $v \in V^m$ is locally Lipschitz-continuous. Let $C^m \equiv \{\phi \in Z | \ \phi$ is an excess score function for some $v \in V^m\}$. (Note that any $\phi \in C^m$ can be described as $\phi = (\phi_1, ..., \phi_{S_n})$ where each ϕ_j is an excess score map for set $A_j \in A$, i.e., there exist some monotonic and continuous function H_j such that $\phi_{jk} = H_j \circ Q_{a_k,A_j} - H_j \circ Q_{a_{k+1},A_j}$.) Finally, let $R^m \equiv \{\phi \in C^m | \ \phi$ is regular at $e_{\Pi}\}$. Our purpose is to show:

Theorem 3 *R^m is finitely prevalent in C^m (thus, prevalent in C^m).*

Our proof parallels the proof of Theorem 3.7 in Shannon [2006]. A similar proof technique is used to show Theorem 5.2 in Anderson and Zame [2001]. We start by introducing a finite-dimensional subspace in which the desired property holds "Lebesgue almost everywhere" in that subspace.[16] Let N^s be the set of excess score functions associated to the set of scoring rules. This set is not a subspace as can be shown that it is not closed to multiplication by negative scalars. Let $\overline{N}^s \equiv \{\phi \in Z | -\phi \in N^s\}$. By construction, $N \equiv N^s \cup \overline{N}^s$ is a subspace of Z. Clearly, $N \subset Z$ and this set is finite dimensional. Indeed, it is isomorphic to set of matrices of $\Pi \times S_n$. We use N as our probe.

Before showing our main result we need some preliminary steps:

Lemma 10 *Almost any $z \in N$ is regular at e_Π.*

The proof is omitted. It follows directly from Saari [1989, 2000] who shows that the set of scoring rules that are not regular (and do not span all choice functions) are a lower dimensional algebraic set of N^s.

Lemma 11 *$L_N(N^s) > 0$ and thus $L_N(C^m) > 0$.*

Proof. Note that $L_N(N^s) = L_N(\overline{N}^s)$. Since $L_N(N^s \cup \overline{N}^s) = 1$ we have that $L_N(N^s) \geq 1/2$. ■

Lemma 12 *Let E be a compact subset of a Euclidean space. Let $M(E)$ be the space of monotonic and continuous functions with domain E empowered with the sup norm ρ. The metric space $(M(E), \rho)$ is complete and convex.*

Proof. Convexity is straightforward: The convex combination of two continuous and monotonic functions is continuous and monotonic. We focus on completeness.

Since $M(E)$ is a subset of the subspace $C(E)$ of continuous functions and $C(E)$ is complete, any Cauchy sequence $\{H^k\} \subset M(E)$ converges to some continuous function H. This convergence is uniform as E is compact. Since H^k is monotonic for all k, if $Q \geq Q'$ then $H^k(Q) \geq H^k(Q')$ for all k. Taking the limit we have that the limit function satisfies $H(Q) \geq H(Q')$, which means that $H^k \in M(E)$. ■

[16] In Hunt, Sauer and (1992), and Hunt (1993) this space is called a *probe*.

Lemma 13 *C^m is a convex Borel set of Z.*

Proof. Throughout the proof, we use $E \equiv [0,1]^{|A|}$, for some integer $1 \leq |A| \leq n-1$ (i.e., a set of cumulative distribution vectors). As in the previous lemma, let $M(E)$ be the space of continuous and monotonic functions with domain E. Fix any pair of linear maps $Q, Q' : \Delta^\Pi \to E$, $Q \neq Q'$ and define $C_0^m = \{\phi : \Delta^\Pi \to \mathbb{R} | \ \phi = H \circ Q - H \circ Q', \ H \in M(E)\}$. The space of "single coordinate" excess score functions of an excess score function in C^m has exactly the same structure of C_0^m for some A, Q, and Q'. We show that (for any A, Q, and Q') the set of "component" excess scores C_0^m is a convex Borel set of Z_0, the set of locally Lipschitz continuous functions from Δ^Π to $\mathbb{R}$. This is sufficient to establish the result as C^m is the Cartesian product of spaces with structure of C_0^m and, respectively, $Z = \times_{i=1}^{S_n} Z_0$.

