



Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Identification in Stochastic Choice

Peter Caradonna and Christopher Turansick

Working Paper n. 723

This Version: February 23, 2026

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
<http://www.igier.unibocconi.it>

The opinions expressed in the working papers are those of the authors alone, and not those of the Institute, which takes non institutional policy position, nor those of CEPR, NBER or Università Bocconi.

Identification in Stochastic Choice*

Peter Caradonna[†] and Christopher Turansick[‡]

February 23, 2026

Abstract

We characterize the identified sets of a wide range of stochastic choice models, including random utility, various models of boundedly-rational behavior, and dynamic discrete choice. In each of these settings, we show two distributions over choice rules are observationally equivalent if and only if they can be obtained from one another via a finite sequence of simple swapping transforms. We leverage this to obtain complete descriptions of both the defining inequalities and extreme points of these identified sets. In cases where choice frequencies vary smoothly with some parameters, we provide a novel global-inverse result for practically testing identification.

1 Introduction

A fundamental question for any model of behavior is to what degree its underlying primitives can be uniquely determined from individual or aggregate-level data. Even when exact identification fails, knowledge of the identified set can provide crucial information for the design of policy, counterfactual simulations, and welfare analysis.¹

*We are grateful to Roy Allen, Jose Apesteguia, Pierpaolo Battigalli, Andrés Carvajal, Simone Cerreia-Vioglio, Christopher Chambers, Emel Filiz-Ozbay, Satoshi Fukuda, Christopher Kops, SangMok Lee, Jay Lu, Yusufcan Masatlioglu, Paulo Natenzon, Axel Niemeyer, Luciano Pomatto, Justus Preusser, Collin Raymond, Kota Saito, Fedor Sandomirskiy, Bumin Yenmez, Kemal Yildiz, and Xinhan Zhang as well as seminar participants at Bilkent, BRIC 11, BSE Summer Forum Workshop on Choice and Decision, EC’25, Italian Junior Workshop on Economic Theory 2025, Liverpool, RUD 2025, UC Davis, and WUSTL for discussions and helpful comments. This paper subsumes “Identifying Restrictions on the Random Utility Model”.

[†]Division of the Humanities and Social Sciences, Caltech. Email: ppc@caltech.edu.

[‡]Department of Decision Sciences and IGIER, Bocconi University. Email: christopher.turansick@unibocconi.it.

¹E.g. [Manski \(2007\)](#); [Agarwal and Somaini \(2018\)](#); [Kalouptside et al. \(2021\)](#). See also [Marschak \(1953\)](#) and [Allen and Rehbeck \(2019\)](#).

Despite their practical importance, the identification properties of many workhorse models of stochastic choice, including the random utility model, have remained poorly understood.² In response, an extensive literature has emerged, studying additional restrictions under which analysts can uniquely recover an underlying distribution of parameters governing behavior from choice share data.³

In this paper, we provide complete characterizations of the identified sets of a wide range of stochastic choice theories, under arbitrary supplemental restrictions. Our primary observation is that, in each of these settings, each set of empirically indistinguishable distributions (e.g. over preferences) can be straightforwardly described in terms of a simple family of transformations. We term these *Ryser swaps*, in light of a connection to the discrete tomography literature.⁴ In particular, we show two distributions generate identical choice shares over any menu if, and only if, they are related by a finite sequence of such swaps.

For expositional clarity, we focus our discussion on the random utility model. However our findings apply straightforwardly to many other settings, including to models of incomplete or boundedly-rational behavior, correlated decision-making, and dynamic stochastic choice.⁵

Example 1. A population of voters are faced with four potential urban planning policies. Policy a calls for regional high-speed rail expansion, and b for large-scale investment in local bus and subway systems, whereas c for moderate congestion-pricing, and d an expansion of bike lanes and pedestrian zones.

The population is comprised of four types of voters. The first are ‘big picture’ activists, who prefer large-scale investment in public transit and a centralized policies ($a \succ b \succ c \succ d$), whereas incrementalists prefer more funding to less, but prize more immediate, practical transit options, and a ground-up approach to policy ($b \succ a \succ d \succ c$). Similarly, suburban commuters prefer rail to buses and dislike taxes on their vehicles ($a \succ b \succ d \succ c$), whereas urban commuters primarily use public transit and prioritize less congested streets ($b \succ a \succ c \succ d$). We summarize these four preferences below.⁶

Suppose straw-poll data over various subsets of these policies is consistent with the societal distribution of preferences being $\mu = (\frac{1}{4}, \frac{1}{4}, \frac{3}{8}, \frac{1}{8})$, and consider now a new policy, likely

²See, e.g., [Fishburn \(1998\)](#); [McClellon \(2015\)](#); [Strzalecki \(2025\)](#).

³E.g. [Luce \(1959\)](#); [McFadden \(1972\)](#); [Gul and Pesendorfer \(2006\)](#); [Apesteguia et al. \(2017\)](#); [Manzini and Mariotti \(2018\)](#); [Filiz-Ozbay and Masatlioglu \(2023\)](#); [Yang and Kopylov \(2023\)](#); [Suleymanov \(2024\)](#). Similar questions arise in the context of demand invertibility, see [Berry and Haile \(2024\)](#).

⁴E.g. [Ryser \(1957\)](#); [Fishburn et al. \(1991\)](#); [Kong and Herman \(1999\)](#). These tools have found recent application in economic theory, see [He et al. \(Forthcoming\)](#).

⁵We discuss these extensions in [Section 7](#). See also [Online Appendix B](#) for further details.

⁶The observation that a distribution with full support on these four preferences is unidentified from choice frequency data is due to [Fishburn \(1998\)](#). Moreover, such examples are minimal: the random utility model is known to be identified whenever there are three or fewer alternatives, e.g. [Block and Marschak \(1959\)](#).

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
a	b	a	b
b	a	b	a
c	d	d	c
d	c	c	d

to benefit suburban commuters.⁷ What can be said about the size of this share of the population?

Our results show suburban commuters constitute anywhere between $\frac{1}{4}$ and $\frac{5}{8}$ of the population. Moreover, these bounds are tight: for any fraction $\frac{1}{4} \leq \alpha \leq \frac{5}{8}$, there is a distribution over these four preferences where the population share of suburban commuters is precisely α , and which yields identical vote shares to μ over *any* menu of policies.

The key feature in this example is that all four preferences rank a and b above c and d , but disagree about the relative rankings of the alternatives within these groupings. In particular, $\succ_3$ and $\succ_4$ can be obtained by taking $\succ_1$ and $\succ_2$, and exchanging their respective rankings between c and d , while preserving all other comparisons.

