Random Utility Maximization with Indifference†

Faruk Gul

and

Wolfgang Pesendorfer

Princeton University

October 2013

Abstract

A random utility maximizing random choice function assigns probabilities to outcomes as if the decision maker randomly "chooses" a strict utility function and then picks from each option set the maximal element. We extend this notion to utility functions that permit indifference and show that a random choice function maximizes a random utility if and only if it satisfies a property that we call *double total monotonicity.* In the no-indifference case, the first total monotonicity condition is equivalent to the well-know Block-Marschak conditions and implies the second. Hence, our theorem is an extension of Falmagne's Theorem.

† This research was supported by grants from the National Science Foundation.

1. Introduction

The possibility of indifference plays an important role in the theory of deterministic choice. In structured settings, such as von Neumann-Morgenstern's theory of choice under risk, indifference arises from the continuity of preferences. To avoid indifference, the modeler would either have to impose artificial and inconvenient restrictions on the domain of preferences or abandon one or more of the key assumptions the theory. Even in the theory of discrete choice, permitting indifference enables the modeler to identify patterns in behavior that would be unrecognizable and would appear as inconsistencies were he committed to a theory that precludes indifference.

A fixed tie-breaking rule, for example, one that asserts that indifference between alternatives 1 and 2 is always resolved in favor of 1, renders the theory indistinguishable from a theory in which the decision-maker strictly prefers 1 to 2. Thus, to take advantage of the added flexibility , the modeler typically remains agnostic about how the decision maker resolves her indifferences. While making the theory more permissive, this creates a challenge when it comes to confronting evidence: what choice evidence supports the hypothesis that the decision-maker is indifferent between 1 and 2 if she is only permitted to choose at most one of these options? Deterministic choice models deal with this issue by implicity assuming that indifference is directly observable, by observing the decision-maker sometimes choose 1 and sometimes choose 2 when both 1 and 2 are available. Thus, a sufficiently large data set in which 2 is never chosen from the set $\{1, 2\}$ would be evidence that the decision maker is not indifferent between the two options.

Permitting choice to be stochastic is an alternative modeling device used to expand the range of choice data consistent with the theory. Here, the modeler interprets the possibility that the decision-maker sometimes choose 1 over 2 and sometimes choose 2 over 1 not (necessarily) as proof of indifference nor as proof of inconsistency but as evidence that the decision-maker's preference or utility function is random. The testable implications of such random utility models follow from the implicit hypothesis that the probability distribution of the randomly drawn utility function is stable and does not depend on the specific choice problem confronting the decision maker.

Most random utility models preclude indifference. In particular, in the context of random choice from subsets of a finite set of alternatives, random utility is synonymous with random strict utility. Hence, in this simple setting, a modeler can either take advantage of the flexibility afforded by indifference or the flexibility afforded by stochastic preferences but not both.

We present a new model of discrete random choice that enables the modeler to permit indifference *and* remain agnostic about how indifference is resolved; that is, our theory of random choice provides a stochastic analog of the (deterministic) multi-valued choice theories and identifies the stochastic counterpart of Houthakker's Axiom (i.e., the weak axiom of preference).

To illustrate our model, consider the simplest setting with two possible choices, 1 and 2. Let $u_1(1) > u_1(2)$, $u_2(1) < u_2(2)$ and $u_3(1) = u_3(2)$ be the three possible utility functions. Assume that 50% of the time, the decision maker 1 (DM1) strictly prefers 1 over 2 (i.e., has utility function u_1), 20% of the time strictly prefers 2 to 1 (i.e., has utility function u_2) and the remaining 30% of the time she is indifferent between the two options (i.e., has utility function u_3). Hence, π such that $\pi(u_1) = .5$, $\pi(u_2) = .2$ and $\pi(u_3) = .3$ is DM1's random utility function.

If the random utility function described above were the correct model of a particular decision maker, what data would we expect to see and how would we distinguish this decision maker from DM2 who is never indifferent and draws utility function u_1 with probability .65 and utility function u_2 with probability .35? If the data in question is a single, undifferentiated sample, then there is no way to distinguish between the two hypotheses. Suppose, however, that we have monthly choice data for the decision maker for twelve months. Each month, she chooses each option a large number of times. The process governing this decision maker's random utility function is stable throughout the year. However, we permit the way she breaks ties to vary across months.[1] This allows the modeler to accommodate choice data that are inconsistent with the hypothesis of monthly i.i.d. draws while maintaining the hypothesis of a stable process of drawing utilities.

[1] For concreteness, assume the choice data represent purchases in a supermarket. The supermarket does not change the goods on offer nor their prices but, once a month, changes the way goods are displayed. This affects how the decision maker resolves her indifferences.

Suppose that the frequency on 1's in the data varies from a high of .8 in June to a low of .3 in September and December with various intermediate values for the remaining months. Also suppose that the aggregate frequency of 1's in the data is 65%. We would say that this data is consistent with DM2 but not with DM1. The random utility model governing DM2's behavior is inconsistent with the significant monthly variation in the data. In contrast, DM1's random utility permits such a variation and, in fact, demands that it be observable when the collections of samples is sufficiently varied. However, our model offers no theory of what causes this variation.

Formally, we identify the data above with the (nonadditive) random choice function ρ such that $\rho(1,12) = .5$; $\rho(2,12) = .2$ and $\rho(1,12) + \rho(2,12) < 1 = \rho(12,12)$. Hence, the possible subadditivity of ρ is what permits indifference (i.e., u_3). Observing this data; that is, observing that indifference occurs 30% of the time means that the *frequency* of 1's in the data varies between .5 and .8. Let $U(A,12)$ be the set of utility functions that attain all their maxima in the set $A \subset \{1,2\}$; that is, $U(1,12) = \{u_1\}$, $U(2,12) = \{u_2\}$ and $U(12,12) = \{u_1, u_2, u_2\}$. Recall the random utility π above. Then, it is easy to verify that

$$\rho(1,12) = \sum_{u \in U(1,12)} \pi(u) = \pi(u_1) = .5$$

$$\rho(1,12) = \sum_{u \in U(2,12)} \pi(u) = \pi(u_2) = .2$$

$$\rho(12,12) = \sum_{u \in U(12,12)} \pi(u) = \pi(u_1) + \pi(u_2) + \pi(u_3) = 1$$

Compare ρ above with the predicted behavior $\hat{\rho}$ of the DM2 whose behavior is governed by $\hat{\pi}$ such that $\hat{\pi}(u_1) = .65$; $\hat{\pi}(u_2) = .35$ and $\hat{\pi}(u_3) = 0$. Then,

$$\rho(1,12) = \sum_{u \in U(1,12)} \hat{\pi}(u) = \hat{\pi}(u_1) = .65$$

$$\rho(2,12) = \sum_{u \in U(2,12)} \hat{\pi}(u) = \hat{\pi}(u_2) = .35$$

$$\rho(12,12) = \sum_{u \in U(12,12)} \hat{\pi}(u) = \hat{\pi}(u_1) + \hat{\pi}(u_2) + \hat{\pi}(u_3) = 1$$

Hence, we would expect DM2's frequency of choosing 1 to remain at or near 65% throughout the 12 months.

Our main theorem characterizes random choice functions that are consistent with random utility maximization with indifference. Our model is a generalization of the random strict utility model studied extensively in the discrete choice literature. However, our main assumption is not weaker than the well-known assumption of this theory. This is because our model, when specialized to additive random choice functions; that is to strict utility functions, is identical to the standard model.

1.1 Related Literature

The idea of modeling psychological (i.e., sensory) responses as stochastic and of analyzing frequencies of such responses is often credited to Fechner (1860). Thurstone (1927) was the first to apply Fechner's insights to more complex responses involving judgment and choice. The typical Thurstone experiment entailed presenting subjects with pairs of options and asking them to assess how they rank according to some criteria. Hence, a subject would assess if object i is heavier, a darker shade of gray or better than j; that is, choose i or $j \neq i$ in a binary comparison from the choice set $\{i, j\} \subset \{1, \ldots, n\}$. One of Thurstone's goals was to develop a subjective cardinal measure of "grayness" by using choice frequencies rather than subjective ordinal assessments.