The convexity of C_0^m follows from the convexity of $M(E)$. Indeed, given $H, \widetilde{H} \in M(E)$ and, any $\alpha \in [0,1]$, let $H(\alpha) = \alpha H + (1-\alpha)\widetilde{H}$. From Lemma 12, $H(\alpha) \in M(E)$. Thus, given $\phi, \widetilde{\phi} \in C_0^m$ defined by $\phi = H \circ Q - H \circ Q'$ and $\widetilde{\phi} = \widetilde{H} \circ Q - \widetilde{H} \circ Q'$ we see that

$$\alpha\phi + (1-\alpha)\widetilde{\phi} = H(\alpha) \circ Q - H(\alpha) \circ Q'$$

is also an element of C_0^m.

We show that C_0^m is Borel in Z_0 by showing it is a closed set. This follows directly Lemma 12 above: The sequence $\{\phi^k\} \subset C_0^m$ with $\phi^k = H^k \circ Q - H^k \circ Q'$ converges to some $\phi \Leftrightarrow H^k$ converges to some H. By Lemma 12 $H \in M(E)$ (as $M(E)$ is complete). It follows that the limit $\phi = H \circ Q - H \circ Q'$ is in C_0^m. Hence, C_0^m is closed. ■

Lemma 14 *R^m is a Borel set of Z (thus, universally measurable).*

Proof. For each $\phi \in R^m$, the generalized Jacobian at e_Π, $\partial\phi(e_\Pi)$, has full rank. This rank is S_n, the dimension of the codomain of all functions in C^m. Given $\phi \in C^m$ (not necessarily regular at e_Π) we can calculate the determinant of all the $S_n \times S_n$ sub-matrices of the generalized Jacobian $\partial\phi(e_\Pi)$. Let $\theta(\phi)$ be the maximum absolute value of these determinants. By construction $\theta(\phi) \geq 0$ and $\theta(\phi) > 0$ if and only if $\partial\phi(e_\Pi)$ has full rank. Let

$$R_k = \{\phi \in Z | \theta(\phi) > 1/k\}.$$

Each of these sets is an open set in Z. It follows that $R^m = \cup_k R_k$ is Borel as it is a countable union of open sets in Z. ■

We can finally prove Theorem 3.

Proof of Theorem 3.

The result is established by showing that $S = C^m \backslash R^m$ is finitely shy in C^m. Note that from Lemma 13, C^m is a convex Borel subset of the ambient space Z. From Lemma 14, S is a Borel set and, thus, universally measurable. Let N be the finite-dimensional subspace of linear excess score functions defined above. From Lemma 11 there exists $a \in Z$ such that $L_N(C^m + a) > 0$. To prove that S is finitely shy in C^m we need to show that $L_N(S + z) = 0$ for every $z \in Z$.

To that end, let $z \in Z$ be arbitrary and consider $(S - z) \cap N$. Let $P_N : Z \to N$ be the projection map on the finite-dimensional space N. It's easy to verify that $\phi \in (S-z) \cap N \Leftrightarrow \phi \in P_N(S-z)$. From the linearity of P_N it follows that $(S - z) \cap N = S_N - z_N$, where $S_N = P_N(S)$ and $z_N = P_N(z)$. Note that S_N is the set of excess score functions in N that are not a regular at e_Π. From Lemma 10, S_N has Lebesgue measure 0 in N. Thus, the translate $S_N - z_N$ has Lebesgue measure 0 in N as well. Since $(S-z) \cap N = S_N - z_N$, $(S - z) \cap N$ has Lebesgue measure 0 in N. As z was arbitrary, S is finitely shy in C^m. ■