Given any distribution over these four preferences with full support, by subtracting some small, but equal amount of probability mass from either $\succ_1$ and $\succ_2$, or $\succ_3$ and $\succ_4$, and re-assigning it equally to the opposing pair (here, obtained by exchanging only the rankings of c and d) one obtains a modified distribution with identical choice probabilities on all menus. Intuitively, such transformations affect only the *correlation* between preferring a to b and c to d , but not the likelihood of these events.

In this setting, it turns out these are the *only* transformations which preserve all choice probabilities. As such, the minimal and maximal possible population frequencies for $\succ_3$ can be obtained by simply maximizing the amount of mass uniformly transferred from the pair $\{\succ_3, \succ_4\}$ to $\{\succ_1, \succ_2\}$ or vice-versa, yielding the extremal, empirically indistinguishable distributions $\underline{\mu} = (\frac{3}{8}, \frac{3}{8}, \frac{1}{4}, 0)$ and $\bar{\mu} = (0, 0, \frac{5}{8}, \frac{3}{8})$. ■

As illustrated by [Example 1](#), failures of identification arise for a straightforward reason: choice probabilities do not, in general, uniquely determine the *correlation* between choices.⁸ Indeed, the identified sets of the random utility model are spanned precisely by the possible correlation patterns that cannot be ruled out by an empiricist with access only to aggregate, rather than individual, level data.

⁷That is, voters who hold $\succ_3$.

⁸For instance, in [Example 1](#), the correlation between choices on menus $\{a, b\}$ and $\{c, d\}$.

Our first main result characterizes the full collection of preference distributions consistent with a given dataset. We show any transformation of a distribution which preserves all choice probabilities is equivalent to applying a finite sequence of simple swaps of the form considered in [Example 1](#). Using this representation, we provide a straightforward, geometric description of the random utility model’s identified sets, both in terms of their defining inequalities, and their extreme points. As consumer welfare measures typically rely on preference information ([Deaton 1980](#); [McKenzie 1983](#)), this can be used, e.g., to obtain tight bounds on the estimated aggregate or distributional effects of policy interventions.⁹

We then investigate how our results specialize under common, structured classes of supplemental restrictions. For example, in many instances an analyst may wish to a priori restrict the *set* of preferences assumed held by a population, but not their relative frequencies.¹⁰ We characterize identified sets subject to arbitrary support restrictions, as well as those families of preferences $\mathcal{S}$ with the property that any data is consistent with at most a single measure whose support belongs to $\mathcal{S}$. In practice, $\mathcal{S}$ often consists of restrictions of some parametric family of utilities.¹¹ In this case, our results provide an exact test for practitioners to determine when choice probabilities uniquely identify a population distribution over parameters.

We also consider cases in which choice probabilities are specified directly, as smooth functions of some vector of parameters. Using topological techniques, we provide a pair of necessary and sufficient conditions for the global invertibility of general, non-linear systems. In contrast to many existing results, we require no monotonicity properties.¹² Thus, for example, applied to the problem of determining the invertibility of a smooth demand, our results are applicable even when goods may fail to be substitutes, and in settings where the law of demand may fail. We demonstrate the practical applicability of our conditions by proving a novel result, that the dynamic, logit habit formation model of [Turansick \(2025\)](#) is not identified from unconditional choice data.

⁹See, e.g., [Deb et al. \(2023\)](#). [Tebaldi et al. \(2023\)](#) and [Kamat and Norris \(2025\)](#) consider similar questions in the context of attribute variation. Our results may be seen as providing complementary tools for settings with menu variation, e.g. [McFadden \(2005\)](#); [Kitamura and Stoye \(2018\)](#), as well as to dynamic, or imperfectly rational, models of behavior; see [Section 7](#).

¹⁰For example, expected utility preferences on domains of lotteries ([Gul and Pesendorfer 2006](#), [Apesteguia and Ballester 2018](#)), exponentially discounted or quasi-hyperbolic preferences over streams ([Lu and Saito 2018](#)), subjective expected utility over acts ([Lu 2016](#); [Duraj 2018](#)), or single-peaked preferences on ordered domains ([Apesteguia et al. 2017](#)). Such restrictions are also implicit in more general, dynamic settings, e.g. [Rust \(1987\)](#); [Lu et al. \(2024\)](#); see also [Frick et al. \(2019\)](#).

¹¹E.g. [Apesteguia and Ballester \(2018\)](#) consider the case of random CARA or CRRA expected utility.

¹²C.f. [Gale and Nikaido \(1965\)](#); [Beckert and Blundell \(2004\)](#); [Simsek et al. \(2005\)](#); [Berry et al. \(2013\)](#); [Lindenlaub \(2017\)](#); [Wang \(2021\)](#); [Allen \(2022\)](#). See also ‘dominant diagonal’ conditions, e.g. [McKenzie \(1960\)](#); [Rosen \(1965\)](#).

Finally, when the set of alternatives is ordered, an elegant approach due to [Apestegui et al. \(2017\)](#) shows that requiring single-crossing structure on the support of a rationalization yields identification.¹³ However, such rationalizations need not always exist, even for data compatible with the random utility model. Using our notion of a Ryser swap, we provide a principled generalization we term *swap progressivity*. Whenever the data admit a single-crossing representation, this will always be the unique swap-progressive rationalization. However, we show a unique swap-progressive representation exists for *any* data compatible with the random utility model. This provides a novel, practical identification strategy that incorporates the natural ordering over alternatives.

The paper proceeds as follows. In [Section 2](#), we formally introduce the random utility model. In [Section 3](#), we formalize our core notion of a Ryser swap, and present our main geometric characterization. [Section 4](#) provides a dual characterization of identified sets in terms of their extreme points, and characterizes identifying support restrictions. In [Section 5](#), we introduce our notion of swap progressivity and establish its properties, and [Section 6](#) considers the identification of general, parametric stochastic choice models. Finally, [Section 7](#) discusses extensions to other environments. To more concretely relate our contributions to existing work, we defer our discussion of related literature to [Section 8](#). [Section 9](#) concludes.

