Marginal stochastic choice

Yaron Azrieli and John Rehbeck*

July 11, 2025

Abstract

Models of stochastic choice typically use conditional choice probabilities given menus as the primitive for analysis, but in the field these are often hard to observe. We consider the case where an analyst has access to a *marginal stochastic choice dataset* containing the marginal distributions of available menus and of choices, but not to conditional choice frequencies. The Random Utility Model (RUM) has no testable implications for such datasets, but any restriction on the domain of feasible preference orders does limit the set of rationalizable marginals. The Luce model can also rationalize essentially any dataset, but unlike RUM its parameters can often be identified. We also demonstrate that additional testable implications for the marginals may arise when the distribution of menus is endogenous.

*Department of Economics, The Ohio State University, 1945 North High street, Columbus, OH 43210. azrieli.2@osu.edu and rehbeck.7@osu.edu. We thank participants of several conferences where this work has been presented for their comments and Julie Holland Mortimer for very useful references. We are especially grateful to three anonymous referees of *Theoretical Economics* and to the Coeditor Todd Sarver for their constructive suggestions. John Rehbeck is supported in part by the National Science Foundation grant NSF-2049749.

1 Introduction

The vast majority of theory papers on stochastic choice assume that the data available to the observer contains choice frequencies conditional on a wide collection of choice sets (menus). Indeed, the early works of Luce [1959], Block and Marschak [1960], Falmagne [1978], Barberá and Pattanaik [1986], and McFadden and Richter [1990] demonstrated the potential of this framework to deliver elegant and intuitive characterizations of choice rules, and subsequent works followed this tradition.[1] In this paper we take a different approach and assume that the researcher has access to a *marginal stochastic choice dataset*. Such datasets consist of a pair of distributions, one over menus (denoted μ) and one over alternatives (denoted λ). Here, $\mu(A)$ is the share of choices made in the Ath menu, while $\lambda(a)$ is the aggregate frequency with which the ath alternative is chosen. We revisit some of the prominent models in the stochastic choice literature that were characterized using conditional choice frequencies, and study their testable implication and identification properties with this new type of data.

Our motivation to consider such datasets stems from the fact that they are closer to the type of datasets usually available in empirical applications. Indeed, it is often hard to obtain data about the set of alternatives that were available or considered by an individual at the time of making their choice.[2] Nevertheless, it is sometimes possible to separately collect data on overall availability rates (μ), which when combined with aggregate choice data (λ) yields a marginal stochastic choice dataset.

As a concrete example, consider the case of grocery products. Variation in assortment across stores, as well as stock-out events, lead to significant menu variation for

[1]More recent papers that use conditional choice probabilities to characterize various decision models include Gul and Pesendorfer [2006], Manzini and Mariotti [2014], Fudenberg and Strzalecki [2015], Fudenberg et al. [2015], Brady and Rehbeck [2016], Aguiar [2017], Apesteguia et al. [2017], Kitamura and Stoye [2018], Frick et al. [2019], Cattaneo et al. [2020], Cattaneo et al. [2021], and Kovach and Tserenjigmid [2022]. See Strzalecki [2025] for a comprehensive survey of the stochastic choice literature.

[2]To quote Manski [1977, page 239], "Current methods for estimating the parameters of random utility functions require ex post observation of a sequence of choice problems for each of which the decision maker, choice set, and chosen alternative are known. Often, however, the survey instrument used in estimation supplies the identities of the decision maker and his chosen alternative but not those of his feasible inferior alternatives."

consumers.[3] However, trying to figure out which menu was available at a particular purchase is challenging, for instance because items in inventory are sometimes not shelved.[4] It is therefore typically impossible to observe choice frequencies of products conditional on the available menu, and indeed the literature for the most part resorts to assuming that there is no menu variation at all [Hausman et al., 1994, Nevo, 2001, Hausman and Leonard, 2002].[5] Still, overall availability rates of grocery products may be observed: Matsa [2011] obtains detailed stock-out rates from data collected by the U.S. Bureau of Labor Statistics used for constructing the consumer price index. This gives a measure of the frequency with which consumers face any given menu.[6]

We say that the marginal stochastic choice dataset (μ, λ) is rationalizable with a given model of stochastic choice if there are conditional choice probabilities $\{\pi(a|A)\}$ consistent with the model such that

$$\lambda(a) = \sum_{A: a \in A} \mu(A)\pi(a|A) \tag{$*$}$$

holds for each alternative a. Equation $(*)$ means that the available dateset (μ, λ) could have been generated by the given stochastic choice model. As a benchmark, we start with the case where no restrictions are imposed on π, and show that (μ, λ) is feasible if and only if

$$\sum_{a \in A} \lambda(a) \geq \sum_{B \subseteq A} \mu(B) \tag{$**$}$$

for every menu A. The inequality $(**)$ is a variant of the condition in the classic 'marriage

[3]Gruen et al. [2002] estimates that overall stock-out rates are about 8%. See Hickman and Mortimer [2016, Section 2.1] for a discussion of the prevalence of assortment variation.

[4]Thus, scanner data might indicate a store has the item although it is not immediately available to the consumer. Gruen et al. [2002] gives additional examples and details on the difficulty of tracking stock-outs.

[5]Tenn and Yun [2008] show that not accounting for availability through stock-outs can significantly bias elasticity estimates.

[6]Another example of a dataset that fits our framework is Bruno and Vilcassim [2008] which studies demand for confectionary chocolate in the UK using data on sales and on how widely each product is distributed. To obtain the distribution over menus μ they assume that availability probabilities are independent across products.

lemma' [Hall, 1934] and simply means that the frequency with which alternatives in A are chosen (the left-hand side) must be at least as large as the frequency with which only alternatives in A are available (the right-hand side).

We proceed by imposing more structure on the decision process. We find that assuming choices are made according to the Random Utility Model (RUM) [Block and Marschak, 1960, Falmagne, 1978] places no additional restrictions on marginal stochastic choice data beyond the collection of inequalities (∗∗). This is in contrast to the case in which conditional frequencies are observable, where RUM does have testable implications. Nevertheless, we prove that under a mild assumption on the support of μ, any restriction of the domain of preference orders does lead to non-trivial testable implications; we illustrate this with a couple of the classic domain restrictions in the literature. Taken together, these results highlight the importance of a-priori knowledge that the researcher has and the assumptions they are willing to make about preferences when only marginal datasets are available.

Next, we find that even the strong assumption of having choices generated by the Luce model [Luce, 1959] essentially places no additional restrictions on marginal stochastic choice data. Since Luce is a special case of many stochastic choice models, this result emphasizes the difficulty of testing such models using only the marginals. However, unlike RUM, under Luce the model's parameters are often identified – we provide a necessary and sufficient condition on the distribution of menus μ for this to be the case. Thus, with the Luce model the marginal stochastic choice dataset (μ, λ) often pins down the conditional choice frequencies given every menu. We note that our proof of this result is not constructive, and there may be non-trivial computational problems in the identification process.

In the models described so far we viewed the distribution of menu availability μ as exogenously given.[7] But in practice agents often choose the menu themselves before

[7] This distribution is clearly an important part of the dataset because it affects what distributions of choice λ can emerge under a given model of behavior, but up to this point μ was not influenced by or influencing how agents choose.

choosing an alternative from the menu, e.g., consumers may choose a grocery store partly based on the products available at that store. When this is the case, the distribution of menus μ reflects maximization behavior, and we can therefore expect that not every μ can be rationalized. Furthermore, the choice of a menu not only determines what alternatives are available in the second period, but is also informative about the preferences over alternatives, and therefore about the alternative that the agent eventually chooses. In the last part of the paper we illustrate the additional restrictions that may arise with endogenous menu choice by considering the temptation and self-control model of Gul and Pesendorfer [2001]. It turns out that this model places intuitive restrictions on marginal stochastic choice data: An agent would never choose a menu A and then an alternative a from that menu if there is a feasible sub-menu of A that also contains a; this property translates into a restriction on the support of μ, as well as a strengthening of the inequalities in $(**)$.

From a technical point of view our analysis heavily relies on classic results and ideas from the theory of transferable utility cooperative games. The connection is that instead of describing the distribution of availability by its probability mass function μ, we can describe it by its cumulative distribution function $v_\mu(A) = \sum_{B \subseteq A} \mu(B)$. The inequalities $(**)$ can then be understood as requiring that λ is in the core of the game v_μ. We leverage known results about the core to derive several of our characterizations. While some connections between cooperative games and stochastic choice have already been pointed out in the literature,[8] they all concern collections of choice probabilities conditional on menus. We hope that the new connections we uncover here will be useful in future work on stochastic choice.

1.1 Literature review

There is a long history of empirical and econometrics papers trying to deal with the problem of unobservable choice sets. This problem is discussed in length in Manski

[8]See for example Monderer [1992], Gilboa and Monderer [1992], and Billot and Thisse [2005].

[1977], while some examples of papers on related issues include Swait and Ben-Akiva [1986], Horowitz [1991], Ben-Akiva and Boccara [1995], Tenn and Yun [2008], Tenn [2009], and Abaluck and Adams-Prassl [2021]. These papers often study identification of conditional probabilities of choosing alternative a from menu A ($\pi(a \mid A)$) that are consistent with marginal choices as in $(*)$. Some of these papers allow for menus to be at least partly chosen by the decision makers, similar to the setup we consider in Section 6. Recently, Barseghyan et al. [2021] study a random utility model where the distribution of availability is not observed but there is a known lower bound on the size of menus. Their characterization of the sharp identification region in Theorem 3.1 has some similarities to our Proposition 1. See also Lu [2022] which analyzes a similar problem in the case where there are both an upper and lower bounds on the feasible menu.

As pointed out above, many of our results build on existing work in cooperative game theory. Specifically, the characterization of RUM rationalizability is based on the characterization of the extreme points of the core of a convex game due to Shapley [1971], and can also be obtained from results in Weber [1988]. See also Ichiishi [1981] for related results. For the Luce model, our proof follows the footsteps of the proof in Monderer et al. [1992] who showed that the core is homeomorphic to the set of weighted Shapley values. Kalai and Samet [1987] study weighted Shapley values axiomatically. Recently, Doval and Eilat [2023] use similar techniques to study the set of feasible marginal distributions of actions (averaged across states of the world) taken by an agent when the analyst can't observe the information that was available at the time of decision making.

Finally, this paper also contributes to the emerging literature taking insights from stochastic choice to new settings and datasets. The recent work of Dardanoni et al. [2020] is motivated by similar considerations to the current paper, namely, that theoretical models of stochastic choice should be based on datasets more likely to be available in the field. In their model agents are heterogeneous in their cognitive ability which determines the number of alternatives they can consider. The researcher observes the aggregate choice distribution (λ of the current paper) and the question is whether the distribution of cognitive types can be identified. Thus, roughly speaking, the main difference from our work is that we assume that the distribution of menus is observable and study conditional choices, while they study whether the menu distribution can be inferred from aggregate choices. Chambers et al. [2024] study correlated preferences within the random utility framework, where the dataset assumed to contain the frequencies of tuples of choices made by a group of agents conditional on the menu that was available to each one of them. Manzini et al. [2019] studies whether or not an individual chooses to approve an option from a list. This stochastic dataset is novel since the sum of approving different options can be more than one, and since choice probabilities vary with the order in which options are presented. In Cheung and Masatlioglu [2021] decision makers obtain a recommended set of alternatives before choosing, but are free to choose any alternative (even outside of the recommended menu).