Thurstone modelled the decision-maker as identifying the larger of two psychological scale values U_i and U_j when choosing from $\{i, j\} \subset \{1, \ldots, n\}$. Thus, when the decision-maker perceives $U_i > U_j$, she picks i. Thurstone assumed that the scale values have a joint normal distribution and thus invented the probit model and more generally the random utility model.

Luce (1959) provided an alternative model of random choice from arbitrary finite subsets. He postulated the choice axiom; i.e., the assumption that the ratio of the frequency with which a decision-maker chooses i over j is the same for any options set that contains both i and j. The axiom led to a sparse model of choice with $n-1$ parameters $x_1 = 1, x_2, \ldots x_n > 0$ such that the probability of choosing $i \in B$ from the option set B is

$$\rho(i, B) := \frac{x_i}{\sum_{j \in B} x_j}$$

for every B and $i \in B$.

Block and Marschak (1960) showed that the Luce model was also a random utility model. They also identified a collection of inequalities, the Block-Marschak conditions, that are necessary for any random choice model to be a random utility model. The first of these inequalities requires that the probability of choosing $i \in B$ gets smaller as more options are added to B. For any collection of Luce parameters $x_1, x_2, \ldots x_n$, Luce and Suppes (1965) provided explicit distributions for the random utilities U_i, so that the resulting random utility model yields the same choice probabilities as Luce model with parameters x_i, $i = 1, \ldots, n$. In particular, they showed that a one parameter family of independent double exponential distribution for random utilities is equivalent to the Luce model.[2]

Falmagne (1978) provided the first characterization of random utility maximization by showing that the Block-Marschak polynomials are sufficient for random utility maximization. Our main theorem extends Falmagne's by extending the definition of random utility maximization to permit indifference. We discuss the relationship between Falmagne's Theorem and ours extensively in sections 2 and 3 below.

Gul and Pesendorfer (2006) provided a characterization of random expected utility maximization over von Neumann-Morgenstern lotteries. Their axioms reveal that in this more structured setting, the first of the Block-Marschak conditions, monotonicity, together with the stochastic counterpart of the independence axiom, continuity and the requirement that only extreme points are chosen, characterizes random expected utility maximization.

Lu (2013) offers the first model of random utility with indifference. His choice objects are Anscombe-Aumann acts and his axioms ensure that only the decision-maker's prior is random. Lu's approach enables the modeler to avoid taking a stance on how indifference is resolved by formulating the random choice rule as a function on an algebra of sets. Thus, if $\{1, 2\}$ is a minimal set in that algebra, the random choice rule specifies only the overall probability of choosing 1 or 2 but does not specify the probability of choosing the individual elements $i = 1, 2$. We adopt Lu's approach, develop an alternative device for choice with indifference and characterize random utility maximization with indifference in the discrete setting.

[2] Luce and Suppes credit E. Holman and E. J. Marley with these results.

In addition to the work discussed above, there is an extensive literature that explores the possibility (and the empirical need) for random choice rules that violate the Block-Marschak conditions or specific random utility models. Luce (1977) provides a detail review of much of the literature up to that date. Gul, Natenzon and Pesendorfer (2012) provide an extension of Luce's model that is (almost) equivalent to random utility maximization (without indifference) and a survey of the more recent literature on the limitations of the Luce model and its extensions.

2. Random Choice Functions

Let $n > 0$ be an integer and let $N = \{1, \ldots, n\}$. Let $\mathcal{N}$ denote the set of all nonempty subsets of N and $X = \{(A, B) \in \mathcal{N} \times \mathcal{N} \,|\, A \subset B\}$. Given any $x = (A, B) \in X$, B is the set of *feasible options*, A is set of *considered options* and x is a *choice problem.* We call $\rho : X \to [0, 1]$ a *Random Choice Function* (RCF) if for all $(B, B) \in X$, $\rho(B, B) = 1$.

We interpret $\rho(A, B)$ as the *minimal* probability of ending up with an element in the considered set A given that the feasible set B. That is, given all the possible ways of "resolving indifferences," from each randomly drawn utility function, $\rho(A, B)$ is the minimal probability of ending up with a subset of A. We let ρ denote a generic RCF, $x = (A, B), y = (A', B')$ etc. denote generic choice problems and A, B denote respectively, the sets of considered and feasible options. When we write $\rho(x)$ or $\rho(A, B)$ it will be understood that ρ is an RCF and $x \in X$. We write $\rho(i, B)$, $\rho(ij, B)$, $\rho(Ai, Bij)$ etc. instead of the more cumbersome alternatives, $\rho(\{i\}, B)$ and $\rho(\{i, j\}, B)$ and $\rho(A \cup \{i\}, B \cup \{i, j\})$ respectively.

Let U be the set of all complete and transitive binary relations on N. That is, $u \subset N \times N$ if and only if (1) iuj or jui and (2) iuj and juk implies iuk. Clearly, a binary relation can be represented by a utility function if and only if it is an element of U. Therefore, we will call U the set of utilities and without risk of confusion, identify u with some real-valued function that represents it. Hence, we will write $u(i) \geq u(j)$ to mean iuj. Define U_s, the set of strict utilities as the subset of U that are antisymmetric; that is, the set of total orders on N. Hence, $U_s = \{u \in U \,|\, u(i) = u(j) \text{ implies } i = j\}$.

For any $u \in U$ and $B \in \mathcal{N}$, define $c_u(B) = \{i \in B \,|\, u(i) \geq u(j) \text{ for all } j \in B\}$. That is, c_u is the set of optimal $i \in B$ for the utility u. Therefore, we say c maximizes u if

$c = c_u$ for some $u \in U$ and we call any c that maximizes some utility rational. For any $(A, B) \in X$, let $U(A, B) = \{u \in U \,|\, c_u(B) \subset A\}$. Hence, $U(A, B)$ is the set of utility functions that would have all their maximizers in set A when the decision maker has the option set B. A function $\pi : U \to [0, 1]$ is a *random utility* if $\sum_{u \in U} \pi(u) = 1$. A random utility π is a *strict random utility* if $\pi(u) > 0$ implies $u \in U_s$. If $\pi(u) = 1$, we identify π with u and call it a utility. Then, ρ is maximizes π if

$$\rho(x) = \sum_{u \in U(x)} \pi(u) \tag{rum}$$

Our main theorem provides conditions that relate some known axioms on (deterministic) choice functions to conditions on RCFs. We briefly review these conditions: a *choice function* is a mapping $c : \mathcal{N} \to \mathcal{N}$ such that $c(B) \subset B$. Such a function maximizes $u \in U$ if $c = c_u$ for some $u \in U$. A choice function is a special case of an RCF: for $(A, B) \in X$, let

$$\rho_c(A, B) = \begin{cases} 1 & \text{if } c(B) \subset A \\ 0 & \text{otherwise.} \end{cases}$$

Hence, each choice function can be viewed as an RCF by identifying the choice function with its "indicator function." Thus, $c \to \rho_c$ maps each choice function to an *extreme* RCF function; i.e., an RCF such that $\rho(x) \in \{0, 1\}$ for all x. Since we have not made any assumptions on RCFs so far, the converse is not true: not every extreme ρ is equal to ρ_c for some choice function. For the equivalence, we need the following consistency condition:

$$\begin{gathered} \rho(A \cup A', B) \geq \rho(A, B) \\ \rho(A \cup A', B) + \rho(A \cap A', B) \geq \rho(A, B) + \rho(A', B) \end{gathered} \tag{c}$$

If ρ satisfies (c), then we say it is *consistent.* If ρ is extreme and consistent, then we say it is a *deterministic.* Note that if $\rho = \rho_c$, the first part of the consistency condition above is simply the observation that if considered set A contains all of the (optimal) choices from feasible set B, then so does any superset of A. The second part is the observation that if A and A' both contain all optimal alternatives, then so does $A \cap A'$. It is clear why consistency is necessary for ρ to equal some ρ_c. Conversely, it is easy to see that if ρ is consistent, then for every $B \in \mathcal{N}$, there is a minimal set $c(B)$ such that $\rho(c(B), B) = 1$.

Hence, this mapping $B \to c(B)$ is the desired choice function. Thus, we have shown that $\rho = \rho_c$ for some choice function c if and only if ρ is deterministic.