2 The Random Utility Model

Let X denote a fixed, finite set of alternatives over which an individual chooses. A **preference** is a linear order on X ; we denote the set of all preferences by $\mathcal{L}$.¹⁴ To conserve on notation, we will often write preferences as strings of alternatives, arranged in descending order, i.e. ‘ $abcd$ ’ denotes the ranking $a \succ b \succ c \succ d$. For any preference $\succ$ and any $0 \leq k \leq |X|$, the k -**initial** and k -**terminal** segments of $\succ$ are the ordered strings consisting of the k most-preferred alternatives and the $|X| - k$ least-preferred.¹⁵ We will denote these by $s_k^\uparrow(\succ)$ and $s_k^\downarrow(\succ)$ respectively. For example, if $\succ$ corresponds to the ordering $abcde$, then:

$$s_2^\uparrow(\succ) = ab \quad \text{and} \quad s_3^\downarrow(\succ) = cde.$$

A function $\rho : X \times 2^X \setminus \{\emptyset\} \rightarrow [0, 1]$ defines a **random choice rule** if, for all non-empty subsets $A \subseteq X$,

$$\sum_{x \in A} \rho(x, A) = 1.$$

¹³See also [Filiz-Ozbay and Masatlioglu \(2023\)](#); [Apestegui and Ballester \(2023\)](#).

¹⁴A linear order is a complete, transitive, and antisymmetric binary relation on X .

¹⁵For $k = 0$ (resp. $k = |X|$) we define $s_k^\uparrow(\succ)$ (resp. $s_k^\downarrow(\succ)$) as the empty string.

Random choice rules are assumed observable; they constitute the basic data in our identification problem. For any finite set A , we use $\Delta(A)$ to denote the set of probability measures over A . A **restriction** of the random utility model is any subset $\mathcal{M} \subseteq \Delta(\mathcal{L})$. We say a random choice rule is rationalizable subject to a restriction $\mathcal{M}$ if there exists some probability distribution $\mu \in \mathcal{M}$ such that, for all $x \in A \subseteq X$, we have:

$$\rho(x, A) = \mu\{\succ \in \mathcal{L} \mid x \text{ is maximal in } A\} = \sum_{\succ \in \mathcal{L}} \mu(\succ) \mathbb{1}_{\{x \succ A \setminus x\}}.$$

Under the interpretation of a restriction $\mathcal{M}$ as describing a set of possible makeups of a heterogeneous society, a collection of observed choice frequencies ρ are $\mathcal{M}$ -rationalizable if and only if they arise as the distribution of constrained-optimal outcomes according to some population composition $\mu \in \mathcal{M}$.

We say two measures $\mu, \nu \in \Delta(\mathcal{L})$ are **observationally equivalent** if, for all $x \in A \subseteq X$,

$$\mu\{\succ \in \mathcal{L} \mid x \text{ is maximal in } A\} = \nu\{\succ \in \mathcal{L} \mid x \text{ is maximal in } A\}, \quad (1)$$

i.e. they generate identical choice frequencies on every choice set $\emptyset \subsetneq A \subseteq X$. Finally, a restriction $\mathcal{M}$ is **identifying** if it contains no pair of distinct, observationally equivalent measures.

3 The Geometry of Identification

The unrestricted random utility model (i.e. $\mathcal{M} = \Delta(\mathcal{L})$) has long been known to be unidentified.¹⁶ Thus, to each distribution $\mu \in \Delta(\mathcal{L})$, there corresponds some generally non-singleton collection of observationally equivalent distributions.

Let $\mathcal{P}$ denote the set of random choice rules on X , and define the mapping $\Phi : \Delta(\mathcal{L}) \rightarrow \mathcal{P}$ via:

$$\Phi(\mu)_{(x,A)} = \begin{cases} \mu\{\succ \in \mathcal{L} \mid x \text{ is maximal in } A\} & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \quad (2)$$

By (1) two distributions $\mu, \nu \in \Delta(\mathcal{L})$ are observationally equivalent if and only if $\Phi(\mu) = \Phi(\nu)$. However, as Φ is linear in μ , there exists some subspace $\mathcal{R} \subseteq \mathbb{R}^{\mathcal{L}}$ such that, whenever $\Phi(\mu) = \rho$, the identified set $\Phi^{-1}(\rho)$ is precisely $(\mu + \mathcal{R}) \cap \Delta(\mathcal{L})$.¹⁷ In particular, every identified set is fully characterized by this subspace.

¹⁶See Fishburn (1998).

¹⁷Here we identify $\mathbb{R}^{\mathcal{L}}$ with the space of all signed measures over linear orders on X .

3.1 Compatible Pairs and the Ryser Subspace

We say that a pair of preferences are **k -compatible**, $0 \leq k \leq |X|$, if both preferences agree on the set of k -best alternatives.¹⁸ In other words, $\succ_1$ and $\succ_2$ are k -compatible if $s_k^\uparrow(\succ_1)$ can be obtained by permuting the terms of the sequence $s_k^\uparrow(\succ_2)$, and analogously $s_k^\downarrow(\succ_1)$ from $s_k^\downarrow(\succ_2)$. This ensures the concatenated sequences:

$$\succ'_1 = s_k^\uparrow(\succ_1) s_k^\downarrow(\succ_2) \quad \text{and} \quad \succ'_2 = s_k^\uparrow(\succ_2) s_k^\downarrow(\succ_1) \quad (3)$$

define valid preferences. We refer to the pair of preferences obtained in this manner as the **k -conjugates** of $\succ_1$ and $\succ_2$. If $\succ_1$ and $\succ_2$ disagree on both (i) the ranking of the k -best alternatives and (ii) the ranking of the $(|X| - k)$ -worst, $\succ'_1$ and $\succ'_2$ will be distinct from either $\succ_1$ or $\succ_2$; in this case we refer to $\succ_1$ and $\succ_2$ as **non-trivially** k -compatible.

Our interest in compatible pairs of preferences is motivated by the following proposition, which says that any uniform distribution supported on a compatible pair of preferences is behaviorally equivalent to the uniform distribution over the pair obtained via (3).

Proposition 1. Let $\succ_1, \succ_2 \in \mathcal{L}$ be k -compatible, with k -conjugates $\succ'_1$ and $\succ'_2$. Then the uniform distributions on $\{\succ_1, \succ_2\}$ and $\{\succ'_1, \succ'_2\}$ are observationally equivalent.

As the following example illustrates, the observational indeterminacy highlighted by Proposition 1 is a consequence of the fact that a joint distribution is not, in general, uniquely determined by the requirement that it possess uniform marginals.