2 Preliminaries

For any finite set Y we denote by $\Delta(Y)$ the set of probability distributions on Y. If $y \in Y$ and $p \in \Delta(Y)$, then we write $p(y)$ instead of $p(\{y\})$. We sometimes identify $\Delta(Y)$ with the standard simplex in $\mathbb{R}^Y$, i.e. with the set of $p \in \mathbb{R}^Y$ such that $p(y) \geq 0$ for every y and $\sum_y p(y) = 1$. We use $Int(D)$ to denote the relative interior of a set $D \subseteq \mathbb{R}^Y$, where $\mathbb{R}^Y$ is endowed with its standard topology. In particular, $Int(\Delta(Y))$ is the set of $p \in \Delta(Y)$ such that $p(y) > 0$ for all y.

Throughout the paper we denote by X the finite set of alternatives and by $\mathcal{X} = 2^X \setminus \emptyset$ the collection of all non-empty subsets of X. Alternatives in X are typically denoted by $a, b, \ldots$, while elements of $\mathcal{X}$ are called menus and are typically denoted by $A, B, \ldots$. The cardinality of X is $|X| = n$. If $A \in \mathcal{X}$ then $A^c = X \setminus A$ is the complement of A.

In this paper, we study what can be learned from the *marginal stochastic choice dataset* $(\mu, \lambda) \in \Delta(\mathcal{X}) \times \Delta(X)$, where μ is the frequency with which menus are available and λ is the aggregate frequency of choices of alternatives. We sometimes refer to λ

as the marginal distribution over alternatives and μ as the marginal distribution over menus. We highlight that these are not choice frequencies conditional on a given menu as is common in standard models of stochastic choice.

Much of our analysis relies on classic results from the theory of cooperative games, so we now briefly present the essential definitions and results from that theory used in later sections. More details can be found in Grabisch [2016]. At a high-level, we map the observed marginal menu distribution μ to a cooperative game and then use insights from cooperative game theory to characterize the model.

A *cooperative game* is a set function $v : 2^X \to \mathbb{R}$ satisfying $v(\emptyset) = 0$. Throughout the paper we assume that any cooperative game is normalized so that $v(X) = 1$. The core of a game v, denoted $Core(v)$, is the set

$$Core(v) = \left\{ p \in \mathbb{R}^X : \sum_{a \in X} p(a) = 1 \ and \sum_{a \in A} p(a) \geq v(A) \ \ \forall A \in \mathcal{X} \right\}.$$

A game v is *convex* (or super-modular) when $v(A \cup B) + v(A \cap B) \geq v(A) + v(B)$ for all menus $A, B \in \mathcal{X}$. A game is *strictly convex* when the above inequality is strict whenever neither A contains B nor B contains A. Given any real vector $z = (z(B))_{B \in \mathcal{X}}$ with $\sum_{B \in \mathcal{X}} z(B) = 1$ we define the game v_z by $v_z(A) = \sum_{B \subseteq A} z(B)$. Conversely, given a game v, by defining $z_v(B) = \sum_{A \subseteq B} (-1)^{|B \setminus A|} v(A)$ one has that $v(A) = \sum_{B \subseteq A} z_v(B)$ holds for all A, and there is no other vector z with this property. The vector z_v is known as the Möbius transform (or Harsanyi Dividend) of v. A cooperative game v is *totally monotone* when $z_v \geq 0$.[9] It is well-known that any totally monotone game is convex. We omit the proof of the following simple lemma.

Lemma 1. Suppose that v is totally monotone (in particular, convex). Then v is strictly convex if and only if $z_v(\{a,b\}) = v(\{a,b\}) - v(\{a\}) - v(\{b\}) > 0$ for every $a, b \in X$.

[9] Totally monotone games are also known as 'belief functions' in the theory of Dempster [1968a] and Shafer [1976].

3 Consistent marginals

We say that a marginal stochastic choice dataset (μ, λ) is *consistent* when it could have been generated by some stochastic choice function. The formal definition follows.

Definition 1. The marginal stochastic choice dataset $(\mu, \lambda) \in \Delta(\mathcal{X}) \times \Delta(X)$ is consistent when there exists a collection $\pi = (\pi(\cdot|A))_{A \in \mathcal{X}}$ with each $\pi(\cdot|A) \in \Delta(A)$, such that

$$\lambda(a) = \sum_{A: a \in A} \mu(A)\pi(a|A)$$

for every $a \in X$.

Consistency is a minimal feasibility requirement that places no restrictions on the form of the stochastic choice function π. The only constraints are those of availability: If $a \notin A$, then we must have $\pi(a|A) = 0$. An inconsistent dataset suggests issues with the data and our paper is silent on how to proceed in such cases.

It turns out that consistency can be characterized by a simple collection of inequalities reflecting the availability constraints. The following proposition can be deduced from classic results of Strassen [1965], and has already appeared in several previous works (e.g. Chateauneuf and Jaffray [1989, Corollary 3]). It serves as a useful benchmark for later sections where more restrictions are imposed on π.

Proposition 1. For a marginal stochastic choice dataset $(\mu, \lambda) \in \Delta(\mathcal{X}) \times \Delta(X)$, the following are equivalent:

1. (μ, λ) is consistent.
2. For all $A \in \mathcal{X}$, $\sum_{a \in A} \lambda(a) \geq \sum_{B \subseteq A} \mu(B)$.
3. For all $A \in \mathcal{X}$, $\sum_{a \in A} \lambda(a) \leq \sum_{B: B \cap A \neq \emptyset} \mu(B)$.

Condition 2 of the proposition states that the total frequency with which elements of A are chosen must be at least as large as the frequency with which *only* alternatives from A are available. Thus, this condition is clearly necessary for consistency. Similarly, condition 3 states that the frequency of choices from A cannot exceed the frequency that *some* alternative from A is available. These two collections of inequalities are equivalent since the inequality in 2 for a menu A is the same as the inequality in 3 for the menu A^c. The fact that these inequalities are sufficient for consistency can be proven using the max-flow min-cut duality theorem – see Proposition 9 for a more general result.[10]

Recall from the previous section that for a given $\mu \in \Delta(\mathcal{X})$ we define the game v_μ by $v_\mu(A) = \sum_{B \subseteq A} \mu(B)$, and that the core of v_μ is the set of distributions in $\Delta(X)$ that are point-wise above v_μ.[11] In other words, condition 2 of Proposition 1 can be rewritten as $\lambda \in Core(v_\mu)$. We therefore have the following.

Corollary 1. The marginal stochastic choice dataset (μ, λ) is consistent if and only if $\lambda \in Core(v_\mu)$.

Figure 1 illustrates the shape of $Core(v_\mu)$ for a particular distribution μ in the case where X contains three alternatives. By Corollary 1 this is precisely the set of choice distributions λ such that (μ, λ) is consistent.

It is worth pointing out that the stochastic choice function π in Definition 1 is typically not unique whenever the number of alternatives is at least 3. Here is a concrete example.

Example 1. Let $X = \{a, b, c\}$ and suppose that μ assigns probability of $\frac{1}{3}$ to each of the binary menus. Consider a (deterministic) choice π that selects a from $\{a, b\}$, b from $\{b, c\}$, and c from $\{a, c\}$. This π induces the uniform distribution over X. However, π' that chooses the other alternative in each menu also induces the uniform distribution.

In what follows, we impose more structure on the stochastic choice function π by considering some of the prominent models from the stochastic choice literature. We char-

[10]Proposition 1 is a generalized version of the classic 'marriage lemma' [Hall, 1934]. Theorem 3.1 in Barseghyan et al. [2021] also applies a generalization of the marriage lemma to characterize the sharp identification region when the distribution of availability is unobservable but there is a known bound on the minimal size of menus.

[11]Since μ is a probability distribution the game v_μ is non-negative and therefore the core only contains non-negative vectors.

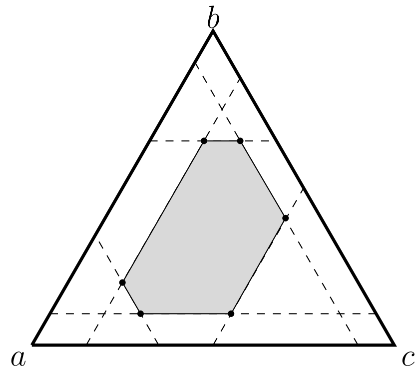

Figure 1: The solid line outer triangle represents the set of choice distributions λ over $X = \{a, b, c\}$. The shaded area is $Core(v_\mu)$ when μ is given by $\mu(a) = 0.1$, $\mu(b) = 0.1$, $\mu(c) = 0.15$, $\mu(\{a, b\}) = 0.3$, $\mu(\{a, c\}) = 0.1$, $\mu(\{b, c\}) = 0.1$, and $\mu(\{a, b, c\}) = 0.15$. Each dashed line corresponds to a constraint $\sum_{a \in A} \lambda(a) \geq v_\mu(A)$ for some menu A.

acterize the testable implications of each model for marginal stochastic choice datasets and study what can be inferred about the model parameters.

4 Random utility

Suppose that each individual in a population has a strict preference ordering over X and chooses their top alternative from the available menu. Formally, let O be the set of strict total orders on X, with typical element $\succ$. Given a menu A and an alternative $a \in A$ denote $T[a, A] = \{\succ \in O \ : \ a \succ b \ \ \forall b \in A \setminus \{a\}\}$ the collection of orders that rank a above any other element in A.

Definition 2. The marginal stochastic choice dataset (μ, λ) is RUM rationalizable when there is a distribution $\nu \in \Delta(O)$ such that

$$\lambda(a) = \sum_{A : a \in A} \mu(A) \nu(T[a, A])$$

for every $a \in X$. When ν satisfies the above equality, we say that ν RUM rationalizes the marginal stochastic choice dataset.[12]

The definition of RUM rationalizability is the same as saying that there exists π as in Definition 1 with $\pi(a|A) = \nu(T[a, A])$. It is well-known [Block and Marschak, 1960, Falmagne, 1978] that not every π can be generated in this way. The novelty of our situation is that we only observe marginal frequencies, so it is unknown whether the random utility model adds any additional restrictions on top of unrestricted consistency. Namely, even when the marginal choice data (μ, λ) can be generated by a stochastic choice function π inconsistent with RUM, it might be possible to find another stochastic choice function π' that is consistent with RUM and rationalizes the same marginals. The next proposition says that this is indeed the case.

Proposition 2. The marginal stochastic choice dataset (μ, λ) is RUM rationalizable if and only if it consistent.

All proofs can be found in the Appendix. Proposition 2 is a simple consequence of the following result of Shapley [1971], which characterizes the extreme points of the core of a convex game.[13] Given $\succ \in O$ and an alternative a, denote by $L_{\succ}(a) = \{b : a \succ b\}$ the lower contour set of a according to $\succ$.