Consider following condition on a choice function:

$$c(B) \cup c(B') \subset B \cap B' \text{ implies } c(B) = c(B \cup B') \tag{w_-}$$

$$c(B) \cap B' \neq \emptyset \text{ implies } c(B') \cap B \subset c(B) \tag{w}$$

The second condition above, (w) is the well-known *weak axiom of revealed preference* or Houthakker's axiom which is known to be necessary and sufficient for c to be a utility maximizer; that is, for c to be rational.[3] Hence, if c is rational, then there exists a utility u such that

$$c(B) = c_u(B) := \{i \in B \,|\, u(i) \geq u(j) \text{ for all } j \in B\}$$

The first condition is, in general, weaker. For example, if $N = \{1, 2, 3\}$ and $c(N) = N, c(\{1, 2\}) = \{1\}, c(\{2, 3\}) = \{2\}, c(\{1, 3\}) = \{3\}$ then c satisfies (w_-) but not (w). To see that (w) implies (w_-) it suffices to observe that, for any u, c_u satisfies (w_-).

Though (w_-) is weaker than (w) in general, there is one special case when they are equivalent: if c is single-valued; that is, if $c(B)$ is a singleton for every $B \in \mathcal{N}$. In this case, define u as follows: for $i \neq j$, let iuj if and only if $\{i\} = c(\{i, j\})$. For any c that satisfies (w_-), it is easy to verify that this u is a strict utility and $c = c_u$. We say that c is *almost rational* if it satisfies (w_-).

Recall that an ρ is deterministic if and only if $\rho = \rho_c$ for some c and whenever ρ is deterministic this c is unique. We say that ρ is *rational deterministic* whenever $\rho = \rho_c$ for some rational c. It follows that ρ is rational deterministic if and only if ρ_{c_u} for some $u \in U$. When ρ is rational deterministic, we write $\rho = \rho_u$ rather than ρ_{c_u}.

Note that $\rho_u(x) = 1$ if $u \in U(x)$; otherwise $\rho_u(x) = 0$. Therefore, we can restate (rum) as follows:

$$\rho(x) = \sum_{u \in U} \pi(u) \cdot \rho_u(x) \tag{rr}$$

Hence, ρ is maximizes a random utility if and only if it is a convex combination of rational deterministic RCFs in which case we say it is *random rational*. Our main goal is to

[3] See Houthakker (1950).

characterize random rational choice functions and to relate this characterization to Falmagne's Theorem which establishes necessary and sufficient conditions for random strict rationality; that is, conditions that ensure that the RCF ρ can be expressed as a convex combination of strict utilities:

$$\rho(x) = \sum_{u \in U_s} \pi(u) \cdot \rho_u(x) \tag{rsr}$$

for some strict random utility π.

2.1 Total Monotonicity and Double Total Monotonicity

A finite set Y with a binary relation $\succeq$ is partially ordered set (poset) if $\succeq$ is reflexive antisymmetric and transitive; that is, (i) $x \succeq x$, (ii) $x \succeq y$ and $y \succeq x$ implies $x = y$ and (iii) $x \succeq y \succeq z$ implies $x \succeq y$. Let $[x] = \{y \,|\, x \succeq y\}$; that is, $[x]$ is the set of all predecessors of x. An element y is an immediate predecessor of x if $x \succeq y$ and $x \succeq z \succeq y$ implies $y = z$ or $x = z$. We let $[x-1]$ denote the set of immediate predecessors of x. Without risk of confusion, we write $x-1$ if this set is a singleton. We say x is interior if $[x-1] \neq \emptyset$; hence if x is not a minimal element of $\succeq$.

For any poset $(Y, \succeq)$ and real-valued functions F, f on Y, we say that f is the derivative of F and conversely, F is the integral of f if:

$$F(x) := \sum_{y \in [x]} f(y)$$

for all x. Clearly, every f has a unique integral F. It can be verified, inductively, that converse is also true: every F has a unique derivative f. We let ΔF denote this derivative.

If $\succeq$ is a total order; that is, if it is a complete partial order, then $[x-1]$ has at most one element, $x-1$, so that $\Delta F(x) = F(x) - F(x-1)$ for all interior x and $f(x) = F(x)$ for all minimal x. More generally,

$$\begin{aligned}\Delta F(x) = F(x) &- \sum_{x_1 \in [x-1]} F(x) + \sum_{x_1 \in [x-1]} \sum_{x_2 \in [x_1-1]} F(x_2) \\ &- \sum_{x_1 \in [x-1]} \sum_{x_2 \in [x_1-1]} \sum_{x_3 \in [x_2-1]} F(x_3) \cdots\end{aligned} \tag{m}$$

where we interpret a sum over an empty set as 0. Equation (m) above is known as the *Möbius inversion* formula.

A function F is *monotone* if $F(x) - F(y) \geq 0$ for all $y \in [x-1]$ and *totally monotone* if $\Delta F \geq 0$. A totally monotone function F is a *cumulative* if $F(x) = 1$ for every maximal element $x \in Y$. When Y is the set of all nonempty subsets of a finite set Ω ordered by set inclusion, a monotone F such that $F(\emptyset) = 0$ and $F(\Omega) = 1$ is called a *capacity.* In this context, totally monotone functions; that is, cumulatives are called *belief functions.* Our notion of totally monotonicity extends this familiar concept to arbitrary posets.

We can have multiple partial orders $\succeq_1, \succeq_2 \ldots$ on the same set Y. Then, we let $\Delta_k F$ denote the derivative of F with respect to partial order $\succeq_k$. For our main theorem, we define two partial orders on X:

$$\begin{aligned} (A,B) \succeq_1 (A',B') &\text{ if and only if } A' \subset A \text{ and } B \subset B' \\ (A,B) \succeq_2 (A',B') &\text{ if and only if } A \subset A' \text{ and } B \backslash A = B' \backslash A' \end{aligned} \tag{1}$$

Definition: *A random choice function is double totally monotone (DTM) if* $\Delta_2 \Delta_1 \rho \geq 0$.

To interpret DTM, consider the extreme RCF ρ_u that maximizes a single utility $u \in U$. In that case, $\Delta_1 \rho_u(A,B) \in \{0,1\}$ and $\Delta_u \rho_u(A,B) = 1$ if and only if

$$\begin{aligned} i, j \in A &\Rightarrow u(i) = u(j) \\ i \notin B \backslash A, j \in B \backslash A &\Rightarrow u(i) > u(j) \end{aligned} \tag{2}$$

To see why (2) must hold, define the function $f : X \to \{0,1\}$ such that $f(A,B) = 1$ if and only if (2) holds and $f(A,B) = 0$ otherwise. Note that $\rho_u(A,B) = 1$ implies that there is a unique $A' \subset A, B' \supset B$ such that $f(A',B') = 1$. Conversely, if $f(A',B') = 1$ for some $A' \subset A, B' \supset B$ then $\rho_u(A,B) = 1$. Recall that $(A,B) \succeq_1 (A',B')$ if $A' \subset A, B' \supset B$ and, therefore,

$$\rho_u(A,B) = \sum_{\{(A',B') | (A,B) \succeq_1 (A',B')\}} f(A,B)$$

It follows that $f = \Delta_1 \rho$, as desired.

Condition (2) identifies choice problems (A,B) that capture part of an "indifference curve" of u; the considered set A consists of the $u-$maximal elements in B and every alternative not in the feasible set B is weakly preferred to any alternative in B.