Example 2. Consider again Example 1, where $X = \{a, b, c, d\}$, and there are four preferences: $\succ_1: abcd$, $\succ_2: badc$, $\succ_3: abdc$, and $\succ_4: bacd$. Note that the pair $\succ_1$ and $\succ_2$ are 2-compatible, with 2-conjugates $\succ_3$ and $\succ_4$.¹⁹ We may view these four preferences as elements of a product set $\{ab, ba\} \times \{cd, dc\}$, by associating an ordered pair of strings with their concatenation, e.g.:

$$(ab, cd) \mapsto abcd.$$

A joint distribution over $\{ab, ba\} \times \{cd, dc\}$ generates the same choice probabilities on every menu as the uniform distribution on $\{\succ_1, \succ_2\}$ if and only if both of its marginals are themselves uniform.²⁰ Thus, in this setting, Proposition 1 asserts the observational equivalence of the two extremal joint measures with uniform marginals: μ_{12} , the uniform on $\{\succ_1, \succ_2\}$, which perfectly correlates the choice of a from $\{a, b\}$ with c from $\{c, d\}$, and μ_{34} , the uniform on $\{\succ_3, \succ_4\}$ which perfectly anti-correlates these choices. See Figure 1. ■

¹⁸In particular, any pair of preferences is vacuously 0-compatible.

¹⁹That is, $\succ_3 = s_2^\uparrow(\succ_1) s_2^\downarrow(\succ_2)$, and $\succ_4 = s_2^\uparrow(\succ_2) s_2^\downarrow(\succ_1)$.

²⁰This can be verified by direct computation or, e.g., follows from Theorem 5 of Falmagne (1978).

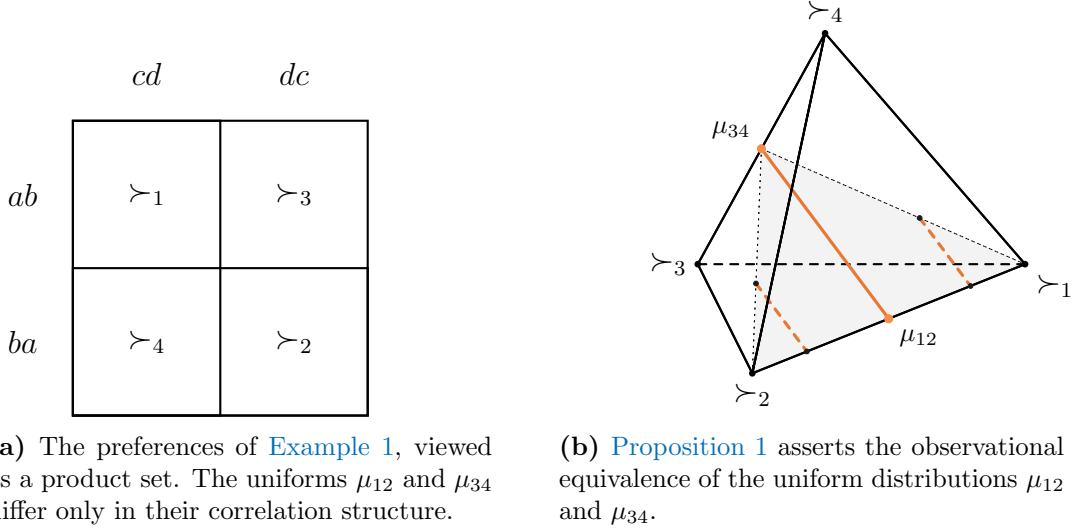


Figure 1: By Proposition 1, μ_{12} and μ_{34} are observationally equivalent. By the linearity of (2), both are also equivalent to any distribution in their convex hull, and any pair of distributions belonging to some common translation of this set (e.g. the dashed segments) are observationally equivalent.

We say a signed measure $R \in \mathbb{R}^{\mathcal{L}}$ defines a **Ryser swap** if:

$$R = \mathbb{1}_{\{\succ'_1, \succ'_2\}} - \mathbb{1}_{\{\succ_1, \succ_2\}},$$

where $\succ_1$ and $\succ_2$ are k -compatible, and $\succ'_1$ and $\succ'_2$ are k -conjugate for some k .²¹ A signed measure is a **weighted Ryser swap** if it is proportional to a Ryser swap. Adding a weighted Ryser swap to some initial distribution μ may be regarded as a compound transformation, which first separates out from μ a product measure, applies to it a marginal-preserving transformation, then recombines the result with the remaining mass from μ to obtain a modified measure μ' . As in Example 1, whenever μ' is a probability distribution, it differs from μ only in the *correlation* between choices, not their frequencies.

Finally, we define the **Ryser subspace** $\mathcal{R} \subseteq \mathbb{R}^{\mathcal{L}}$ as the span of the Ryser swaps:

$$\mathcal{R} = \text{span}\{R \in \mathbb{R}^{\mathcal{L}} \mid R \text{ is a Ryser swap}\}.$$

Every signed measure in $\mathcal{R}$ corresponds to a transformation which applies a finite sequence of weighted swaps. By linearity of Φ , given $\bar{R} \in \mathcal{R}$ and $\mu \in \Delta(\mathcal{L})$, if $\mu + \bar{R}$ defines a probability distribution, Proposition 1 implies it is observationally equivalent to μ .²² Our next theorem

²¹Note $\succ_1$ and $\succ_2$ are trivially k -compatible if and only if $R = 0$.

²²Proposition 1 asserts that, given compatible $\succ_1, \succ_2$ with conjugates $\succ'_1, \succ'_2$, the uniforms μ_{12} and $\mu_{1'2'}$ are observationally equivalent, and hence $\Phi(\mu_{12}) = \Phi(\mu_{1'2'})$. By linearity:

$$\Phi(\mu + \varepsilon(\mu_{1'2'} - \mu_{12})) = \Phi(\mu) + \varepsilon(0) = \Phi(\mu).$$

shows that, in fact, *every* pair of observationally equivalent distributions are related in this manner.

Theorem 1. Two distributions $\mu, \mu' \in \Delta(\mathcal{L})$ are observationally equivalent if and only if $\mu - \mu' \in \mathcal{R}$. In particular, for any $\mu \in \mathcal{M} \subseteq \Delta(\mathcal{L})$, the set of observationally equivalent distributions in $\mathcal{M}$ is precisely $(\mu + \mathcal{R}) \cap \mathcal{M}$.

[Theorem 1](#) shows that not only do Ryser swaps preserve choice probabilities, they in fact generate the set of *all* transformations which do so. In other words, two distributions over preferences are observationally equivalent if and only if the first can be generated from the second (or vice-versa) by a finite sequence of such swaps. Geometrically, this implies the set of distributions observationally equivalent to some $\mu \in \Delta(\mathcal{L})$ is simply the intersection of the affine subspace $\mu + \mathcal{R}$ with $\Delta(\mathcal{L})$.