Proposition 3. [Shapley, 1971] Let v be a convex game. Then $p \in \mathbb{R}^X$ is an extreme point of $Core(v)$ if and only if there is $\succ \in O$ such that for every a

$$p(a) = v(L_{\succ}(a) \cup \{a\}) - v(L_{\succ}(a)).$$

Moreover, if v is strictly convex then the mapping $\succ \longrightarrow p$ is one-to-one.

In words, each of the extreme points of $Core(v_\mu)$ is the distribution of choices induced by a homogeneous population of agents all having the same preference order. Now, if

[12]Our definition of RUM rationalization is related to the Random Attention and Utility Model (RAUM) of Kashaev and Aguiar [2022], see Definition 2 of that paper. There are however two important differences: First, in our framework the marginal over menus μ is part of the data, while in their model the distribution of attention is a free parameter. On the other hand, they require that choice data is rationalizable conditional on every menu, while we restrict attention to the grand set of alternatives.

[13]See Grabisch [2016, Remark 3.16] for other papers that independently proved similar results.

(μ, λ) is consistent then λ is in the core and can therefore be represented as a convex combination of its extreme points. Since the mapping that sends each distribution of preferences ν to the resulting λ is linear (for fixed μ), the corresponding distribution of preference orders RUM rationalizes (μ, λ). We note that Proposition 2 can also be deduced from the results of Dempster [1968b] on belief functions and of Weber [1988] on random order values.

4.1 Restricted RUM

While Proposition 2 shows that RUM has no testable implications for the marginals, this is no longer the case whenever the researcher has some a-priori information that restricts the set of possible preference orders in the population. Indeed, given a set of feasible orders $O' \subseteq O$, say that (μ, λ) is O'-RUM rationalizable when there is a distribution $\nu \in \Delta(O')$ such that for every $a \in X$

$$\lambda(a) = \sum_{A: a \in A} \mu(A) \nu(T[a, A]).$$

Then it follows from the above discussion that (μ, λ) is O'-RUM rationalizable if and only if λ is in the convex hull of the extreme points of $Core(v_\mu)$ that correspond to the orders in O'. Applying Lemma 1 then gives the next result.

Corollary 2. Suppose that μ assigns positive probability to each of the binary menus and let O' be a strict subset of O. Then the set of λ's such that (μ, λ) is consistent but not O'-RUM rationalizable is a non-empty and relatively open subset of $Core(v_\mu)$.

To illustrate this result, consider the Single-Peaked Random Utility Model (SPRUM) of Apesteguia et al. [2017]. According to this model, there is an exogenously given order $\succ^* \in O$, and the set of feasible orders $O' \subseteq O$ is the set of single-peaked preferences relative to $\succ^*$.[14] Even with just three alternatives, this domain restriction significantly

[14]The preference $\succ$ is single-peaked relative to $\succ^*$ if there is $a \in X$ such that $a \succ b \succ c$ whenever $a \succ^* b \succ^* c$ and $c \succ b \succ a$ whenever $c \succ^* b \succ^* a$.

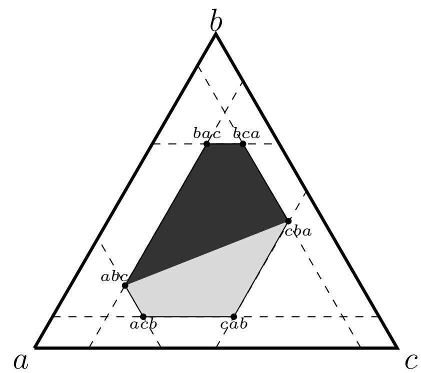

Figure 2: The entire shaded area is $Core(v_\mu)$ for the same μ as in Figure 1. The dark-shaded area is the set of λ's such that (μ, λ) is O'-RUM rationalizable when O' is the domain of single-peaked preferences relative to the order $a \succ^* b \succ^* c$. Any λ in the light-shaded area is consistent but not O'-RUM rationalizable.

reduces the set of rationalizable distributions, as can be seen in Figure 2.[15]

As a second example, suppose that alternatives in X are lotteries over a finite set of prizes, and that the dataset represents the choices of a population of expected utility maximizers as in Gul and Pesendorfer [2006]. If a lottery $a \in X$ is in the interior of the convex-hull of the other lotteries $X \setminus \{a\}$, then none of the decision makers ranks a at the top. This yields a restriction on the domain of possible preference orders, and Corollary 2 then implies that marginal datasets have non-trivial testable implications.

Third, suppose that there is a partition of the set of alternatives into 'quality categories', and it is known that all agents prefer alternatives in higher categories over those in lower ones.[16] For this type of domain restriction we have the following characterization of rationalizability.

Proposition 4. Suppose that μ assigns positive probability to each of the binary menus. Let $(A_1, \ldots, A_K)$ be a partition of X, and let O' be the set of orders $\succ$ such that $a \succ b$

[15] Note that in this example the analyst assumes a specific reference order $\succ^*$. If instead they only assume that preferences are single-peaked with respect to *some* reference order, then we conjecture that any consistent dataset would be rationalizable.

[16] For example, suppose that alternatives are money lotteries and that they can be partitioned into categories such that lotteries in higher categories first-order dominate lotteries in lower categories. Assuming that agents have expected utility preferences generates this kind of domain restriction.

whenever $a \in A_i$, $b \in A_j$, and $i > j$. Then (μ, λ) is O'-RUM rationalizable if and only if it is consistent and $v_\mu\left(\cup_{i=1}^k A_i\right) = \sum_{i=1}^k \sum_{a \in A_i} \lambda(a)$ for all $k = 1, \ldots, K$.

The characterizing equality in the proposition reflects the fact that alternatives in lower quality categories $i = 1, \ldots, k$ are chosen only when none of the alternatives in higher categories is available. Note that this equality means that some of the constraints that define $Core(v_\mu)$ are binding (whenever there are at least two categories), and hence that only λ's on the relative boundary of the core can be rationalized.

4.2 Identification

It is typically impossible to back out the distribution over preferences ν from the marginal stochastic choice dataset (μ, λ). Indeed, it is well-known that point identification fails in the RUM even when conditional choice frequencies are observable.[17] If ν and ν' generate the same stochastic choices in every menu, then clearly they cannot be separated based on the marginals. Of course, when only (μ, λ) is available the situation is even worse, since ν and ν' may not be distinguishable even when they are distinguishable based on conditional choice frequencies. We illustrate this with the following example.

Example 2. Let $X = \{a, b, c\}$ and suppose that μ assign probability of 1/4 to each of the menus $\{a, b\}$, $\{a, c\}$, $\{b, c\}$ and X. Let ν assign probability of 1/3 to each of the orderings $a \succ b \succ c$, $b \succ c \succ a$, and $c \succ a \succ b$. Let ν' assign probability of 1/3 to each of the other three orderings $a \succ c \succ b$, $b \succ a \succ c$, and $c \succ b \succ a$. Then ν and ν' induce different choice frequencies in each of the binary menus, but the resulting marginal distribution of choices λ is uniform for both of them.

The failure of identification can be easily understood by looking at Figure 1. Any λ in the interior of the core can be represented in multiple ways as a convex combination

[17] Barberá and Pattanaik [1986] and Fishburn [1998] give simple examples showing the non-uniqueness of the rationalizing distribution of preferences when there are four alternatives. These examples demonstrate that even the support of ν cannot be identified. Turansick [2022] characterizes those conditional stochastic choice functions that admit a unique representation, and shows that this is equivalent to the support of the representation being unique.

of its extreme points, and any such representation is a possible RUM rationalization of (μ, λ).

When the domain of preferences is restricted, identification depends on the location of the extreme points corresponding to the feasible orders in O'. Specifically, the model is identified if and only if these extreme points are affinely independent. In particular, identification fails whenever O' contains more than n elements. For the case of three alternatives and a full-support μ, the model is identified if and only if O' contains at most three orders. We were unable to obtain a full characterization of identified domains with more than three alternatives.

While point identification is very demanding under RUM, it may be that some properties of the distribution of preferences can be inferred from (μ, λ). It is known for example that the probabilities of contour sets are pinned down by the (conditional) stochastic choices from menus [Falmagne, 1978, Barberá and Pattanaik, 1986]. This is no longer true when the dataset contains only the marginals (μ, λ). Indeed, in Example 2 the probability that a is ranked above b and below c is $1/3$ under ν but is 0 under ν'. However, we now show that it is still possible to obtain an upper bound on the frequency with which a given set of alternatives is top-ranked in the population. Given $A \in \mathcal{X}$, let $T[A]$ be the set of orders that rank any element of A above any element of A^c.

Proposition 5. Suppose that ν RUM rationalizes the marginal stochastic choice dataset (μ, λ). Then whenever $A \in \mathcal{X}$ satisfies $1 - v_\mu(A) - v_\mu(A^c) \neq 0$ it holds that

$$\nu(T[A]) \leq \frac{\sum_{a \in A} \lambda(a) - v_\mu(A)}{1 - v_\mu(A) - v_\mu(A^c)}.$$

5 The Luce model

The following definition formalizes the idea that the marginal stochastic choice dataset (μ, λ) is consistent with an individual who behaves according to the Luce model of stochastic choice.

Definition 3. The marginal stochastic choice dataset (μ, λ) is Luce rationalizable when there is $u \in Int(\Delta(X))$ such that for every $a \in X$

$$\lambda(a) = \sum_{A: a \in A} \mu(A) \frac{u(a)}{\sum_{b \in A} u(b)}. \tag{1}$$

We say u Luce rationalizes (μ, λ) when (1) holds.[18]

In standard stochastic choice, it is well-known that the RUM can rationalize more datasets than the Luce model. Nevertheless, we now show that RUM and Luce rationalizations are essentially the same in terms of the marginals they can generate. However, in contrast to RUM, if the marginal distribution over menus μ satisfies a mild richness condition, then under Luce we can identify the parameter u and therefore deduce the conditional choice probabilities in every menu by observing only the marginals. To formulate these results we need the following definition.

Definition 4. Given $\mu \in \Delta(\mathcal{X})$, the set $A \subseteq X$ is μ-separating if there is no menu B in the support of μ such that both $B \cap A \neq \emptyset$ and $B \cap A^c \neq \emptyset$. We say that μ is rich if there are no μ-separating sets except X and $\emptyset$.

Put differently, A is μ-separating if every menu in the support of μ is either contained in A or contained in its complement A^c. Richness is satisfied for example when the grand menu X is in the support of μ. We can now state the key result of this section.

Proposition 6. Suppose that μ is rich. Then the mapping $u \longrightarrow \lambda$ given by (1) is a bijection between $Int(\Delta(X))$ and $Int(Core(v_\mu))$.

Proposition 6 allows us to obtain the following characterization of Luce rationalizability.

Corollary 3. The marginal stochastic choice dataset (μ, λ) is Luce rationalizable if and only if $\lambda \in Int(Core(v_\mu))$.

[18] The definition requires that the sum of coordinates of u is 1. Clearly, any positive vector can be normalized without changing the resulting stochastic choice function.