The second derivative $\Delta_2\Delta_1\rho_u(A,B)$ is either 0 or 1 with $\Delta_2\Delta_1\rho_u(A,B) = 1$ if and only if

$$\begin{aligned} i \in A, j \in A &\Rightarrow u(i) = u(j) \\ i \in A, j \in B\backslash A &\Rightarrow u(i) > u(j) \\ i \notin B, j \in B &\Rightarrow u(i) > u(j) \end{aligned} \tag{3}$$

An analogous argument to the one given to prove (2) can be used to proof (3): Define $g(A,B) = 1$ if and only if (3) holds and $g(A,B) = 0$ otherwise. Then, $\Delta_1\rho(A,B) = 1$ implies that there is a unique A', B' with $B\backslash A = B'\backslash A', A' \supset B'$ such that $g(A',B') = 1$. Conversely, if there is a A', B' with $B\backslash A = B'\backslash A', A' \supset B'$ such that $g(A',B') = 1$ then $\Delta_1\rho(A,B) = 1$. Recall that $(A,B) \succeq_2 (A',B')$ if $B\backslash A = B'\backslash A'$ and $A \subset A'$. Therefore,

$$\Delta_1\rho_u(A,B) = \sum_{\{(A',B')|(A,B)\succeq_2(A',B')\}} g(A,B)$$

It follows that $g = \Delta_2\Delta_1\rho(A,B)$ as desired.

Condition (3) identifies choice problems (A,B) that capture a complete indifference curve of u; the considered set A consists of an indifference class; every alternative not in the feasible set B is strictly preferred to A and any alternative in $B\backslash A$ is strictly worse. If ρ maximizes a random utility, then $\Delta_2\Delta_1\rho(A,B)$ corresponds to the probability of an (A,B) indifference curve. With this interpretation, it is clear that double total monotonicity is a *necessary* condition for random rationality. After all, the set of utility functions that give rise to a particular indifference curve cannot have negative probability.

It is easy to verify that double totally monotonicity implies total monotonicity. We say that ρ is *additive* if $\rho(A \cup A', B) = \rho(A,B) + \rho(A',B)$ whenever $A \cap A' = \emptyset$; ρ is totally monotone if $\Delta_1\rho \geq 0$. Lemma 2 below establishes that for an additive RCF, totally monotonicity is equivalent to the condition that Block and Marschak (1960) establish as necessary for random utility maximization. Falmagne (1978) establishes the sufficiency of these conditions.

Colonius (1984) is the first to use the Möbius inversion formula to analyze the random choice problem. He does so to simplify parts of Falmagne's original proof. The underlying poset in his application is the restriction of $\succeq_1$ to the set $X^1 := \{(i,B) \in X\}$. Fiorini

(2004) uses Colonius' formulation to provide an alternative proof of Falmagne's theorem; one that takes advantage of the totally unimodularity of matrices describing simple acyclic digraphs. Without additivity; that is, with indifference, total monotonicity is not sufficient; the stronger condition, double total monotonicity is required.

3. Random Rationality and Random Strict Rationality

Characterizing random utility maximization or random strict utility maximization amounts to identifying conditions that ensure that the RCFs that satisfies these conditions are convex combinations of extreme RCFs – i.e., RCFs that only take on the values 0 or 1 – that satisfy the conditions. The following Lemma summarizes the implications of total monotonicity and dual total monotonicity for extreme RCFs.

Lemma 1: *An extreme ρ is totally monotone (double totally monotone) if and only if $\rho = \rho_c$ for some almost rational (rational) c.*

Our main theorem establishes that double totally monotone RCFs are convex combinations of extreme double totally monotone RCFs. That is, double totally monotone RCFs are convex combinations or rational choice function. Lemma 3 below establishes the impossibility of an analogous result for totally monotone RCFs.

Theorem: *An RCF is random rational if and only if it is double totally monotone.*

Next, we relate the theorem above to Falmagne (1978). Recall that Falmagne characterizes additive RCF by showing that the Block-Marschak conditions are sufficient for random strict utility maximization: for all $(i, B) \in X$,

$$\Delta_1 \rho(i, B) \geq 0 \qquad (bm)$$

Block and Marschak (1960) establish that these conditions are necessary for random utility maximization. To relate our Theorem to Falmagne's we need to characterizes the second derivatives of additive RCFs:

Lemma 2: *If ρ is additive, then $\Delta_1 \rho(A, B) = \sum_{i \in A} \Delta_1 \rho(i, B)$ and*

$$\Delta_2 \Delta_1 \rho(A, B) = \begin{cases} \Delta_1 \rho(A, B) & \text{if } A \text{ is a singleton} \\ 0 & \text{otherwise} \end{cases}$$

Formally, Falmange's Theorem deals with a subset of the RCFs covered in our Theorem (additive versus arbitrary), imposes a weaker condition than our Theorem (total monotonicity versus double total monotonicity) and yields a stronger result than our Theorem (random strict utility maximization versus random utility maximization). Lemma 2 above establishes that for additive RCF totally monotonicity is in fact equivalent to double total monotonicity. We note below that if an additive RCF maximizes a random utility, this random utility must be strict. These two observations enable us to state Falmagne's results as a corollary of ours.

Corollary (Falmagne): *An additive RCF is random strict rational if and only if it is totally monotone.*

To see how Falmagne's Theorem follows from ours, note that if $\Delta_1 \rho \geq 0$ and ρ is additive, then Lemma 2 ensures that $\Delta_2 \Delta_1 \rho \geq 0$. Then, our Theorem establishes that ρ maximizes some random utility. To conclude the proof, it is enough to show that if ρ is additive and maximizes the random utility π, then π must be strict. To see that this is true, suppose ρ maximizes π and $\pi(u) > 0$ for some $u \in U \backslash U_s$. Then, there exists i, j such that $u(i) = u(j)$ and hence $0 = \rho(i, ij) = \rho(j, ij)$ while $\rho(ij, ij) = 1$ proving that ρ is not additive.

Finally, we consider the implication of total monotonicity without double total monotonicity on general (i.e., not necessarily additive) RCFs. In particular, we ask the following: does total monotonicity yield a theory of random almost rationality analogous to the way double total monotonicity yields a theory of random rationality? Lemma 1 above and the linearity of the Δ operator ensure that every if ρ is a convex combination of almost rational extreme RCFs (that is, random choice functions of the form ρ_c for some $c \in C$), then $\Delta_1 \rho \geq 0$. The converse turns out the be false: not every totally monotone RCF is a convex combination of extreme totally monotone RCFs. Formally, let C be the set of all choice function c that satisfy (w_-). A function $\nu : C \to [0, 1]$ is a probability if $\sum_{c \in C} \nu(c) = 1$ and an RCF ρ is random almost rational if

$$\rho = \sum_c \nu(c) \cdot \rho_c$$

for some probability ν. As we noted above, every random almost rational choice function is totally monotone. In the appendix, we provide an example proving that the converse is not true: that is, we establish the following:

Proposition: *There exists a totally monotone ρ such that $\rho \neq \sum_{c \in C} \nu(c) \cdot \rho_c$ for any probability ν.*

4. Appendix

Let $|T|$ denote the cardinality of any set T. Let $\mathcal{G}$ be the set of all $\succeq_1$-totally monotone RCFs and recall that $\mathcal{F}$ is the set of double totally monotone RCFs. For any collection of functions $\mathcal{H}$ and any Δ, let $\Delta\mathcal{H} := \{\Delta h \,|\, h \in \mathcal{H}\}$. Define

$$\mathcal{F}_o = \{F : X \to I\!R_+\}$$

$$\mathcal{F}' = \Delta_1 \mathcal{F}$$

$$\mathcal{F}'' = \Delta_2 \mathcal{F}'$$

For $F \in \mathcal{F}$, let F' denote $\Delta_1 F$ and F'' denote $\Delta_2 F'$.