This provides a practical means of computing features of the identified set. Since $\mathcal{R}$ does not depend on the data, once a choice of X is specified, a basis $R_1, \dots, R_N$ for $\mathcal{R}$ can be obtained from the set of Ryser swaps. Then, given any rationalization $\mu \in \Delta(\mathcal{L})$, optimal bounds on, e.g., the sizes of different shares of the population across all rationalizations, can be computed by evaluating a straightforward linear program over the weights $\alpha_1, \dots, \alpha_N$ assigned to these basis elements, subject only to the constraint $\mu + \sum_i \alpha_i R_i \geq \mathbf{0}$.²³

To illustrate, consider again [Example 1](#). As there are only four alternatives, any non-trivially compatible pair must agree on the best (and worst) two alternatives, but not their ordering. As a consequence, if the measure $\mu + R$ is a probability distribution for some $R \in \mathcal{R}$, R must subtract mass only from $\succ_1, \dots, \succ_4$, and reassign it to preferences formed by matching terminal segments $s_{\downarrow}^2 \in \{cd, dc\}$ to initial segments $s_{\uparrow}^2 \in \{ab, ba\}$. But $\succ_1, \dots, \succ_4$ are the *only* preferences formed this way. It follows R must be proportional to the sole (up to sign) non-trivial Ryser swap supported on these preferences, $\bar{R} = \mathbb{1}_{\{\succ_1, \succ_2\}} - \mathbb{1}_{\{\succ_3, \succ_4\}}$. In particular, this implies the orange line segments in [Figure 1](#) are precisely the identified sets.²⁴

Example 3. In more complex environments, elements of $\mathcal{R}$ may have richer form. Suppose $X = \{a, b, c, d, e, f\}$, and consider the following six preferences.

Since every $\bar{R} \in \mathcal{R}$ is a weighted sum of Ryser swaps, μ and $\mu + \bar{R}$ are observationally equivalent.

²³One could evaluate a similar linear program over $\Delta(\mathcal{L})$ using $\Phi(\cdot) = \rho$ as a constraint, without appealing to [Theorem 1](#) (e.g. [Deb et al. 2023](#)). However, doing so requires $|X| - 1$ variables and hence may become computationally difficult in larger environments. In contrast, identified sets are generally far lower dimensional; in [Example 1](#), $\Delta(\mathcal{L})$ is 23-dimensional whereas the identified set was 1-dimensional. Thus the explicit description provided by [Theorem 1](#) may allow for significant computational savings in practice.

²⁴Indeed, given some rationalization μ , tight bounds on *any* property of the identified set can be computed simply by optimizing $\mu + \alpha \bar{R}$ over the scalar α , subject to the constraint this vector belong to the simplex.

$\succ_1$	$\succ_2$	$\succ_3$	$\succ'_1$	$\succ'_2$	$\succ'_3$
a	b	c	a	b	c
b	a	d	b	a	d
c	e	b	e	c	b
d	f	a	f	d	a
e	c	f	c	f	e
f	d	e	d	e	f

Here, μ_{123} , the uniform distribution on $\{\succ_1, \succ_2, \succ_3\}$, is observationally equivalent to μ'_{123} , the uniform on $\{\succ'_1, \succ'_2, \succ'_3\}$. However, unlike in [Example 1](#), these distributions do not differ by any single weighted Ryser swap.²⁵ Nevertheless, they do so by a *sequence* of such swaps.

$\succ_1$	$\succ_2$	$\succ_3$		$\succ'_1$	$\succ$	$\succ_3$		$\succ'_1$	$\succ'_2$	$\succ'_3$
a	b	c		a	b	c		a	b	c
b	a	d		b	a	d		b	a	d
c	e	b	$\rightarrow$	e	c	b	$\rightarrow$	e	c	b
d	f	a		f	d	a		f	d	a
e	c	f		c	e	f		c	f	e
f	d	e		d	f	e		d	e	f

To construct such a sequence, first transfer all mass from $\succ_1$ and $\succ_2$ under μ_{123} to their 2-conjugates (the pair formed by swapping their terminal segments $cdef$ and $efcd$). Label these resulting preferences $\succ'_1$ and $\succ$. Then, transfer all mass on $\succ$ and $\succ_3$ to their 4-conjugates (the pair obtained from exchanging ef and fe) to yield μ'_{123} . ■

4 Support Restrictions and Extreme Points

Perhaps the simplest class of restrictions are those in which an analyst constrains the set of preferences held in a population. We say $\mathcal{M} \subseteq \Delta(\mathcal{L})$ defines a **support restriction** if:

$$\mathcal{M} = \{\mu \in \Delta(\mathcal{L}) : \text{supp}(\mu) \subseteq \mathcal{S}\},$$

for some set of preferences $\mathcal{S} \subseteq \mathcal{L}$. Such restrictions are natural when a modeler wishes to a priori restrict the set of preferences, but not their relative frequencies.

²⁵To see this, note any Ryser swap between any pair of these six preferences is either trivial or places non-zero mass on some preference outside the set.

Example 4. Let X consist of the following four monetary lotteries:

$$\begin{aligned} a &= \frac{1}{2}[\$10] + \frac{1}{2}[\$20], & b &= \frac{1}{2}[\$5] + \frac{1}{2}[\$25] \\ c &= \frac{1}{4}[\$10] + \frac{3}{4}[\$20], & d &= \frac{1}{4}[\$5] + \frac{3}{4}[\$25]. \end{aligned}$$

Suppose we wish to consider only those preferences $\mathcal{S} \subseteq \mathcal{L}$ which are consistent with some (monotone) expected utility representation. To determine this subset, note $c \succsim_{FOSD} a$ and $d \succsim_{FOSD} b$, hence any preference in $\mathcal{S}$ ranks $c \succ a$ and $d \succ b$.²⁶ Moreover,

$$c = \frac{1}{2}[a] + \frac{1}{2}[\$20] \quad \text{and} \quad d = \frac{1}{2}[b] + \frac{1}{2}[\$25],$$

thus every preference in $\mathcal{S}$ which ranks $b \succ a$, also obeys $d \succ c$. These are all the restrictions imposed by monotonicity and the expected utility axioms, hence $\mathcal{S}$ consists of the five orders:

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$
d	d	d	c	c
c	c	b	d	a
b	a	c	a	d
a	b	a	b	b

listed here. ■

While [Example 4](#) highlights the case of expected utility preferences, one could likewise consider other restrictions, from parametric sub-families such as CARA or CRRA, to broader classes of risk preferences, such as betweenness or disappointment aversion. Similarly, many other environments suggest natural restrictions, such as exponentially discounted preferences over consumption streams, or subjective expected utility preferences over acts.