Recall that $\lambda \in Core(v_\mu)$ is precisely the condition characterizing consistent datasets, as well as those that can be rationalized by the RUM. The corollary says that 'typically' – when λ is in the relative interior of the core – consistent datasets can also be rationalized by the Luce model. Distributions λ on the relative boundary cannot be rationalized because the weights u must be strictly positive. We emphasize that, unlike Proposition 6, the corollary does not require μ to be rich; the proof shows that it is always possible to partition X into subsets such that the restriction of μ to each of them is rich, and Proposition 6 can then be applied in each of these subsets to deliver the result.

Another simple consequence of Proposition 6 is the following corollary which gives necessary and sufficient conditions for the parameter u to be identified.

Corollary 4. Suppose that (μ, λ) is Luce rationalizable. Then the rationalizing vector $u \in Int(\Delta(X))$ is unique if and only if μ is rich.

Before moving on we should point out that Monderer et al. [1992] prove a very similar result to that of Proposition 6. In cooperative game theory, the distribution λ obtained from the game v_μ via equation (1) is known as a positively weighted Shapley value, where the weights are given by the vector u. When u is the constant vector one obtains the standard Shapley value of the game v_μ. Monderer et al. [1992] prove that for a strictly convex game v the core is homeomorphic to the set of all weighted Shapley values (allowing for lexicographic systems of weights). Our result is weaker in that we only consider positive weights and the games we consider are totally monotone which implies convexity. On the other hand, our richness assumption is weaker than strict convexity of v_μ (recall Lemma 1). Our proof is also somewhat simpler due to the total monotonicity of v_μ, although the heart of the argument is the same as in Monderer et al. [1992].

Even when μ is rich and u can be identified from (μ, λ), it is typically impossible to express the inverse mapping from λ to u in closed form. It is therefore natural to ask whether we can still say something about properties of u, at least in some special cases.

The next result gives an example where this is the case. We show that if μ satisfies a certain symmetry property between two alternatives, then the alternative that has higher marginal choice probability also has a higher conditional choice probability in every menu that contains both.

Definition 5. For $a, b \in X$, the marginal distribution of menu availability $\mu \in \Delta(\mathcal{X})$ is ab-exchangeable when $\mu(A \cup \{a\}) = \mu(A \cup \{b\})$ for every $A \subseteq X \setminus \{a, b\}$. The distribution μ is exchangeable when it is ab-exchangeable for every pair $a, b \in X$.

Proposition 7. Suppose that u Luce rationalizes (μ, λ) and that μ is ab-exchangeable. Then $\lambda(a) \geq \lambda(b)$ if and only if $u(a) \geq u(b)$. In particular, when μ is exchangeable the ranking of menu-contingent choice probabilities is the same as the ranking of marginal choice probabilities.

6 Endogenous menu choice

In the models considered up to now the marginal distribution of menu availability μ was viewed as exogenously given and choices made by the agent (or population of agents) were only affecting the marginal choice probabilities λ. We now consider a conceptually different scenario in which the agent first chooses a menu of alternatives and then chooses an alternative from that menu in the second period. Our goal here is not to provide a comprehensive analysis of this case, but rather to point out that prominent 'behavioral' models of menu preferences may imply significant testable implications not only for the observable menu distribution (μ), but also for the resulting distribution over alternatives (λ). Indeed, the choice of a menu in the first period is informative about the agent's preferences over alternatives, and therefore also about the alternative that the agent eventually chooses.

We illustrate this possibility using the model of Gul and Pesendorfer [2001], where preferences over menus are affected by anticipated temptation in the second stage; agents are worried about lack of self-control and therefore prefer to limit their future choices.

Formally, an agent is characterized by a pair of functions $u, v : X \to \mathbb{R}$. The function u describes the agent's value for the alternatives when temptation is absent, while the function v can be interpreted as the agent's urges when choosing an alternative from a menu. Given (u, v), preferences over menus are represented by the utility function

$$U_{u,v}(A) = \max_{a \in A}[u(a) + v(a)] - \max_{b \in A}[v(b)].$$

After choosing a menu, the agent chooses an alternative from that menu that maximizes the sum $u(a) + v(a)$.

The following definition formalizes the idea that the marginals (μ, λ) are consistent with a population of individuals who behave according to the above temptation and self-control model.

Definition 6. Let $\mathcal{X}' \subseteq \mathcal{X}$ be a non-empty collection of feasible menus. The marginal stochastic choice dataset (μ, λ) is temptation and self-control rationalizable given $\mathcal{X}'$ ($\mathcal{X}'$-TSC rationalizable, for short) when there is a distribution ψ over $\mathbb{R}^X \times \mathbb{R}^X$ such that:

(i) For every $A \in \mathcal{X}'$,

$$\mu(A) = \psi\Big(\big\{(u, v) : \; U_{u,v}(A) > U_{u,v}(B) \;\; \forall B \in \mathcal{X}' \setminus \{A\}\big\}\Big).$$

(ii) For every $a \in X$,

$$\lambda(a) = \sum_{A \in \mathcal{X}' : a \in A} \psi\Big(\big\{(u, v) \;\; : \;\; u(a) + v(a) > u(b) + v(b) \;\; \forall b \in A \setminus \{a\} \; \textit{and} \;\; U_{u,v}(A) > U_{u,v}(B) \;\; \forall B \in \mathcal{X}' \setminus \{A\}\big\}\Big).$$

Condition (i) in Definition 6 means that the observed marginal distribution over menus matches the population preference distribution over menus, while condition (ii) is the analogous condition for alternatives. Note that we require strict inequalities at both stages so that individuals strictly prefer their chosen menu over any other feasible menu, and that they strictly prefer their chosen alternative over any other alternative in the menu they selected. Allowing for ties both complicates the notation and trivializes the problem.[19]

To characterize $\mathcal{X}'$-TSC rationalizability, we need to introduce some additional notation. First, for any feasible menu $A \in \mathcal{X}'$ let $\bar{A}_{\mathcal{X}'} = A \setminus \bigcup_{\{B \in \mathcal{X}' \colon B \subsetneq A\}} B$ be the set of alternatives in A that are not contained in any feasible sub-menu of A. We say that $A \in \mathcal{X}'$ is redundant when $\bar{A}_{\mathcal{X}'} = \emptyset$.

To illustrate, suppose that $X = \{a, b, c\}$ and the collection of feasible menus is $\mathcal{X}' = \{\{a\}, \{c\}, \{a, b\}, \{b, c\}, X\}$. Then $\overline{\{a, b\}}_{\mathcal{X}'} = \overline{\{b, c\}}_{\mathcal{X}'} = \{b\}$ and $\overline{X}_{\mathcal{X}'} = \emptyset$. This means that an agent who chooses either $\{a, b\}$ or $\{b, c\}$ in the first period will necessarily choose b in the second period, since if they wanted a or c they could have chosen the singleton menus in the first period and avoid second-period temptation. In addition, we should not observe X being chosen at all in the first period since it is redundant – any alternative is contained in a feasible sub-menu.

Now, given $\mu \in \Delta(\mathcal{X}')$ we define a game $v_\mu^{\mathcal{X}'}$ by $v_\mu^{\mathcal{X}'}(A) = \sum_{\{B \in \mathcal{X}' : \bar{B}_{\mathcal{X}'} \subseteq A\}} \mu(B)$ for every non-empty $A \subseteq X$. The definition of $v_\mu^{\mathcal{X}'}$ differs from that of v_μ of Section 3 in that here we sum up the probabilities of all menus B such that $\bar{B}_{\mathcal{X}'} \subseteq A$ rather than only menus B that are contained in A themselves. Thus, we clearly have that $v_\mu^{\mathcal{X}'} \geq v_\mu$.

Proposition 8. The marginal stochastic choice dataset (μ, λ) is $\mathcal{X}'$-TSC rationalizable if and only if $\mu(A) = 0$ for every redundant menu A and $\lambda \in Core(v_\mu^{\mathcal{X}'})$.

Compared with Proposition 1, the conditions for TSC rationalizability are stronger in two ways than the conditions for (unrestricted) consistency. The first is that redundant menus cannot be in the support of μ,[20] and the second is that the core constraints are

[19] TSC-rationalizability depends on the collection of feasible menus $\mathcal{X}'$. Restricted domains of feasible menus are natural in many applications and are often used in the stochastic choice literature.

[20] This property corresponds to the 'Set Betweenness' axiom that Gul and Pesendorfer [2001] use in their characterization of the model. Indeed, if a menu is redundant then it is equal to the union of its feasible sub-menus. Set Betweenness implies that at least one of these sub-menus must be preferred to the union.

calculated based on the game $v_\mu^{\mathcal{X}'}$ which, as explained above, is larger than the game v_μ.

Going back to the above example, recall that X is redundant and hence $\mathcal{X}'$-TSC rationalizability requires that $\mu(X) = 0$; all other menus in $\mathcal{X}'$ can be included in the support of μ. The second condition of the proposition, $\lambda \in Core(v_\mu^{\mathcal{X}'})$, requires that $\lambda(a) \geq \mu(a)$, $\lambda(c) \geq \mu(c)$, as well as $\lambda(b) \geq \mu(\{a,b\}) + \mu(\{b,c\})$.[21] The latter inequality reflects the fact mentioned above that agents who chose one of the menus $\{a,b\}$ or $\{b,c\}$ in the first stage necessarily choose alternative b in the second stage. It follows that for every μ there is a unique λ such that (μ,λ) is $\mathcal{X}'$-TSC rationalizable, namely, $\lambda(a) = \mu(a)$, $\lambda(c) = \mu(c)$, and $\lambda(b) = \mu(\{a,b\}) + \mu(\{b,c\})$.

In contrast, if any stochastic choice function π is allowed, then the inequalities that characterize consistency are given by $\lambda(a) \geq \mu(a)$, $\lambda(c) \geq \mu(c)$, $\lambda(a) + \lambda(b) \geq \mu(a) + \mu(\{a,b\})$, and $\lambda(b)+\lambda(c) \geq \mu(c)+\mu(\{b,c\})$. Thus, for typical distributions μ there would be many different λ's such that (μ,λ) is consistent. Figure 3 illustrates the difference between (unrestricted) consistency and $\mathcal{X}'$-TSC rationalizability for this example in the case where $\mu(\{a\}) = \mu(\{c\}) = \mu(\{a,b\}) = \mu(\{b,c\}) = \frac{1}{4}$.

7 Final comments

This paper studies properties of various models of stochastic choice when available data is limited to the marginal distributions of menus and choices. We focused on the two most prominent models, RUM and Luce. As can be expected, refuting these models based solely on the marginals is typically hard. But, somewhat surprisingly, one may be able to infer quite a lot about the models' parameters. For example, it is often possible to completely recover the value of alternatives for the Luce rule from marginal data.

We have also shown that restricting the domain of preferences in the RUM does lead to significant testable implications for the marginals. A possible direction for future work is to more carefully explore the implications of such domain restrictions. For example,

[21] The inequalities corresponding to other sets $A \subseteq X$ are implied by those for singleton menus in this example, but this is not the case in general.