Let $\mathcal{D} = \{D \in \mathcal{N} \,|\, D \neq N\}$ and for $B \in \mathcal{N}$, $D \in \mathcal{D}$,

$$[x]_1 = \{y \in X \,|\, x \succeq_1 y\}$$

$$[y]_2 = \{z \in X \,|\, y \succeq_2 z\}$$

$$X(A,B) = \{(\hat{A},\hat{B}) \in X \,|\, \emptyset \neq \hat{A} \cap B \subset A \text{ and } B \subset \hat{B}\}$$

$$X_-(D) = \{(A,B) \in X \,|\, B \backslash A = D\}$$

$$X_+(D) = \{(A,D) \in X\}$$

$$Z(B) = \{(\hat{A},\hat{B}) \in X \,|\, \emptyset \neq \hat{A} \cap B \text{ and } B \subset \hat{B}\}$$

Lemma A1: *For all $F \in \mathcal{F}_o$ and (i) $x \in X$ implies $F(x) = \sum_{y \in X(x)} F''(y)$; (ii) $B \in \mathcal{N}$ implies $F(B,B) = \sum_{x \in Z(B)} F''(x)$.*

Proof: By definition, $F(A,B) = \sum\limits_{y \in [A,B]_1} F'(y)$ and $F'(y) = \sum\limits_{z \in [y]_2} F''(z)$. Therefore,

$$F(A,B) = \sum_{y \in [A,B]_1} \sum_{z \in [y]_2} F''(z)$$

To conclude the proof of (i), we will show that

$$\sum_{y \in [A,B]_1} \sum_{z \in [y]} F''(z) = \sum_{z \in X(A,B)} F''(z)$$

To establish the equality above, will show that (1) for any $(\hat{A}, \hat{B}) \in X$ such that $\emptyset \neq \hat{A} \cap B \subset A$ and $B \subset \hat{B}$, there exists (A', B') such that $(A, B) \succeq_1 (A', B') \succeq_2 (\hat{A}, \hat{B})$, (2) $(A, B) \succeq_1 (A', B') \succeq_2 (\hat{A}, \hat{B})$ implies $(\hat{A}, \hat{B}) \in X(A, B)$ and (3) $(A, B) \succeq_1 (A_t, B_t) \succeq_2 (\hat{A}, \hat{B})$ for $t = 1, 2$ implies $A_1 = A_2$ and $B_1 = B_2$. The assertions, (1), (2) and (3), together imply that each term of the summation on right-hand side appears exactly once on the left-hand side and conversely, each term on the left-hand side appears exactly once on the right-hand side.

For (1), let $D = \hat{A} \backslash A$, $A' = \hat{A} \backslash D$ and $B' = \hat{B} \backslash D$. Then, it is easy to verify that $(A, B) \succeq_1 (A', B') \succeq_2 (\hat{A}, \hat{B})$.

For (2), assume $(A, B) \succeq_1 (A', B') \succeq_2 (\hat{A}, \hat{B})$. Hence, $\emptyset \neq A' \subset A$, $B \subset B'$ and for some $D \subset N \backslash B'$, $\hat{A} = A' \cup D$, $\hat{B} = B' \cup D$. That is, $\emptyset \neq A' = \hat{A} \cap B \subset A$ and $B \subset B' \subset \hat{B}$ as desired.

For (3), assume $(A, B) \succeq_1 (A_t, B_t) \succeq_2 (\hat{A}, \hat{B})$ for $t = 1, 2$. Let $D_t = \hat{A} \backslash A$. We first show that $A_1 = A_2$. If not, assume without loss of generality that $i \in A_1 \backslash A_2$. Then, $i \in A_1 \cup D_1$ and $i \in B$. Hence, $i \notin D_2$ and therefore $i \notin A_2 \cup D_2$, a contradiction. Hence, $A_1 = A_2$, $A_1 \cap D_1 = A_2 \cap D_2 = \emptyset$ and $A_1 \cup D_1 = A_1 \cup D_2$. Therefore, $D_1 = D_2$ and hence $D_1 = D_2 \subset (N \backslash B_1) \cap (N \backslash B_2)$ and $B_1 \cup D_1 = B_2 \cup D_2 = B_2 \cup D_1$ and therefore $B_1 = B_2$ as desired.

To prove (ii), we note that $Z(B) = X(B, B)$ and appeal to (i). $\square$

A binary relation u on N is a *partial utility* if there exists $D^u, E^u \subset N$, $D^u \cap E^u = \emptyset$, $D^u \cup E^u \neq N$ such that u restricted to $N \backslash (D^u \cup E^u)$ is complete and transitive and for $i, j \in N$ such that $\{i, j\} \cap (D^u \cup E^u) \neq \emptyset$,

$$iuj \text{ if and only if } i \in N \backslash (D^u \cup E^u) \text{ and } j \in D^u$$

Note that for any partial utility u, the sets D^u, E^u are unique. Let V be the set of all partial utilities. Note that $U \subset V$. In particular, $U = \{u \in V \mid D^u \cup E^u = \emptyset\}$.

For $x := (A, B) \in X$, define u_x as follows:

$$iu_xj \text{ if (and only if) } i \in A \text{ and } j \in B$$

Then, let $V_X = \{u_x \mid x \in X\}$. Lemma A2 below establishes that $V_X \subset X$.

Recall that for $u \in U$ and $B \in \mathcal{N}$, $c_u(B)$ is the set of maximizers of u in B. We extend this definition to $u \in V$ as follows:

$$c_u(B) = \{i \in B \mid iuj \text{ for all } j \in B\}$$

Note that $c_u(B)$ may be empty. For any $(A, B) \in X$, let $V(A, B) = \{u \in V \mid c_u(B) \subset A\}$. A function $\phi : V \to [0, 1]$ is a *partial random utility* if

$$\sum_{u \in V(B,B)} \phi(u) = 1$$

for all $B \in \mathcal{N}$. Let Φ denote the set of all partial random utilities. For $F \in \mathcal{F}$, let

$$\phi_F(u) = \begin{cases} F''(x) & \text{whenever } u = u_x \text{ for } x \in X \\ 0 & \text{if } u \in V \backslash V_X \end{cases}$$

For $D \in \mathcal{D}$, let

$$V_-(D) = \{v \in V \mid D^v = D\}$$
$$V_+(D) = \{v \in V \mid N \backslash E^v = D\}$$

for all $N \neq D \in \mathcal{N}$. Finally, for all $\phi \in \Phi$ and $D \in \mathcal{D}$, let

$$F_\phi(x) = \sum_{u \in V(x)} \phi(u)$$
$$\Phi_o(D, \phi) = \{\phi_o \in \Phi \mid \phi_o(u) = 0 \text{ for all } u \in V_-(D) \cup V_+(D) \text{ and } F_{\phi_o} = F_\phi\}$$

Let $\Phi(D, \phi)$ be the set of all $\phi_o \in \Phi(D, \phi)$ such that $\sum_{u \in V_-(D')} \phi_o(u) = \sum_{u \in V_-(D')} \phi(u))$ and $\sum_{u \in V_-(D')} \phi_o(u) = \sum_{u \in V_+(D')} \phi(u)$ for all $D' \in \mathcal{D} \backslash \{D\}$.

Lemma A2: *(i)* $V_X \subset V$. *(ii) For* $x = (A, B)$, $D^{u_x} = B \backslash A$ *and* $E^{u_x} = N \backslash B$. *(iii)* $F(x) = \sum_{u \in V(x)} \phi_F(u) = \sum_{u \in V(x)} \phi_F(u) \rho_u$ and $\phi_F \in \Phi$.

Proof: (i) and (ii) follow immediately from the definitions of V, V_X, u_x, D^{u_x} and E^{u_x}. To prove (iii), we note that $u_x \in V(A,B)$ if and only if $x \in X(A,B)$ and hence

$$\sum_{u \in V(A,B)} \phi(u) = \sum_{x \in X(A,B)} F''(x) = F(A,B)$$

by Lemma 1(i). Then, since $F \in \mathcal{F}$, we have $\sum_{u \in V(B,B)} \phi(u) = \sum_{x \in X(B,B)} F''(x) = F(B,B) = 1$. □

Lemma A3: *If $\phi \in \Phi$, $D \in \mathcal{D}$ and $\sum_{u \in V_-(D)} \phi(u) = \sum_{u \in V_+(D)} \phi(u) > 0$, then $\Phi(D,\phi) \neq \emptyset$.*

Proof: Let $\gamma = \sum_{u \in V_-(D)} = \sum_{u \in V_+(D)} \phi(u) > 0$. For $u \in V_-(D)$ and $v \in V_+(D)$, define $u \cdot v$ as follows:

$$iuvj \text{ if } iuj \text{ or } ivj$$

Let $V_*(D) = \{uv \,|\, u \in V_-(D^u), v \in V_+(D^v)\}$. It is easy to verify that $uv \in V$ and $E^{uv} = E^u$ and $D^{uv} = D^v$. It follows that $uv \notin V_-(D^u) \cup V_+(D^v)$. Also, note that if $w \in V_*(D)$, then there is a unique pair $u \in V_-(D^u), v \in V_+(D^u)$ such that $w = uv$. Let $W(D) = V \backslash (V^*(D) \cup V_-(D) \cup V_+(D))$. Define ϕ_o as follows:

$$\phi_o(w) = \begin{cases} 0 & \text{if } w \in V_-(D) \cup V_+(D) \\ \phi(w) + \phi(u) \cdot \phi(v)/\gamma & \text{if } w = uv \in V_*(D) \\ \phi(w) & \text{if } w \in W(D) \end{cases}$$