The question we now consider is how such restrictions affect the set of rationalizing distributions, and when the probability weights on these and other families of orders can be uniquely recovered from choice frequency data alone.

Let $\succ_1, \dots, \succ_N$ be a (finite) sequence of preferences, allowing repetition. For each $1 \leq k \leq |X|$, let Π_k denote the partition of $\{1, \dots, N\}$ such that i and j belong to the same partition element if and only if $\succ_i$ and $\succ_j$ are $(k-1)$ -compatible.²⁷ A **rearrangement** is a collection of permutations $\{\sigma_k\}_{k=1}^{|X|}$ of $\{1, \dots, N\}$, such that σ_k fixes each partition element of Π_k . In other words, a rearrangement is a choice of permutations, where each σ_k exchanges

²⁶Here, $\succsim_{FOSD}$ corresponds to first order stochastic dominance.

²⁷In particular, as every pair of preferences are trivially 0-compatible, we always have $\Pi_1 = \{\{1, \dots, N\}\}$.

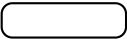
the labels of preferences in $\succ_1, \dots, \succ_N$, but only among those which agree on the set of $(k-1)$ best alternatives.

Let x_n^k denote the k -th most preferred alternative of $\succ_n$. We say two sequences $\succ_1, \dots, \succ_N$ and $\succ'_1, \dots, \succ'_N$ are **rearrangement equivalent** if there exists a rearrangement $\{\sigma_k\}_{k=1}^{|X|}$ such that, for all $1 \leq n \leq N$:

$$\succ'_n = x_{\sigma_1^{-1}(n)}^1 x_{(\sigma_2 \circ \sigma_1)^{-1}(n)}^2 \cdots x_{(\sigma_{|X|} \circ \cdots \circ \sigma_1)^{-1}(n)}^{|X|},$$

Rearrangement equivalence describes when one sequence of preferences can be transformed into the other by iteratively replacing compatible pairs with their conjugates. As the following example illustrates, this has a simple visual intuition: two sequences are rearrangement equivalent precisely when the strings of one can be ‘braided together’ to form the other.

Example 5. Suppose $X = \{a, b, c, d, e, f\}$, and consider again the sequence of preferences $\succ_1, \succ_2, \succ_3$ from [Example 3](#) given by:

$\succ_1$	$\succ_2$	$\succ_3$	Π		$\succ'_1$	$\succ'_2$	$\succ'_3$
a	b	c			a	b	c
b	a	d	* * *		b	a	d
c	e	b	 *	$\xrightarrow{\sigma}$	e	c	b
d	f	a	* * *		f	d	a
e	c	f	 * 		c	f	e
f	d	e	* * *		d	e	f

We also plot the partitions $\Pi_1, \dots, \Pi_6$ of $\{1, 2, 3\}$ defined above, where $*$ denotes any singleton partition element. The sequence $\succ'_1, \succ'_2, \succ'_3$ is rearrangement equivalent to $\succ_1, \succ_2, \succ_3$, under the collection of permutations where $\sigma_1, \sigma_2, \sigma_4$, and σ_6 are the identity, and σ_3 (resp. σ_5) swaps the two entries of the unique non-singleton partition element of Π_3 (resp. Π_5). To see this, note that σ_1 and $(\sigma_2 \circ \sigma_1)$ are the identity permutation $(1, 2, 3)$, while $(\sigma_3 \circ \sigma_2 \circ \sigma_1) = (\sigma_4 \circ \cdots \circ \sigma_1) = (2, 1, 3)$.²⁸ Finally, $(\sigma_5 \circ \cdots \circ \sigma_1) = (\sigma_6 \circ \cdots \circ \sigma_1) = (3, 1, 2)$. Thus, for example, the fifth-best alternative under $\succ'_2$ is the same as for $\succ_3$, as $(\sigma_5 \circ \cdots \circ \sigma_1) = (3, 1, 2)$ and therefore $(\sigma_5 \circ \cdots \circ \sigma_1)^{-1} = (2, 3, 1)$. ■

When $\mathcal{M}$ is a support restriction, (2) implies the set of measures in $\mathcal{M}$ observationally equivalent to some $\mu \in \Delta(\mathcal{L})$ forms a polytope. Our next result shows that an observationally

²⁸We use one-line notation for permutations, e.g., $(2, 3, 1)$ represents the permutation $\sigma(1) = 2$, $\sigma(2) = 3$, and $\sigma(3) = 1$. In particular, this means $\sigma^{-1}(n)$ corresponds to which component of the vector has value n .

equivalent $\mu' \in \mathcal{M}$ is an extreme point of this set if and only if its support contains no pair of distinct, rearrangement-equivalent sequences.²⁹ In particular, such a pair of sequences exist if and only if there are a pair of distinct rationalizations in $\mathcal{M}$ which average to μ' .

Theorem 2. Suppose $\mathcal{M}$ is a support restriction. Then $\mu \in \mathcal{M}$ is an extreme point of the identified set $(\mu + \mathcal{R}) \cap \mathcal{M}$ if and only if $\text{supp}(\mu)$ contains no pair of distinct, rearrangement-equivalent sequences.

Whereas [Theorem 1](#) characterizes the identified sets of the random utility model in terms of Ryser swaps, in the case of support restrictions, [Theorem 2](#) provides a complementary description via their extreme points.³⁰ These extremal distributions are of natural interest, as they constitute the primitive, irreducible rationalizations from which any others may be constructed. In practice, they often correspond to the simplest, or most parsimonious, population models which explain the data. For example, single-crossing ([Apesteguia et al. 2017](#)), progressive ([Filiz-Ozbay and Masatlioglu 2023](#)), and swap-progressive ([Section 5](#)) random utility representations are always extreme points of their identified sets.³¹

[Theorem 2](#) also resolves an open question in the theory of rational *joint* choice behavior.³² [Kashaev et al. \(2024\)](#) show that, while the revealed preference implications of such models are generally intractable, they can be simply axiomatized under any restriction to sets of preferences with linearly independent choice functions.³³ To date, no characterization of these sets has been known. However, a key ingredient in our proof is showing they are precisely the sets which contain no pair of distinct, rearrangement-equivalent sequences. Thus our results also describe when testing for stochastically rational joint behavior is easy.

In addition, this allows us to completely describe identifying support restrictions.

Corollary 1. A support restriction defined by $\mathcal{S} \subseteq \mathcal{L}$ is identifying if and only if $\mathcal{S}$ contains no pair of distinct, rearrangement-equivalent sequences.