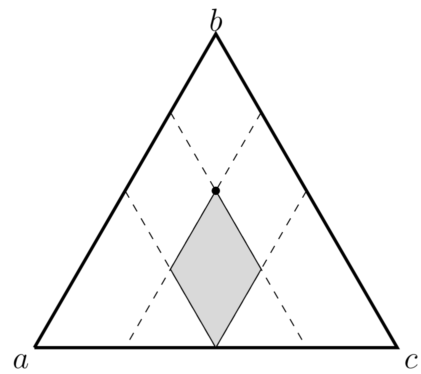

Figure 3: An illustration of the example with $X = \{a, b, c\}$, $\mathcal{X}' = \{\{a\}, \{c\}, \{a, b\}, \{b, c\}, X\}$, and $\mu(\{a\}) = \mu(\{c\}) = \mu(\{a, b\}) = \mu(\{b, c\}) = \frac{1}{4}$. The shaded area contains all choice distributions λ such that (μ, λ) is consistent (the core of v_μ). The only λ such that (μ, λ) is $\mathcal{X}'$-TSC rationalizable is $\lambda(a) = \lambda(c) = \frac{1}{4}$ and $\lambda(b) = \frac{1}{2}$ – the marked point in the figure.

one can study the case where alternatives are lotteries and preferences are assumed to be represented by expected utility [Gul and Pesendorfer, 2006], or the case where alternatives are bundles of goods and preferences satisfy monotonicity [Kitamura and Stoye, 2018]. A slightly different example is Filiz-Ozbay and Masatlioglu [2023], where agents may be 'boundedly rational'.

A different scenario in which marginal datasets naturally arise is when the available menu is fixed but individuals consider only a subset of the alternatives. In this case μ describes the distribution of *consideration sets* in the population. A previous version of this paper included an analysis of the model introduced by Manzini and Mariotti [2014] in this context. Other models of stochastic consideration sets can be explored as well.

When the distribution of menus is endogenous, additional restrictions may arise since the choice of a menu can contain information about which alternative will eventually be chosen. We have demonstrated this possibility using the model of Gul and Pesendorfer [2001]. Another interesting direction would be to study two-period models in which an agent chooses a distribution over menus as well as a distribution over alternatives from each menu. These distributions may arise as the result of maximization behavior that trades off the benefits and costs of distributions of menus/alternatives similar to perturbed utility models such as Fudenberg et al. [2015] and Allen and Rehbeck [2022]. If the maximization problems across the two periods are connected, then second-period stochastic choices would be constrained by the menu distribution chosen in the first period. That could generate non-trivial testable implications for the pair (μ, λ).

Finally, while we assumed throughout that the observables are μ and λ, in applications it may be that more or less data can be accessed. One plausible scenario is that the researcher cannot observe how often each menu is available. Instead of observing μ, the data only shows the availability of each alternative $a \in X$, namely, only $\xi(a) = \sum_{A:a\in A} \mu(A)$ for each a is observable.[22] Say that the marginal distribution of choices $\lambda \in \Delta(X)$ is potentially–consistent given $\xi \in [0,1]^X$ when there exists $\mu \in \Delta(\mathcal{X})$ such that (μ, λ) is consistent and such that $\xi(a) = \sum_{A:a\in A} \mu(A)$ for every $a \in X$. The following is a simple consequence of Proposition 1.

Corollary 5. The distribution of choices $\lambda \in \Delta(X)$ is potentially–consistent given $\xi \in [0,1]^X$ if and only if $\lambda(a) \leq \xi(a)$ for every $a \in X$.

It is also plausible that, in addition to the aggregate data, the researcher has choice frequencies in some menus (e.g., sales data from some retailers) but not in others; or it may be that conditional choice probabilities for a particular alternative are available from its producer, while for the rest of the alternatives only the aggregate is known. It appears that much of our analysis would carry over to such situations, but we do not pursue these directions here.

References

Jason Abaluck and Abi Adams-Prassl. What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Eco-*

[22]As mentioned before, this is the data available in Bruno and Vilcassim [2008]. They assume independence of availability across products to generate the menu distribution.

nomics, 136(3):1611–1663, 2021.

Victor H Aguiar. Random categorization and bounded rationality. *Economics Letters*, 159:46–52, 2017.

Roy Allen and John Rehbeck. Revealed stochastic choice with attributes. *Economic Theory*, pages 1–22, 2022.

Jose Apesteguia, Miguel A Ballester, and Jay Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, 2017.

Salvador Barberá and Prasanta K Pattanaik. Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, 54(3):707–715, 1986.

Levon Barseghyan, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048, 2021.

Moshe Ben-Akiva and Bruno Boccara. Discrete choice models with latent choice sets. *International journal of Research in Marketing*, 12(1):9–24, 1995.

Antoine Billot and Jacques-François Thisse. How to share when context matters: The möbius value as a generalized solution for cooperative games. *Journal of Mathematical Economics*, 41(8):1007–1029, 2005.

H.D. Block and I. Marschak. Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, pages 97–132, 1960.

Richard L Brady and John Rehbeck. Menu-dependent stochastic feasibility. *Econometrica*, 84(3):1203–1223, 2016.

Hernán A Bruno and Naufel J Vilcassim. Research note—structural demand estimation with varying product availability. *Marketing Science*, 27(6):1126–1131, 2008.

Matias D Cattaneo, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. A random attention model. *Journal of Political Economy*, 128(7):2796–2836, 2020.

Matias D Cattaneo, Paul Cheung, Xinwei Ma, and Yusufcan Masatlioglu. Attention overload. *arXiv preprint arXiv:2110.10650*, 2021.

Christopher P Chambers, Yusufcan Masatlioglu, and Christopher Turansick. Correlated choice. *Theoretical Economics*, 19(3):1087–1117, 2024.

Alain Chateauneuf and Jean-Yves Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of möbius inversion. *Mathematical social sciences*, 17(3):263–283, 1989.

Paul Cheung and Yusufcan Masatlioglu. Decision making with recommendation. working paper, 2021.

Valentino Dardanoni, Paola Manzini, Marco Mariotti, and Christopher J Tyson. Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, 88(3):1269–1296, 2020.

Arthur P Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232, 1968a.

Arthur P Dempster. Upper and lower probabilities generated by a random closed interval. *The Annals of Mathematical Statistics*, 39(3):957–966, 1968b.

Laura Doval and Ran Eilat. The core of bayesian persuasion. *arXiv preprint arXiv:2307.13849*, 2023.

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

Emel Filiz-Ozbay and Yusufcan Masatlioglu. Progressive random choice. *Journal of Political Economy*, 131(3):716–750, 2023.

Peter C Fishburn. Stochastic utility. *Handbook of utility theory*, 1:273–319, 1998.

Mira Frick, Ryota Iijima, and Tomasz Strzalecki. Dynamic random utility. *Econometrica*, 87(6):1941–2002, 2019.

Drew Fudenberg and Tomasz Strzalecki. Dynamic logit with choice aversion. *Econometrica*, 83(2):651–691, 2015.

Drew Fudenberg, Ryota Iijima, and Tomasz Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.

Itzhak Gilboa and Dov Monderer. A game-theoretic approach to the binary stochastic choice problem. *Journal of Mathematical Psychology*, 36(4):555–572, 1992.

Michel Grabisch. *Set functions, games and capacities in decision making*, volume 46. Springer, 2016.

T.W. Gruen, D. Corsten, and S. Bharadwaj. Retail out of stocks: A worldwide examination of causes, rates, and consumer responses. Technical report, Grocery Manufacturers of America, 2002.

Faruk Gul and Wolfgang Pesendorfer. Temptation and self-control. *Econometrica*, 69 (6):1403–1435, 2001.

Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74(1): 121–146, 2006.

P Hall. On representation of subsets. *Journal of the London Mathematical Society*, 10: 26–30, 1934.

Jerry Hausman, Gregory Leonard, and J Douglas Zona. Competitive analysis with differenciated products. *Annales d'Economie et de Statistique*, pages 159–180, 1994.

Jerry A Hausman and Gregory K Leonard. The competitive effects of a new product introduction: A case study. *The Journal of Industrial Economics*, 50(3):237–263, 2002.

William Hickman and Julie Holland Mortimer. Demand estimation with availability variation. In *Handbook on the Economics of Retailing and Distribution*, pages 306–340. Edward Elgar Publishing, 2016.

Joel L Horowitz. Modeling the choice of choice set in discrete-choice random-utility models. *Environment and Planning a*, 23(9):1237–1246, 1991.

Tatsuro Ichiishi. Super-modularity: Applications to convex games and to the greedy algorithm for lp. *Journal of Economic Theory*, 25(2):283–286, 1981.

Ehud Kalai and Dov Samet. On weighted shapley values. *International journal of game theory*, 16:205–222, 1987.

Nail Kashaev and Victor H Aguiar. A random attention and utility model. *Journal of Economic Theory*, 204(C):105487, 2022.

Yuichi Kitamura and Jörg Stoye. Nonparametric analysis of random utility models. *Econometrica*, 86(6):1883–1909, 2018.

Jon Kleinberg and Eva Tardos. *Algorithm design*. Pearson Education India, 2006.

Matthew Kovach and Gerelt Tserenjigmid. Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, 130(9):2411–2461, 2022.

Zhentong Lu. Estimating multinomial choice models with unobserved choice sets. *Journal of Econometrics*, 226(2):368–398, 2022.

R Duncan Luce. *Individual choice behavior: A theoretical analysis*. John Wiley & Sons, Inc., 1959.

Charles F Manski. The structure of random utility models. *Theory and decision*, 8(3):229, 1977.

Paola Manzini and Marco Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176, 2014.

Paola Manzini, Marco Mariotti, and Levent Ulku. Sequential approval: A model of "likes", paper downloads and other forms of click behaviour. working paper, 2019.

David A Matsa. Competition and product quality in the supermarket industry. *The Quarterly Journal of Economics*, 126(3):1539–1591, 2011.

Daniel McFadden and Marcel K Richter. Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, pages 161–186, 1990.

Dov Monderer. The stochastic choice problem: A game-theoretic approach. *Journal of Mathematical Psychology*, 36(4):547–554, 1992.

Dov Monderer, Dov Samet, and Lloyd S Shapley. Weighted values and the core. *International Journal of Game Theory*, 21(1):27–39, 1992.

Aviv Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342, 2001.

Glenn Shafer. A mathematical theory of evidence. In *A mathematical theory of evidence*. Princeton university press, 1976.

Lloyd Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

Tomasz Strzalecki. Stochastic choice theory. *Cambridge Books*, 2025.

Joffre D Swait and Moshe Ben-Akiva. Analysis of the effects of captivity on travel time and cost elasticities. In *Behavioural Research for Transport Policy, International Association for Travel Behaviour*, 1986.

Steven Tenn. Demand estimation under limited product availability. *Applied Economics Letters*, 16(5):465–468, 2009.

Steven Tenn and John M Yun. Biases in demand analysis due to variation in retail distribution. *International Journal of Industrial Organization*, 26(4):984–997, 2008.