We claim that $\phi_o \in \Phi(D,\phi)$. First, we will show that $F_{\phi_o} = F_\phi$. Let $V_-^D(A,B) = V_-(D) \cap V(A,B)$, $V_+^D(A,B) = V_+(D) \cap V(A,B)$, $V_*^D(A,B) = V_*(D) \cap V(A,B)$ and $W^D(A,B) = W(D) \cap V(A,B)$. Take $(A,B) \in \mathcal{N}$ such that $B \cap [N \backslash D] \neq \emptyset$. Then,

$$\sum_{w \in V(A,B)} \phi(w) = \sum_{w \in V_-^D(A,B)} \phi(w) + \sum_{w \in V_*^D(A,B)} \phi(w) + \sum_{w \in W^D(A,B)} \phi(w) \text{ and}$$
$$\sum_{w \in V(A,B)} \phi_o(w) = \sum_{w \in V_*^D(A,B)} \phi_o(w) + \sum_{w \in W^D(A,B)} \phi_o(w)$$

But $\sum_{w \in W^D(A,B)} \phi(w) = \sum_{w \in W^D(A,B)} \phi_o(w)$ by construction and

$$\sum_{w \in V_*^D(A,B)} \phi_o(w) = \sum_{w \in V_*^D(A,B)} \phi(D) + \sum_{u \in V_-^D(A,B)} \sum_{v \in V_+^D(A,B)} \phi(u) \cdot \phi(v)/\gamma \qquad (2)$$

Note that

$$\sum_{u\in V_-^D(A,B)}\sum_{v\in V_+^D(A,B)}\phi(u)\cdot\phi(v)/\gamma=\sum_{u\in V_-^D(A,B)}\phi(u)\cdot\left(\sum_{v\in V_+(D)}\phi(v)\right)/\gamma=\sum_{u\in V_-D(A,B)}\phi(u)$$

Hence, $\sum_{w\in V(A,B)}\phi_o(w)=\sum_{w\in V(A,B)}\phi(w)$. Next, take $(A,B)\in\mathcal{N}$ such that $B\subset D$. Then,

$$\sum_{w\in V(A,B)}\phi(w)=\sum_{w\in V_+^D(A,B)}\phi(w)+\sum_{w\in V_*^D(A,B)}\phi(w)+\sum_{w\in W^D(A,B)}\phi(w)\text{ and}$$
$$\sum_{w\in V(A,B)}\phi_o(w)=\sum_{w\in V_*^D(A,B)}\phi_o(w)+\sum_{w\in W^D(A,B)}\phi_o(w)$$

Again, $\sum_{w\in W^D(A,B)}\phi(w)=\sum_{w\in W^D(A,B)}\phi_o(w)$ by construction and

$$\sum_{w\in V_*^D(A,B)}\phi_o(w)=\sum_{w\in V_*^D(A,B)}\phi(D)+\sum_{u\in V_-^D(A,B)}\sum_{v\in V_+^D(A,B)}\phi(u)\cdot\phi(v)/\gamma$$

Note that

$$\sum_{u\in V_-^D(A,B)}\sum_{v\in V_+^D(A,B)}\phi(u)\cdot\phi(v)/\gamma=\sum_{v\in V_+^D(A,B)}\phi(v)\cdot\left(\sum_{u\in V_-(D)}\phi(u)\right)/\gamma=\sum_{u\in V_+^D(A,B)}\phi(u)$$

Hence, equation (1) yields $\sum_{w\in V(A,B)}\phi_o(w)=\sum_{w\in V(A,B)}\phi(w)$ and we have shown that $F_{\phi_o}=F_\phi$. In particular, $F_{\phi_o}(B,B)=F_\phi(B,B)$ and since $\phi\in\Phi$, we conclude that $\phi_o\in\Phi$. Clearly, $\phi(u)=0$ for all $u\in V_-(D)\cup V_+(D)$ and hence $\phi_o\in\Phi_o(D,\phi)$.

To conclude the proof, we will verify that $\sum_{w\in V_-(D')}\phi_o(w)=\sum_{w\in V_-(D')}\phi_o(w)$ and $\sum_{w\in V_+(D')}\phi_o(w)=\sum_{w\in V_+(D')}\phi_o(w)$ for all $D'\in\mathcal{D}\backslash\{D\}$. Note that for $D'\in\mathcal{D}\backslash\{D\}$ such that $D'\cap(N\backslash D)\neq\emptyset$, $V_-(D')\subset W(D)$ and since $\phi_o(w)=\phi(w)$ for all $w\in W(D)$, we have $\sum_{w\in V_-(D')}\phi_o(w)=\sum_{w\in V_-(D')}\phi(w)$. For $D'\in\mathcal{D}$ such that $D\neq D'\subset D$, $V_-(D)\cap V_-(D')=\emptyset$ but $V_-(D)\cap V_+(D')$ and $V_*(D)\cap V_-(D')$ may be nonempty. In particular, $uv\in V_*(D)\cap V_-(D')$ if and only if $u\in V_-(D)$ and $v\in V_+(D)\cap V_-(D')$. Hence,

$$\sum_{w\in V_-(D')}\phi_o(w)=\sum_{w\in V_*(D)\cap V_-(D')}\phi(w)+\sum_{u\in V_-(D)}\sum_{v\in V_-(D)\cap V(D')}\phi(u)\phi(v)/\gamma$$
$$+\sum_{w\in W(D)\cap V_-(D')}\phi(w)$$

Then,

$$\sum_{u \in V_-(D)} \sum_{v \in V_+(D) \cap V_-(D')} \phi(u)\phi(v)/\gamma = \sum_{v \in V_+(D) \cap V_-(D')} \phi(v) \left(\sum_{u \in V_-(D)} \phi(u) \right) /\gamma$$
$$= \sum_{v \in V_+(D) \cap V_-(D')} \phi(v)$$

The last two display equations yield

$$\sum_{w \in V_-(D')} \phi_o(w) = \sum_{w \in V_*(D) \cap V_-(D')} \phi(w) + \sum_{v \in V_+(D) \cap V_-(D')} \phi(v)$$
$$+ \sum_{w \in W(D) \cap V_-(D')} \phi(w)$$

and since $V_-(D) \cap V_-(D') = \emptyset$, we conclude that $\sum_{w \in V_-(D')} \phi_o(w) = \sum_{w \in V_-(D')} \phi(w)$ as desired.

To prove that $\sum_{w \in V_+(D')} \phi_o(w) = \sum_{w \in V_+(D')} \phi(w)$, we again consider two cases, $D \cap (N \backslash D') \neq \emptyset$ or $D' \neq D \subset D$. In the former case, $V_+(D') \subset W(D)$ and since $\phi_o(w) = \phi(w)$ for all $w \in W(D)$, we have $\sum_{w \in V_-(D')} \phi_o(w) = \sum_{w \in V_-(D')} \phi(w)$. In the latter case, $V_+(D) \cap V_+(D') = \emptyset$ but $V_-(D) \cap V_+(D')$ and $V_*(D) \cap V_+(D')$ may be nonempty. In particular, $uv \in V_*(D) \cap V_+(D')$ if and only if $u \in V_-(D) \cap V_+(D')$ and $v \in V_+(D)$. Hence,

$$\sum_{w \in V_-(D')} \phi_o(w) = \sum_{w \in V_*(D) \cap V_-(D')} \phi(w) + \sum_{u \in V_-(D) \cap V_+(D')} \sum_{v \in V_+(D)} \phi(u)\phi(v)/\gamma$$
$$+ \sum_{w \in W(D) \cap V_-(D')} \phi(w)$$

Then,

$$\sum_{u \in V_-(D) \cap V_+(D')} \sum_{v \in V_+(D)} \phi(u)\phi(v)/\gamma = \sum_{v \in V_-(D) \cap V_+(D')} \phi(u) \left(\sum_{v \in V_+(D)} \phi(v) \right) /\gamma$$
$$= \sum_{u \in V_-(D) \cap V_+(D')} \phi(u)$$

The last two display equations yield

$$\sum_{w\in V_+(D')} \phi_o(w) = \sum_{w\in V_*(D)\cap V_+(D')} \phi(w) + \sum_{u\in V_-(D)\cap V_+(D')} \phi(u) + \sum_{w\in W(D)\cap V_+(D')} \phi(w)$$

and since $V_+(D)\cap V_+(D') = \emptyset$, we conclude that $\sum_{w\in V_+(D')} \phi_o(w) = \sum_{w\in V_+(D')} \phi(w)$ as desired.