To illustrate, consider again the support restriction $\mathcal{S}$ consisting of the four preferences in [Example 1](#). As noted before, the only admissible permutation in any rearrangement that can generate new preferences is σ_2 , thus two sequences are rearrangement equivalent if and only if one can be obtained from the other by shuffling terminal segments $s_{\downarrow}^2$. As a consequence,

²⁹That is, one cannot construct two such sequences by drawing (with replacement) from $\text{supp}(\mu')$.

³⁰When $\mathcal{M}$ is a support restriction, [Theorem 1](#) and [Theorem 2](#) provide dual characterizations of the identified polytopes, in terms of their faces and extreme points, respectively. We note, however, that [Theorem 1](#) remains valid without assumptions on $\mathcal{M}$.

³¹See [Proposition 2](#). However, there exist identified sets whose extreme points do not all arise in this manner; see [Proposition 3](#).

³²E.g. [Chambers et al. \(2024\)](#).

³³Here, we are identifying choice functions with their representation as 0–1 vectors in $\mathcal{P}$.

while $\mathcal{S}$ is not identifying, any proper subset is. In particular, a distribution is an extreme point of its identified set if and only if it does not have full support; see [Figure 1](#).

Similarly, [Corollary 1](#) also implies that the restriction to monotone expected utility preferences in [Example 4](#) is identifying. As in [Example 1](#), it once again suffices to consider only rearrangements with σ_2 the sole non-trivial permutation. As a consequence, proving identification reduces to verifying one cannot construct two distinct sequences of preferences from $\mathcal{S}$ with identical initial 2-segments, but whose corresponding sequences of terminal 2-segments are distinct. But such sequences cannot exist, essentially due to the fact each preference in $\mathcal{S}$ which ranks $b \succ a$ does not also rank $c \succ d$. Thus $\mathcal{S}$ is identifying.

5 Order-Based Restrictions

In many economic environments, the underlying set of alternatives carries a natural order structure. For example, monetary lotteries may be ordered by various measures of risk or stochastic dominance, vehicles by gas mileage, or government policies by level of public good provision. In such contexts, behavioral types often possess an analogous ordering, e.g. by risk aversion, environmental conscientiousness, or civic-mindedness.

Throughout this section, we will assume that $\succeq$ is a linear order on X . Following [Apesteguia et al. \(2017\)](#), a distribution $\mu \in \Delta(\mathcal{L})$ is a **single-crossing** representation if $\text{supp}(\mu)$ can be ordered $\succ_1, \dots, \succ_N$ such that if (i) $x \succeq y$ and (ii) $x \succ_j y$, then for all $i > j$, $x \succ_i y$ as well. When a single-crossing representation exists, [Apesteguia et al. \(2017\)](#) show is necessarily unique; however there exist choice frequencies which are consistent with the random utility model but incompatible with any single-crossing representation.

An alternative approach to identification due to [Filiz-Ozbay and Masatlioglu \(2023\)](#) considers a broader class of primitives. Let $\mathcal{C}$ denote the set of all choice functions, including those not rationalizable by any strict preference. A distribution $\mu \in \Delta(\mathcal{C})$ is said to be a **progressive** random choice representation if $\text{supp}(\mu)$ can be ordered $c_1, \dots, c_N$ such that $i > j$ implies $c_i(A) \succeq c_j(A)$ for all menus $\emptyset \subsetneq A \subseteq X$. In contrast to single-crossing representations, progressive random choice rules exist and are unique for any choice shares. However, they may not be supported on the rational choice functions, and hence fail to be a random utility representation, even for data compatible with the random utility model.

Our notion of compatibility provides a novel, order-based identifying restriction which preserves the best features of both the single-crossing and progressive frameworks. We say that a random utility representation $\mu \in \Delta(\mathcal{L})$ is **swap-progressive** if $\text{supp}(\mu)$ can be

ordered $\succ_1, \dots, \succ_N$ such that, whenever $\succ_i$ and $\succ_j$ are k -compatible, $0 \leq k \leq |X| - 1$, then $i > j$ implies $x_i^{k+1} \succeq x_j^{k+1}$.³⁴

Our next result shows swap-progressive representations are unique, and exist if and only if the data admit a random utility rationalization. Thus the empirical content of swap-progressivity coincides with that of the baseline random utility model. Moreover, when either single-crossing or progressive random utility representations exist, both do, and both coincide with the unique swap-progressive representation.³⁵

Theorem 3. Let $\succeq$ be a linear order on X . Then:

- (i) A swap-progressive rationalization exists if and only if the data are compatible with the random utility model. Moreover, if such a representation exists, it is unique.
- (ii) For any order $\succeq$, every single-crossing representation is swap-progressive. Conversely, a swap-progressive representation is single-crossing if and only if it is progressive.

In light of [Theorem 3](#), we interpret swap-progressivity as a principled refinement of both notions: it agrees with the random utility predictions of both single-crossing and progressivity whenever these exist, but generates unique predictions even when these other frameworks cannot. This provides a practical, but general, method for selecting rationalizations, that systematically incorporates the order structure of the environment.

In particular, in many applications it is valuable to obtain rationalizations whose support is ‘1-dimensional,’ or able to be linearly ordered according to some economic criterion.³⁶ As the following example illustrates, this is always true of swap-progressive rationalizations.³⁷

Example 6. Consider again the setting of [Example 1](#), and suppose policies are ranked according to their fiscal cost $a \succeq b \succeq d \succeq c$. [Theorem 3](#) guarantees the policy maker is able to recover a unique swap-progressive rationalization from straw-poll data, even though infinitely many rationalizations exist.

The power of swap progressivity is that it severely limits the support of a rationalization. For example, no swap-progressive distribution can place positive weight on both $\succ_1 = abcd$ and $\succ_2 = badc$. As both are trivially 0-compatible, $a \succeq b$ implies $\succ_1$ must be ranked higher

³⁴Recall that x_n^k denotes the k -th most preferred alternative of $\succ_n$, and note $x_i^{k+1} \succeq x_j^{k+1}$ explicitly allows for the equality of x_i^{k+1} and x_j^{k+1} .

³⁵The first part of this claim is Remark 1 in [Filiz-Ozbay and Masatlioglu \(2023\)](#).

³⁶See, e.g., [Chiappori et al. \(2019\)](#); [Barseghyan et al. \(2021\)](#). For further examples and a discussion of these ideas, see [Filiz-Ozbay and Masatlioglu \(2023\)](#).