Christopher Turansick. Identification in the random utility model. *Journal of Economic Theory*, 203(C):105489, 2022.

Robert J Weber. Probabilistic values for games. *The Shapley Value. Essays in Honor of Lloyd S. Shapley*, pages 101–119, 1988.

A Proofs

Proof of Proposition 2

Clearly, we only need to show that if (μ, λ) is consistent, then it is RUM rationalizable. By Proposition 1, if (μ, λ) is consistent then $\lambda \in Core(v_\mu)$ and recall from Section 2 that v_μ is totally monotone and hence convex. Since $Core(v_\mu)$ is a convex and compact polytope, λ can be written as a convex combination of its extreme points. Thus, by Proposition 3 there is a distribution ν over O such that for every $a \in X$,

$$\lambda(a) = \sum_{\succ \in O} \nu(\succ)[v_\mu(L_\succ(a) \cup \{a\}) - v_\mu(L_\succ(a))].$$

Now,

$$v_\mu(L_\succ(a) \cup \{a\}) - v_\mu(L_\succ(a)) = \sum_{\{A:\ a \in A \subseteq L_\succ(a) \cup \{a\}\}} \mu(A) = \sum_{\{A:\ a \in A,\ \succ \in T[a,A]\}} \mu(A),$$

where the first equality is by the definitions of v_μ and $L_\succ(a)$, and the second equality is by the definition of the set $T[a, A]$. Combining the above two equations, we get that

$$\lambda(a) = \sum_{\succ \in O} \nu(\succ) \sum_{\{A:\ a \in A,\ \succ \in T[a,A]\}} \mu(A) = \sum_{\{A: a \in A\}} \mu(A)\nu(T[a, A]),$$

where the last equality is just a change in the order of summation.

Proof of Proposition 4

Suppose that ν O'-RUM rationalizes (μ, λ) and fix $k \in \{1, \ldots, K\}$. We have

$$\begin{aligned}\sum_{a \in \cup_{i=1}^k A_i} \lambda(a) &= \sum_{a \in \cup_{i=1}^k A_i} \sum_{\{B: a \in B\}} \mu(B)\nu(T[a, B]) = \\ & \sum_{a \in \cup_{i=1}^k A_i} \sum_{\{B: a \in B \subseteq \cup_{i=1}^k A_i\}} \mu(B)\nu(T[a, B]) = \\ & \sum_{\{B: B \subseteq \cup_{i=1}^k A_i\}} \mu(B) \sum_{a \in B} \nu(T[a, B]) = \sum_{\{B: B \subseteq \cup_{i=1}^k A_i\}} \mu(B) = \upsilon_\mu(\cup_{i=1}^k A_i),\end{aligned}$$

where the first equality holds since ν rationalizes (μ, λ), the second equality follows from $\nu(T[a, B]) = 0$ whenever $a \in B \nsubseteq \cup_{j=1}^k A_j$ (since the support of ν is contained in O'), the next is a change in the order of summation, and the next follows from $\sum_{a \in B} \nu(T[a, B]) = 1$ for every set B.

Conversely, suppose (μ, λ) is consistent and that $\upsilon_\mu(\cup_{i=1}^k A_i) = \sum_{a \in \cup_{i=1}^k A_i} \lambda(a)$ for all k. By Proposition 2 there is $\nu \in \Delta(O)$ that RUM-rationalizes (μ, λ). Thus,

$$\begin{aligned}\sum_{a \in \cup_{i=1}^k A_i} \lambda(a) &= \sum_{a \in \cup_{i=1}^k A_i} \sum_{\{B: a \in B\}} \mu(B)\nu(T[a, B]) \geq \\ & \sum_{a \in \cup_{i=1}^k A_i} \sum_{\{B: a \in B \subseteq \cup_{i=1}^k A_i\}} \mu(B)\nu(T[a, B]) = \upsilon_\mu(\cup_{i=1}^k A_i).\end{aligned}$$

By assumption, the first and last expressions are equal, so the inequality must in fact hold as equality. This implies that $\mu(B)\nu(T[a, B]) = 0$ for every pair a, B with $a \in \left(\cup_{i=1}^k A_i\right) \cap B$ and $B \nsubseteq \cup_{i=1}^k A_i$. In particular, for any pair of alternatives $a \in \cup_{i=1}^k A_i$ and $b \notin \cup_{i=1}^k A_i$ we have that $\mu(\{a, b\})\nu(T[a, \{a, b\}]) = 0$. The proposition assumes that $\mu(\{a, b\}) > 0$, which implies that any $\succ$ with $a \succ b$ is not in the support of ν. We have shown that for any k, the support of ν does not contain orders that rank alternatives from $\cup_{i=1}^k A_i$ above alternatives from the complement of this set. This proves that the support of ν is contained in O' as needed.

Proof of Proposition 5

Suppose that ν Rum-rationalizes (μ, λ) and that $A \in \mathcal{X}$ satisfies $1 - v_\mu(A) - v_\mu(A^c) \neq 0$. Then the total probability that some alternative from A is chosen satisfies

$$\sum_{a \in A} \lambda(a) \geq \nu(T[A])(1 - v_\mu(A^c)) + (1 - \nu(T[A]))v_\mu(A).$$

Indeed, with probability $\nu(T[A])$ the realized order ranks A at the top, in which case an element of A will be chosen so long as an element of A is available. And with the complementary probability $1 - \nu(T[A])$ an element of A must be chosen whenever only elements of A are available. Rearranging the above inequality gives the bound in the proposition.

Proof of Proposition 6

Suppose that μ is rich. We start by showing that the mapping $u \longrightarrow \lambda$ given by (1) is one-to-one. Suppose that u, u' induce the same marginal distribution of choices λ. Let $A_\circ = \arg\max_{a \in X} \frac{u(a)}{u'(a)}$. We claim that $A_\circ = X$. Suppose by contradiction that this is not the case. Then, since μ is rich, there is a menu B in the support of μ that intersects both $A_\circ$ and $A_\circ^c$. Let $a_\circ \in B \cap A_\circ$ and $a_1 \in B \cap A_\circ^c$.

Now, for every menu A that contains $a_\circ$ we have

$$\frac{u(a_\circ)}{\sum_{a \in A} u(a)} = \frac{1}{\sum_{a \in A} \frac{u(a)}{u(a_\circ)}} \geq \frac{1}{\sum_{a \in A} \frac{u'(a)}{u'(a_\circ)}} = \frac{u'(a_\circ)}{\sum_{a \in A} u'(a)},$$

where the inequality follows since $\frac{u(a)}{u(a_\circ)} \leq \frac{u'(a)}{u'(a_\circ)}$ for every $a \in X$. Since by assumption

$$\sum_{\{A:\ a_\circ \in A\}} \mu(A) \frac{u(a_\circ)}{\sum_{a \in A} u(a)} = \lambda(a_\circ) = \sum_{\{A:\ a_\circ \in A\}} \mu(A) \frac{u'(a_\circ)}{\sum_{a \in A} u'(a)},$$

it follows that for every menu A in the support of μ that contains $a_\circ$ we have $\frac{u(a_\circ)}{\sum_{a \in A} u(a)} = \frac{u'(a_\circ)}{\sum_{a \in A} u'(a)}$. In particular, this equality holds for the menu B: $\frac{u(a_\circ)}{\sum_{a \in B} u(a)} = \frac{u'(a_\circ)}{\sum_{a \in B} u'(a)}$. But this in turn implies that $\frac{u(a_1)}{u(a_\circ)} = \frac{u'(a_1)}{u'(a_\circ)}$, contradicting the fact $a_1 \in A_\circ^c$. Thus, we have shown that $A_\circ = X$ so that the ratio $\frac{u(a)}{u'(a)}$ is constant in a. Since both u and u' are in $\Delta(X)$, it follows that $u = u'$.

Next, we argue that if (μ, λ) is Luce rationalizable then $\lambda \in Int(Core(v_\mu))$. Indeed, fix some $u \in Int(\Delta(X))$ and let λ be its image. Let $A \neq X$ be some menu. By assumption, there is B in the support of μ that intersects both A and A^c. Since every alternative in B is chosen with positive probability when B is the available menu, it follows that $\sum_{a\in A} \lambda(a) > v_\mu(A)$ (recall that $\sum_{a\in A} \lambda(a) = v_\mu(A)$ holds if and only if elements of A are never chosen when some element of A^c is available). Thus, $\lambda \in Int(Core(v_\mu))$ as claimed.

Now, the mapping $u \longrightarrow \lambda$ given by (1) is clearly continuous. Therefore, by the Invariance of Domain theorem, the image of $Int(\Delta(X))$ under this mapping is open in $\Delta(X)$. Furthermore, it follows from the previous paragraph that the image is open in $Int(Core(v_\mu))$.

To finish the proof, let C be the set of λ's in $Int(Core(v_\mu))$ that are *not* in the image. We claim that C is also open in $Int(Core(v_\mu))$. Indeed, if it is not then there is $\lambda_0 \in C$ and a sequence $\lambda_n \to \lambda_0$ such that each λ_n is the image of some u_n under Equation (1). By compactness we may assume w.l.o.g. that u_n converges in $\Delta(X)$, say to u_0. It can't be that $u_0 \in Int(\Delta(X))$ since that would contradict the assumption that $\lambda_0 \in C$. Thus, the set $\bar{A} = \{a : \; u_0(a) = 0\}$ is nonempty. We have

$$\sum_{a\in\bar{A}} \lambda_n(a) = \sum_{A\subseteq\bar{A}} \mu(A) + \sum_{A\not\subseteq\bar{A}} \mu(A) \frac{\sum_{a\in A\cap\bar{A}} u_n(a)}{\sum_{a\in A} u_n(a)} = v_\mu(\bar{A}) + \sum_{A\not\subseteq\bar{A}} \mu(A) \frac{\sum_{a\in A\cap\bar{A}} u_n(a)}{\sum_{a\in A} u_n(a)}.$$

When $n \to \infty$ the ratio $\frac{\sum_{a\in A\cap\bar{A}} u_n(a)}{\sum_{a\in A} u_n(a)}$ converges to 0 since the numerator goes to zero while the denominator does not. We thus get that $\sum_{a\in\bar{A}} \lambda_0(a) = v_\mu(\bar{A})$, contradicting the assumption that λ_0 is in $Int(Core(v_\mu))$. This shows that C is open in $Int(Core(v_\mu))$.

We have shown that both the image of the mapping $u \longrightarrow \lambda$ and its complement C are open in $Int(Core(v_\mu))$. But since $Int(Core(v_\mu))$ is connected, it must be that $C = \emptyset$ and that the image is the entire set $Int(Core(v_\mu))$. This completes the proof.