□

Lemma A4: *For $F \in \mathcal{F}$ and $D \in \mathcal{D}$, we have $\sum_{x\in X_-(D)} F''(x) = \sum_{x\in X_+(D)} F''(x)$.*

Proof: For all $D \in \mathcal{D}$, let $\mathcal{N}^0(D) = \{B \in \mathcal{N} \mid D \subset B \text{ and } |B\backslash D| \text{ is even}\}$ and $\mathcal{N}^1(D) = \{B \in \mathcal{N} \mid D \subset B \text{ and } |B\backslash D| \text{ is odd}\}$. Note that $D \in \mathcal{N}^0$; that is, 0 is even and that $N\backslash B$ has exactly $2^{|N\backslash B|}$ subsets half of which have an even number of elements while the other half has an odd number of elements. It follows that

$$\alpha := \sum_{B\in\mathcal{N}^0(D)} F(B,B) = 2^{|N\backslash B|-1} = \sum_{B\in\mathcal{N}^1(D)} F(B,B) := \beta$$

By Lemma A1(ii), $F(B,B) = \sum_{x\in Z(B)} F''(x)$. Then, note that for each $x \in X_+(D)$, $F''(x)$ appears exactly once in the summation that defines α and does not appear at all in the summation that defines β. Conversely, for any $(A,D) \in X_-(D)$, $F''(A,D)$ appears exactly $2^{|A|-1}$ times in the summation that defines β and $2^{|A|-1} - 1$ times in the summation that defines α (since there is no element of $X_-(D)$ corresponding to $(\emptyset, D)$). Hence, cancelling the common terms in $\alpha = \beta$ yields $\sum_{x\in X_-(D)} F(x) = \sum_{x\in X_+(D)} F(x)$ as desired. □

Lemma A5: $\Delta_2\Delta_1\rho_u(x) = 1$ *if* $u = u_x$ *and* $\rho_u(x) = 0$ *otherwise.*

Proof: Let $G(x) = 1$ if $u \in u_x$ and $G(x) = 0$ otherwise. Then, let F' be the unique function such that $G = \Delta_2 F'$. Suppose $\Delta_1\rho_u(A,B) = 1$. Pick $i \in A$ and let $D = \{j \in N\backslash B \mid u(j) = u(i)\}$. Set $x = (A\cup D, B\cup D)$ and note that $u \in U^x$. Hence, $G(x) = 1$ and therefore $F'(A,B) \geq \Delta_1\rho_u(A,B)$ for all $(A,B) \in X$. Next, assume $G(x) > \Delta_2\Delta_1\rho_u(x)$ for some $x \in X$. Then, we have either (1) $\Delta_1\rho_u(A,B) = 0$ and $G(A',B') = 1$ for (A',B') such that $(A,B) \succeq_2 (A',B')$ or (2) $\Delta_1\rho_u(A,B) = 1$ and $G(A',B') = G(A'',B'') = 1$ for $(A',B') \neq (A'',B'')$ such that $(A,B) \succeq_2 (A',B')$ and $(A,B) \succeq_2 (A'',B'')$.

In case (1), we have either $A \neq c_u(B)$; that is, (a) there exist $j \in A$, $i \in B$ such that $u(i) > u(j)$ or (b) there exists $j \in N \backslash B$ and $i \in A$ such that $u(i) > u(j)$ *and* $D \subset N \backslash B$ such that $G(A \cup D, B \cup D) = 1$. If (a) were true, then there exists $j \in A \cup D$, $i \in B \cup D$ such that $u(i) > u(j)$ which implies $u \notin U^x$ for $x = (A \cup D, B \cup D)$, a contradiction. If (b) were true, then either there is $j \in N \backslash (B \cup D)$ and $i \in A \cup D$ such that $u(i) > u(j)$ or there is $j, i \in A \cup D$ such that $u(i) > u(j)$; in either case, $u \notin U^y$ for $x = (A \cup D, B \cup D)$, a contradiction.

Finally, note that in case (2), if $\Delta_1 \rho_u(x) = 1$ and $G(y) = 1$ for some y such that $x \succeq_2 y$, then we must have $x = y$. It follows that $(A', B') = (A'', B'')$, a contradiction. Hence, we have shown that $G = \Delta_2 \Delta_1 \rho_u$ and proven the lemma. □

4.1 Proof of the Theorem

Let $\mathcal{D}(\phi) = \{D \in \mathcal{D} \,|\, \phi(v) > 0 \text{ for some } v \in V_-(D)\}$. For any $F \in \mathcal{F}$, let $\phi^1 = \phi_F$. Then, let $\pi = \phi^1$ if $\mathcal{D}(\phi^1) = \emptyset$. Otherwise, choose $D^1 \in \mathcal{D}(\phi^1)$ and $\phi^2 \in \Phi(D^1, \phi^1)$. By Lemma A3, this can be done. Again, let $\pi = \phi^2$ if $\mathcal{D}(\phi^2) = \emptyset$; otherwise, choose $D^2 \in \mathcal{D}(\phi^2)$ and $\phi^3 \in \Phi(D, \phi^2)$. Since $\mathcal{D}$ is finite, the process must end so that we find $\pi = \phi^k$ for some finite k such $\mathcal{D}(\phi^k) = \emptyset$. It follows that π is a random utility. Since $\phi^{l+1} \in \Phi(D^l, \phi^l)$ for all $l = 1, \dots, k-1$ and $\phi^1 = \phi_F$, Lemma A3 ensures that

$$\sum_{y \in U(x)} \pi(x) = \sum_{y \in V(x)} \phi^k(x) = \sum_{y \in V(x)} \phi^1(x)$$

But Lemma A2(iii) yields $\sum_{y \in V(x)} \phi^1(x) = F(x)$ proving that F maximizes π.

For the converse, let $F_\pi = \sum_{u \in U} \rho_u$ and hence $F''_\pi = \sum_{u \in U} \pi(u) \cdot \Delta_2 \Delta_1 \rho_u$. By Lemma A5, $\Delta_2 \Delta_1 \rho_u(x) = 0$ or 1 and hence $F'' \geq 0$. Since $\rho_u(B, B) = 1$ for all $B \in \mathcal{N}$, we also have $F_\pi \in \mathcal{F}$. □

4.2 Proof of Lemma 1

We will show that if c is almost rational, then $\Delta_1 \rho_c \geq 0$. Define f as follows:

$$f(A, B) = \begin{cases} 1 & \text{if } c_u(B) = A \text{ and } B' \neq B \subset B' \text{ implies } c(B') \neq A \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $f \geq 0$ and $f(x) \in \{0, 1\}$ for all $x \in X$. To complete the proof, we will show that $\rho_c(A, B) = \sum_{y \in [x]_1} f(y)$. First, we will show that $\rho_c(A, B) = 0$, $A' \subset A$ and $B \subset B'$

implies $\rho_c(A', B') = 0$. If $\rho_c(A, B) = 0$, $A' \subset A$, then the rationality of ρ_c ensures that $\rho_c(A', B) = 0$. On the other hand, since $A' \subset A \subset B \subset B'$, (w_-) and $\rho_c(A', B') = 1$ would imply $\rho_c(A', B) = 1$. Since ρ_c is extreme, we conclude that $\rho_c(A', B') = 0$.