Proof of Corollary 3

Fix μ and consider the collection of all μ-separating sets. If A is μ-separating then clearly A^c is μ-separating as well; it is also immediate to check that if A, B are μ-separating then the union $A \cup B$ is μ-separating. Thus, this collection is an algebra of subsets of X. Since X is finite this algebra is generated by a partition, namely, there exists a partition $\{A_1, \ldots A_I\}$ of X such that each A_i is μ-separating and such that no strict subset of any of the A_i's is μ-separating. We denote by μ_i and v_μ^i the restrictions of μ and v_μ to subsets of A_i, respectively. Note that $v_\mu^i(A_i)$ may be strictly less than one, in which case the definition of $Core(v_\mu^i)$ is adjusted accordingly to require that the sum of coordinates of core elements is equal to $v_\mu^i(A_i)$. Clearly, we have that

$$Core(v_\mu) = \bigtimes_{i=1}^{I} Core(v_\mu^i). \tag{2}$$

For each $1 \leq i \leq I$, $u_i \in Int(\Delta(A_i))$ and $a \in A_i$ define $f_i(u_i)(a) = \sum_{A:a\in A\subseteq A_i} \mu(A)\frac{u_i(a)}{\sum_{b\in A} u_i(b)}$. Similarly, for $u \in Int(\Delta(X))$ and $a \in X$ let $f(u)(a) = \sum_{A:a\in A} \mu(A)\frac{u(a)}{\sum_{b\in A} u(b)}$. Notice that given any u and $a \in A_i$, if we define $u_i \in Int(\Delta(A_i))$ by $u_i(a) = \frac{1}{\sum_{b\in A_i} u(b)}u(a)$ then $f_i(u_i)$ coincides with $f(u)$ for alternatives in A_i. Conversely, given $(u_1, \ldots, u_I) \in \bigtimes_{i=1}^{I} Int(\Delta(A_i))$, defining $u \in Int(\Delta(X))$ by $u(a) = \frac{1}{I}u_i(a)$ when $a \in A_i$ gives $f(u) \equiv f_i(u_i)$ on A_i for all i. In other words, we have that

$$f\left(Int(\Delta(X))\right) = \bigtimes_{i=1}^{I} f_i\left(Int(\Delta(A_i))\right). \tag{3}$$

Finally, since μ_i is rich by construction, Proposition 6 implies that for every i

$$f_i\left(Int(\Delta(A_i))\right) = Int(Core(v_\mu^i)). \tag{4}$$

Overall we obtain

$$f\left(Int(\Delta(X))\right) = \bigtimes_{i=1}^{I} f_i\left(Int(\Delta(A_i))\right) = \bigtimes_{i=1}^{I} Int(Core(v_\mu^i)) = Int(Core(v_\mu)),$$

where the first equality is by (3), the second is by (4), and the last is by (2) and by the fact that the relative interior of the product of convex sets is equal to the product of their relative interiors.

Proof of Corollary 4

We have already shown in Proposition 6 that if μ is rich then the mapping $u \longrightarrow \lambda$ is one-to-one. For the converse, suppose that there is a μ-separating set $A \neq \emptyset, X$. Let u be a Luce rationalization of (μ, λ). For some $0 < \alpha < \frac{1}{u(A)}$ define the vector u' by $u'(a) = \alpha u(a)$ for $a \in A$ and $u'(a) = \frac{1-\alpha u(A)}{u(A^c)} u(a)$ for $a \in A^c$. It is straightforward to check that u' is another Luce rationalization of (μ, λ).

Proof of Proposition 7

Given a, b, define the collections of menus $F(a,b) = \{A \in \mathcal{X} \ : \ a \in A, \ b \notin A\}$ and $F(b,a) = \{A \in \mathcal{X} \ : \ b \in A, \ a \notin A\}$. Let $g : F(a,b) \to F(b,a)$ be defined by $g(A) = (A \cup \{b\}) \setminus \{a\}$. Then clearly g is a bijection, and $\mu(A) = \mu(g(A))$ for every $A \in F(a,b)$ by ab-exchangeability.

Now, suppose that u Luce rationalizes the marginal stochastic choice dataset (μ, λ). It follows that,

$$\lambda(a) = \sum_{\{A : a \in A\}} \mu(A) \frac{u(a)}{\sum_{c \in A} u(c)} = \sum_{\{A : \{a,b\} \subseteq A\}} \mu(A) \frac{u(a)}{\sum_{c \in A} u(c)} + \sum_{A \in F(a,b)} \mu(A) \frac{u(a)}{\sum_{c \in A} u(c)}. \tag{5}$$

If $u(a) \geq u(b)$, then the first sum satisfies

$$\sum_{\{A : \{a,b\} \subseteq A\}} \mu(A) \frac{u(a)}{\sum_{c \in A} u(c)} \geq \sum_{\{A : \{a,b\} \subseteq A\}} \mu(A) \frac{u(b)}{\sum_{c \in A} u(c)}. \tag{6}$$

As for the second sum, we have

$$\sum_{A\in F(a,b)}\mu(A)\frac{u(a)}{\sum_{c\in A}u(c)} \geq \sum_{A\in F(a,b)}\mu(A)\frac{u(b)}{\sum_{c\in g(A)}u(c)} = \sum_{A\in F(a,b)}\mu(g(A))\frac{u(b)}{\sum_{c\in g(A)}u(c)} = \sum_{A\in F(b,a)}\mu(A)\frac{u(b)}{\sum_{c\in A}u(c)}, \quad (7)$$

where the inequality is since $t\to\frac{t}{t+c}$ is increasing (when $c>0$), the first equality follows from $\mu(A)=\mu(g(A))$ for all $A\in F(a,b)$, and the last equality follows from g being a bijection. Combining (5), (6), and (7) we get

$$\lambda(a)\geq \sum_{\{A:\{a,b\}\subseteq A\}}\mu(A)\frac{u(b)}{\sum_{c\in A}u(c)}+\sum_{A\in F(b,a)}\mu(A)\frac{u(b)}{\sum_{c\in A}u(c)}=\lambda(b).$$

The converse, that $\lambda(a)\geq\lambda(b)$ implies $u(a)\geq u(b)$, easily follows.

Proof of Proposition 8

Since the collection of feasible menus $\mathcal{X}'$ is fixed, we omit it from the notation and write $\bar{A}$ instead of $\bar{A}_{\mathcal{X}'}$. We still write $v_{\mu}^{\mathcal{X}'}$ to avoid confusion with the game v_μ as defined earlier in the paper. We denote the collections of redundant and non-redundant menus in $\mathcal{X}'$ by R and NR, respectively.

The first step of the proof is the following lemma.

Lemma 2. Fix $A\in\mathcal{X}'$ and $a\in A$. There exists $(u,v)\in\mathbb{R}^X\times\mathbb{R}^X$ such that $U_{u,v}(A)>U_{u,v}(B)$ for all $B\in\mathcal{X}'\setminus\{A\}$ and such that $u(a)+v(a)>u(b)+v(b)$ for all $b\in A\setminus\{a\}$ if and only if $a\in\bar{A}$. Moreover, if $A\in R(\mathcal{X}')$ then there is no $(u,v)\in\mathbb{R}^X\times\mathbb{R}^X$ such that $U_{u,v}(A)>U_{u,v}(B)$ for all $B\in\mathcal{X}'\setminus\{A\}$.

Proof. Suppose first that $a\in\bar{A}=A\setminus\bigcup_{\{B\in\mathcal{X}':B\subsetneq A\}}B$. Define u,v by $u(a)=2$, $u(b)=0$ for all $b\neq a$, $v(b)=0$ for all $b\in A$, and $v(b)=1$ for all $b\notin A$. First, we have that $u(a)+v(a)=2$ while $u(b)+v(b)\leq 1$ for any other b, so that in particular $u(a)+v(a)>$

$u(b) + v(b)$ for all $b \in A \setminus \{a\}$. Second,

$$U_{u,v}(A) = \max_{a' \in A}[u(a') + v(a')] - \max_{b \in A}[v(b)] = 2 - 0 = 2.$$

Let $B \in \mathcal{X}'$ be any other feasible menu and consider two possible cases: If $a \notin B$ then

$$U_{u,v}(B) = \max_{a' \in B}[u(a') + v(a')] - \max_{b \in B}[v(b)] \leq 1 - 0 = 1;$$

if $a \in B$ then by assumption B must contain an element from A^c, so that

$$U_{u,v}(B) = \max_{a' \in B}[u(a') + v(a')] - \max_{b \in B}[v(b)] \leq 2 - 1 = 1.$$

It follows that in either case $U_{u,v}(A) > U_{u,v}(B)$, as needed.

Conversely, suppose that $a \in B \subsetneq A$ for some $B \in \mathcal{X}'$. Consider any pair of functions (u, v) that satisfies $u(a) + v(a) > u(b) + v(b)$ for all $b \in A \setminus \{a\}$. Then in particular $u(a) + v(a) > u(b) + v(b)$ for all $b \in B \setminus \{a\}$, so that

$$U_{u,v}(B) = \max_{a' \in B}[u(a')+v(a')]-\max_{b \in B}[v(b)] = [u(a)+v(a)]-\max_{b \in B}[v(b)] \geq [u(a)+v(a)]-\max_{b \in A}[v(b)] = U_{u,v}(A).$$

Therefore, it can't be that $U_{u,v}(A) > U_{u,v}(B)$ for all $B \in \mathcal{X}' \setminus \{A\}$.

Finally, we need to show that if $A \in R(\mathcal{X}')$ then there is no pair (u, v) such that $U_{u,v}(A) > U_{u,v}(B)$ for any $B \in \mathcal{X}' \setminus \{A\}$. Fix $A \in R(\mathcal{X}')$ and a pair (u, v). Let $a^* \in \arg\max_{a \in A} u(a) + v(a)$. By assumption, there is $B \in \mathcal{X}'$ such that $a^* \in B \subsetneq A$. Then as in the previous paragraph we have that

$$U_{u,v}(B) = [u(a^*) + v(a^*)] - \max_{b \in B}[v(b)] \geq [u(a^*) + v(a^*)] - \max_{b \in A}[v(b)] = U_{u,v}(A),$$

where the first equality is because $a^* \in \arg\max_{a \in B} u(a) + v(a)$, and the inequality is due to $B \subseteq A$. □

The next lemma, which easily follows from the previous one, argues that for (μ, λ) to be $\mathcal{X}'$-TSC rationalizable it is necessary and sufficient to consider only non-redundant menus and only alternatives that cannot be found in sub-menus.

Lemma 3. The pair (μ, λ) is $\mathcal{X}'$-TSC rationalizable if and only if there exists a collection $\pi = \{\pi(\cdot|A)\}_{A \in NR}$, such that the support of each $\pi(\cdot|A)$ is contained in $\bar{A}$, and such that for every $a \in X$

$$\lambda(a) = \sum_{A \in NR} \mu(A)\pi(a|A).$$

Proof. Suppose first that π satisfies the requirements of the lemma. By Lemma 2, for every $A \in NR$ and every $a \in \bar{A}$ there is a pair u, v such that $U_{u,v}(A) > U_{u,v}(B)$ for all $B \in \mathcal{X}' \setminus \{A\}$ and such that $u(a) + v(a) > u(b) + v(b)$ for all $b \in A \setminus \{a\}$. Denote this pair by $(u, v)^{A,a}$. Let ψ be the distribution with support $\left\{(u, v)^{A,a}\right\}_{A \in NR,\ a \in \bar{A}}$, and with $\psi\left((u, v)^{A,a}\right) = \mu(A)\pi(a|A)$ for every a, A. Then it is immediate that ψ satisfies conditions (1) and (2) of Definition 6 and therefore that (μ, λ) is $\mathcal{X}'$-TSC rationalizable.

Conversely, suppose that (μ, λ) is $\mathcal{X}'$-TSC rationalizable and let ψ be a rationalizing distribution over pairs u, v. First, Lemma 2 implies that if $A \in R(\mathcal{X}')$ then $\mu(A) = 0$. Second, for any $A \in NR$ define $\pi(\cdot|A)$ as follows. If $\mu(A) = 0$, then $\pi(\cdot|A)$ is any distribution with support contained in $\bar{A}$; and if $\mu(A) > 0$ then

$$\begin{aligned}\pi(a|A) = \frac{1}{\mu(A)}\psi\Big(\big\{(u, v) \quad : \quad & u(a) + v(a) > u(b) + v(b) \ \ \forall b \in A \setminus \{a\} \ \textit{and} \\ & U_{u,v}(A) > U_{u,v}(B) \ \ \forall B \in \mathcal{X}' \setminus \{A\}\big\}\Big).\end{aligned}$$

Then clearly π satisfies the equality in the lemma and it follows from Lemma 2 that $\pi(a|A) = 0$ whenever $a \notin \bar{A}$. □

The next step is the heart of the proof, so we state it as a separate proposition. Its proof is based on the max-flow min-cut duality theorem.[23]

[23] Proposition 9 implies Proposition 1 of Section 3 as a special case. Indeed, just take $Y = \mathcal{X}$ and $h(A) = A$ for every $A \in \mathcal{X}$. The characterizing condition becomes $\lambda \in Core(v_\mu)$.

Proposition 9. Let Y be a finite set, $\mu \in \Delta(Y)$, $\lambda \in \Delta(X)$, and fix a mapping $h : Y \to \mathcal{X}$ where $\mathcal{X} = 2^X \setminus \emptyset$. There is a conditional probability system $\pi = \{\pi(\cdot|y) \in \Delta(h(y))\}_{y \in Y}$ such that $\lambda(a) = \sum_{y \in Y} \mu(y)\pi(a|y)$ for every $a \in X$ if and only if for every $A \in \mathcal{X}$

$$\sum_{a \in A} \lambda(a) \geq \sum_{\{y \in Y:\ h(y) \subseteq A\}} \mu(y).$$

Proof. 'Only if': Suppose that there is $\pi = \{\pi(\cdot|y) \in \Delta(h(y))\}_y$ such that $\lambda(a) = \sum_y \mu(y)\pi(a|y)$ for every $a \in X$ and fix a menu $A \in \mathcal{X}$. Then

$$\sum_{a \in A} \lambda(a) = \sum_{a \in A} \sum_{y} \mu(y)\pi(a|y) \geq \sum_{\{y:\ h(y) \subseteq A\}} \mu(y) \sum_{a \in A} \pi(a|y) = \sum_{\{y:\ h(y) \subseteq A\}} \mu(y),$$

so that the required collection of inequalities is satisfied.

'If': Consider a directed bipartite graph with sets of nodes $V_1 = X$ and $V_2 = Y$, and where there is an edge from $a \in V_1$ to $y \in V_2$ if and only if $a \in h(y)$. Add two additional nodes, a source and a sink, denoted s and t respectively. Also add edges from s to every $a \in V_1$ and from every $y \in V_2$ to t. Denote by $G = (V, E)$ the resulting augmented graph. For every $e \in E$ set a capacity $c(e) \in \mathbb{R}_+$ as follows: If $a \in X$ then $c(s, a) = \lambda(a)$; If $y \in Y$ then $c(y, t) = \mu(y)$; and $c(a, y) = 1$ for every edge in the original bipartite graph.

We claim that, under the assumption of the proposition, any cut $E' \subseteq E$ that separates s from t has total capacity of at least 1. This is clearly the case if one of the (a, y) edges is in E'; if not, then for every $a \in X$ either $(s, a) \in E'$ or $\{(y, t)\ :\ a \in h(y)\} \subseteq E'$

(or both). Let $A_0 = \{a \in X \ : \ (s,a) \in E'\}$. Then the total capacity of E' satisfies

$$\begin{aligned}\sum_{e\in E'} c(e) &\geq \sum_{a\in A_0} c(s,a) + \sum_{\{y:\ h(y)\not\subseteq A_0\}} c(y,t) \\ &= \sum_{a\in A_0} \lambda(a) + 1 - \sum_{\{y:\ h(y)\subseteq A_0\}} c(y,t) \\ &= \sum_{a\in A_0} \lambda(a) + 1 - \sum_{\{y:\ h(y)\subseteq A_0\}} \mu(y) \\ &\geq 1,\end{aligned}$$

where the final inequality is by the assumption of the proposition.

Clearly, there exists a cut E' of capacity equal to 1. Thus, the optimal value of the min-cut program for G is 1. By the max-flow min-cut theorem of linear programming (see, for example, Theorem 7.13 in Kleinberg and Tardos [2006]), there exists a flow $f : E \to \mathbb{R}_+$ such that $\sum_{a\in X} f(s,a) = \sum_{y\in Y} f(y,t) = 1$. In particular, $f(s,a) = \lambda(a)$ and $f(y,t) = \mu(y)$ for all a and y. For every edge (a,y) define $\pi(a|y) = \frac{f(a,y)}{\mu(y)}$ (and $\pi(\cdot|y)$ arbitrarily on $h(y)$ when $\mu(y) = 0$). For every $a \in X$ we thus have

$$\lambda(a) = f(s,a) = \sum_y f(a,y) = \sum_y \mu(y)\pi(a|y),$$

where the second equality is by the flow conservation constraints at the node a, and the last equality is by the definition of $\pi(a|y)$. This completes the proof. □

We can now finish the proof of the Proposition 8.

'Only If': Suppose that (μ, λ) is $\mathcal{X}'$-TSC rationalizable. By Lemma 2, we have that $\mu(R(\mathcal{X}')) = 0$, which proves the first condition of the proposition. Next, let π be as in Lemma 3. Since the support of μ is contained in NR we can apply Proposition 9 with $Y = NR$ and $h(B) = \bar{B}$ for every $B \in NR$. We thus get that, for every $A \in \mathcal{X}$,

$$\sum_{a\in A} \lambda(a) \geq \sum_{\{B\in NR:\ \bar{B}\subseteq A\}} \mu(B) = \sum_{\{B\in\mathcal{X}':\ \bar{B}\subseteq A\}} \mu(B) = v_\mu^{\mathcal{X}'}(A),$$

where the inequality is from Proposition 9 , the first equality follows from $\mu(R(\mathcal{X}')) = 0$, and the last equality from the definition of the game $v_\mu^{\mathcal{X}'}$. It follows that $\lambda \in Core(v_\mu^{\mathcal{X}'})$ as needed.

'If': Suppose that the conditions of the proposition hold. Applying Proposition 9 with $Y = NR$ and $h(A) = \bar{A}$ for every $A \in NR$ we get that there exists $\pi = \{\pi(\cdot|A) \in \Delta(\bar{A})\}_{A\in NR}$ such that $\lambda(a) = \sum_{A\in NR} \mu(A)\pi(a|A)$. By Lemma 3, (μ, λ) is $\mathcal{X}'$-TSC rationalizable and the proof is complete.

Proof of Corollary 5

First, it is obvious that if λ is potentially–consistent given ξ then $\lambda(a) \leq \xi(a)$ for every $a \in X$. For the converse, fix ξ and λ that satisfy these inequalities. We describe an algorithm that stops after a finite number of steps and produces the desired μ.

Label the alternatives arbitrarily, $X = \{a_1, \ldots, a_n\}$. Given any $\mu \in \Delta(\mathcal{X})$, let $\xi_\mu \in [0, 1]^X$ be defined by $\xi_\mu(a_i) = \sum_{\{A:\ a_i\in A\}} \mu(A)$, and let $D_\mu = \{1 \leq i \leq n :\ \xi_\mu(a_i) < \xi(a_i)\}$ be the set of alternatives for which the probability of availability based on μ, $\xi_\mu(a_i)$, is strictly below the required level $\xi(a_i)$.

The algorithm starts with μ^0 whose support is just the singletons $\{a_i\}_{i=1}^n$, and where $\mu^0(a_i) = \lambda(a_i)$ for every i. Note that, by assumption, $\xi_{\mu^0} \leq \xi$.

For some $k = 1, 2, \ldots$ suppose that a distribution μ^{k-1} is given. If $D_{\mu^{k-1}} = \emptyset$ then stop. Otherwise, we now describe how to obtain μ^k.

Let $i^* = \min D_{\mu^{k-1}}$ so that in particular $\xi_{\mu^{k-1}}(a_{i^*}) < \xi(a_{i^*})$. We claim that there must exist a menu $\tilde{A}$ in the support of μ^{k-1} such that $a_{i^*} \notin \tilde{A}$. Indeed, if a_{i^*} is in every menu in the support of μ^{k-1} then $\xi_{\mu^{k-1}}(a_{i^*}) = 1$, contradicting $\xi_{\mu^{k-1}}(a_{i^*}) < \xi(a_{i^*})$. Let $t = \min\{\mu(\tilde{A}), \xi(a_{i^*}) - \xi_{\mu^{k-1}}(a_{i^*})\} > 0$. Define μ^k by $\mu^k(\tilde{A}) = \mu^{k-1}(\tilde{A}) - t$, $\mu^k(\tilde{A} \cup \{a_{i^*}\}) = \mu^{k-1}(\tilde{A} \cup \{a_{i^*}\}) + t$, and $\mu^k(A) = \mu^{k-1}(A)$ for every other menu A. In words, a mass of t is moved from the menu $\tilde{A}$ to the menu $\tilde{A} \cup \{a_{i^*}\}$.

We have that $\xi_{\mu^k}(a_{i^*}) = \xi_{\mu^{k-1}}(a_{i^*}) + t$, and $\xi_{\mu^k}(a_i) = \xi_{\mu^{k-1}}(a_i)$ for every $i \neq i^*$. Therefore, ξ_{μ^k} is an increasing sequence, bounded above by ξ. Moreover, if i^* is not removed from D_{μ^k}, then $\tilde{A}$ is removed from the support of μ^k, implying that after at most

2^{n-1} iterations an equality $\xi_{\mu^k}(a_{i^*}) = \xi(a_{i^*})$ is achieved. It follows that the algorithm terminates after at most $n \times 2^{n-1}$ iterations.

Let μ^* be the resulting final distribution. By construction, $\xi_{\mu^*}(a_i) = \xi(a_i)$ for every i. It is left to show that (μ^*, λ) is consistent. Note however that $\lambda \in Core(v_{\mu^0})$ and that $v_{\mu^k} \leq v_{\mu^{k-1}}$ for every k since μ^k is obtained from μ^{k-1} by shifting mass from a set to one of its supersets. Thus, $\lambda \in Core(v_{\mu^0}) \subseteq Core(v_{\mu^*})$, so we are done by Proposition 1.