Suppose $\rho_c(x) = 0$. Then, the observation above ensures $f(y) = 0$ for all $y \in [x]_1$ and hence $\rho_c(x) = \sum_{y \in [x]_1} f(y)$. Next, assume $\rho_c(A, B) = 1$. Hence, $c(B) \subset A$ and $\rho_c(c(B), B) = 1$. Let $\mathcal{B} = \{B' \in \mathcal{N} \,|\, c(B') \subset B\}$. It is easy to verify that (w_-) implies $c(B') = c(B)$ for all $B' \in \mathcal{B}$. It is also easy to check that $B', \hat{B} \in \mathcal{B}$ implies $B' \cup \hat{B} \in \mathcal{B}$. Hence, there exists $B^* \in \mathcal{B}$ such that $B' \in B^*$ if and only if $B \subset B' \subset B^*$. Clearly, $\rho_c(c(B), B') = 1$ if and only if $c(B) \subset B' \subset B^*$. By construction, $f(c(B), B^*) = 1$ and $y \neq x := (c(B), B^*)$, $y \in [x]_1$ implies $f(y) = 0$. Then, $\rho_c(x) = \sum_{y \in [x]_1} f(y)$.

Next, we will show that if $f = \Delta_1 \rho \geq 0$, then $\rho = \rho_c$ for some almost rational c. Define $c(B) = A$ for $A \in \mathcal{N}$ such that $A \subset B$, $f(A, B') = 1$ for $B' \in \mathcal{N}$ satisfying $B \subset B'$. Note that $\rho(x) = \sum_{y \in [x]_1} f(y)$ and $f \geq 0$ imply that for all (A, B), there exists a unique $(\hat{A}, \hat{B})$ such that $\hat{A} \subset A, B \subset \hat{B}$ and $f(\hat{A}, \hat{B}) = 1$. Thus, the choice function above is well-defined. Then, $\rho = \rho_c$ is immediate.

That $\rho = \rho_u$ for some u implies $\Delta_2 \Delta_1 \rho \geq 0$ follows immediately from Lemma A5. To conclude the proof of the lemma, we will show that if ρ is extreme and $\Delta_2 \Delta_1 \rho \geq 0$, then there exists $u \in U$ such that $\rho = \rho_u$. If $\Delta_2 \Delta_1 \rho \geq 0$, then $\rho = \sum_{u \in U} \pi(u) \rho_u$. Since ρ is extreme, we must have $\pi(u) = 1$ for some $u \in U$; that is, $\rho = \rho_u$. □

4.3 Proof of Lemma 2

Assume ρ is an additive RCF and let

$$\rho^i(A, B) = \begin{cases} \rho(i, B) & \text{if } i \in A \\ 0 & \text{otherwise} \end{cases}$$

Then, $\rho = \sum_{i \in N} \rho^i$. The linearity of Δ_1 ensures $\Delta_1 \rho = \sum_{i \in N} \Delta_1 \rho^i$. Hence,

$$\Delta_1 \rho(A, B) = \sum_{i \in N} \Delta_1 \rho^i(A, B)$$

If $i \notin A$, then $\rho^i(A, B) = 0$ and hence $\Delta_1 \rho^i(A, B) = 0$ whenever $i \notin A$ and $\rho^i(A, B) = \rho^i(A', B)$ whenever $i \in A \subset A'$. Hence,

$$\Delta_1 \rho^i(A, B) = \begin{cases} \rho^i(i, B) & \text{if } A = \{i\} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

It follows that

$$\Delta_1\rho(A,B) = \sum_{i\in A} \Delta_1\rho^i(i,B) = \sum_{i\in A} \Delta_1\rho(i,B) \tag{2}$$

This proves the first assertion of the Lemma. To prove the second, assume $A \neq \{i\}$. Then, $(A,B) \succeq_2 (A',B')$ implies $A' \neq \{i\}$ and hence equation (1) implies $\Delta_1\rho^i(A',B') = 0$ for all such (A',B'). Then, invoking equation (1) yields $\Delta_2\Delta_1\rho^i(i,B) = \Delta_1\rho^i(i,B)$. Since $\rho = \sum_{i\in N}\rho^i$ we have $\Delta_2\Delta_1\rho = \sum_{i\in N}\Delta_1\Delta_2\rho^i$ and hence the desired result follows. □

4.4 Proof of Proposition

Let $N = \{1,2,3,4,5\}$. Define F as follows: $F'(N,N) = 1$, $F'(B,B) = 1$ for all four-element sets B. Let $\mathcal{B} = \{\{1,2,5\},\{1,2,3\},\{2,3,4\},\{3,4,5\},\{4,5,1\}\}$; $F'(B,B) = 1$ for all three-element sets not in $\mathcal{B}$. For $B \in \mathcal{B}$, $F'(B,B) = 1/2$. Moreover,

$$1/2 = F'(1,\{5,1,2\}) = F'(2,\{1,2,3\}) = F'(3,\{2,3,4\}) = F'(4,\{3,4,5\}) = F'(5,\{4,5,1\})$$

Finally, $F'(x) = 0$ for all x not listed above.

There exists a unique function F such that $F' = \Delta_1 F$. Since $F' \geq 0$, $F \geq 0$. It is easy to verify that $F(B,B) = 1$ for all $B \in \mathcal{N}$. Hence, F is an RCF and therefore $F \in \mathcal{F}$. To conclude the proof we will verify that $F \neq \sum_{c\in C}\nu(c)\cdot\rho_c$ for any probability ν.

Suppose there exists such a probability. Then, $F' = \sum_{c\in C}\nu(c)\Delta_1\rho_c$. Hence, there exists $c \in C$ such that $\Delta_1\rho_c(1,\{512\}) = 1$. Since $1 = \rho_c(\{1,2\},\{1,2\})$, this implies that $\Delta_1\rho_c(2,\{1,2,3\}) = 0$. But since $1 = \rho_c(\{2,3\},\{2,3\})$, we must have $\Delta_1\rho_c(3,\{2,3,4\}) = 1$. Continuing in this fashion, we get $\Delta_1\rho_c(4,\{3,4,5\}) = 0$ and $\Delta_1\rho_c(5,\{4,5,1\}) = 1$. Finally, $\Delta_1\rho_c(5,\{4,5,1\}) = 1$ implies $\rho(5,\{1,5\} = 1$ but $\Delta_1\rho_c(1,\{5,1,2\}) = 1$ implies $\rho_c(1,\{1,5\} = 1$; since $\rho_c(\{1,5\},\{1,5\} = 1$ this is a contradiction. □

References

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

Colonius, H. (1984). Stochastische Theorien individuellen Wahlverhaltens. Berlin: Springer.

Debreu, G., (1960), "Review of Individual choice behaviour by R. D. Luce," *American Economics Review*, 50: 186–188.

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

Fechner, G.T. (1860) Elemente der Psychophysik. Breitkopf and Hartel, Leipzig (reprinted in 1964 by Bonset, Amsterdam); English translation by H.E. Adler (1966): Elements of psychophysics. Holt, Rinehart and Winston, New York.

Gul, F, Natenzon, P. and W. Pesendorfer, (2013) "Random Choice as Behavioral Optimization," working paper.

Gul, F and W. Pesendorfer, (2006) "Random Expected Utility," Econometrica, 74(1), 121–146.

Houthakker, H. (1950), "Revealed preference and the utility function," *Economica,* 17, 159–174

Lu, J. (2013) "Random Choice and Information," working paper.

Luce, R.D., (1959), Individual choice behavior: a theoretical analysis. , Wiley, New York.

Luce, R. D. (1977), "The Choice Axiom after Twenty Years," *Journal of Mathematical Psychology*, 15: 215–233.

Masatlioglu, Y., D. Nakajima and E. Ozbay (2012), "Revealed Attention," *American Economic Review,* 102: 2183-2205.

McFadden, D., (1978) *Modelling the choice of residential location.* In: S. Karlqvist, L. Lundquist, F. Snickers and J. Weibull, Eds., Spatial Interaction Theory and Planning Models, North-Holland, Amsterdam (1978).

Sen, A. (1970) Collective Choice and Social Welfare, Holden-Day, San Francisco.

Thurstone, L. L. (1927) "A law of comparative judgment," *Psychological Review*, 34, (4), 273–286.