³⁷For a more complex example of the swap-progressive representation of data which do not admit a single-crossing representation, see [Online Appendix D.2](#).

within the support than $\succ_2$. However, as both are also 2-compatible, $d \succeq c$ implies $\succ_2$ must also be ranked above $\succ_1$, an impossibility.

In this example, this observation already suffices to pin down the distribution: the $\succeq$ -swap progressive rationalization is precisely $\bar{\mu} = (0, 0, \frac{5}{8}, \frac{3}{8})$. In particular, $\bar{\mu}$ is supported on $bacd$ and $abdc$. These correspond to the two preferences, among those held in *some* rationalization, which (resp.) agree least and most with $\succeq$. As a consequence, we can regard the support of $\bar{\mu}$ as being ordered by fiscal conservativeness. ■

6 Parametric Restrictions

In practical applications, restrictions on random utility, and more general models of stochastic choice, are often specified parametrically. If $\Theta \subseteq \mathbb{R}^d$ denotes a set of parameters, we refer to any one-to-one and onto map $F : \Theta \rightarrow \mathcal{M} \subseteq \Delta(\mathcal{L})$ as a **parametric restriction**.³⁸ By [Theorem 1](#), a parametric restriction is identifying if and only if:

$$F(\theta') \in F(\theta) + \mathcal{R} \implies \theta = \theta',$$

i.e. if $\mathcal{M} = \text{range}(F)$ intersects each translate of the Ryser subspace at most once. In principle, this is a complete characterization of identification. However, in practice, it is often unclear how to verify this from only knowledge of F . Moreover, many times restrictions are defined by specifying choice probabilities directly in terms of parameters, leaving F only implicit. We now consider a simple, practical test for identification in such cases.

We will consider only *smooth* parametric models. However, as it will essentially be costless to do so, we will allow for arbitrary models of parametric stochastic choice, rather than solely random utility. Recall $\mathcal{P}$ denotes the set of all random choice rules. A **parametric stochastic choice model** is a smooth mapping $\varphi : \Theta \rightarrow \mathcal{P}$.³⁹ In general, it is intractable to directly analyze the full vector of functions φ in any realistically-sized setting.⁴⁰ To determine invertibility in practice, one typically restricts to subsystems with equal numbers of parameters and choice probabilities.

We say $\bar{\varphi} : \Theta \rightarrow \mathbb{R}^d$ defines a **smooth submodel** if it is obtained by selecting any d components of the vector-valued map $G \circ \varphi$, where G is any smooth and smoothly invertible

³⁸This is distinct from the case in which a support constraint is defined by the restriction of some parametric family of *utilities* to X . There, parameters correspond to the preferences of the model, and hence rationalizations to distributions over parameters. Here, we consider the case in which the allowable population distributions *themselves* belong to some parametric family $\mathcal{M} = F(\Theta) \subseteq \Delta(\mathcal{L})$.

³⁹Parametric random utility corresponds to the case where $\varphi = \Phi \circ F$, for some one-to-one $F : \Theta \rightarrow \Delta(\mathcal{L})$.

⁴⁰For example, when $|X| = 6$, a parametric stochastic choice model is a vector of 129 functions.

transformation of choice probabilities. As there are typically many fewer parameters than choice probabilities, this entails a meaningful reduction. While our definition allows for general transformations, in practice, G often corresponds to simply normalizing by certain choice probabilities, see [Example 7](#) and [Example 8](#).

A smooth submodel $\bar{\varphi}$ is **parametrically identified** if it admits a smooth inverse $\bar{\varphi}(\Theta) \rightarrow \Theta$. Parametric identification of a submodel $\bar{\varphi}$ implies the same of the full, unrestricted model φ . While in full generality the converse need not be true, in practice it is typically straightforward to use an unidentified $\bar{\varphi}$ to show the full model φ fails to be parametrically identified.⁴¹

Our next result shows a smooth submodel is parametrically identified precisely when it is so locally, and it satisfies a mild, boundary-preservation condition. This reduces the difficult, global problem of verifying identification of a submodel to a simple analytic test.

Theorem 4. Let $\Theta \subseteq \mathbb{R}^d$ be a contractible, open set of parameters.⁴² A smooth map $\bar{\varphi} : \Theta \rightarrow \mathbb{R}^d$ is parametrically identified if and only if:

- (i) The Jacobian matrix $\mathcal{J}_{\bar{\varphi}}$ is everywhere invertible; and
- (ii) For any sequence $(\theta_n)_{n=1}^{\infty} \subset \Theta$ such that each compact $K \subset \Theta$ contains at most finitely many terms, the same is true of $(\bar{\varphi}(\theta_n))_{n=1}^{\infty}$ for any compact $K' \subset \bar{\varphi}(\Theta)$.

Relative to existing global inverse results, [Theorem 4](#) has several key advantages. First, unlike approaches relying on monotonicity, which typically yield sufficient but not necessary conditions, it provides a complete characterization of parametric identification. For example, when studying the invertibility of a smooth demand system, [Theorem 4](#) remains valid even when goods may be complements, and is applicable in settings when the law of demand may fail.⁴³

Second, unlike many existing results, [Theorem 4](#) allows for both the domain and range of $\bar{\varphi}$ to be distinct, proper subsets of $\mathbb{R}^d$, and requires structural assumptions only on the *domain*, rather than range of $\bar{\varphi}$.⁴⁴ Given that the range of a general vector-valued function can be difficult to determine, this property makes [Theorem 4](#) particularly valuable in practice.

⁴¹See, e.g., [Example 8](#) and [Online Appendix D.1](#).

⁴²A set Θ is contractible if there exist a continuous map $H : \Theta \times [0, 1] \rightarrow \Theta$ such that for each $\theta \in \Theta$, $H(\theta, 0) = \theta$, and $H(\theta, 1) = \theta_0$ for some fixed vector θ_0 . This includes all convex (or star-shaped) sets.

⁴³In particular, [Theorem 4](#) remains applicable even in economic environments not amenable to classic demand inversion results, e.g. [Gale and Nikaido 1965](#); [Berry et al. 2013](#); [Allen 2022](#).

⁴⁴C.f. [Palais 1959](#); [McKenzie 1967](#); [Gordon 1972](#); [Krantz and Parks 2002](#); [Komunjer 2012](#).

Example 7. Let $X = \{x_0, \dots, x_K\}$. A stochastic choice rule belongs to the **Luce model** if there exist weights $w_0 \equiv 1$ and $w_k > 0$ for all $k > 0$, such that, for any $x \in A \subseteq X$:

$$\varphi(w_1, \dots, w_K)_{(x_i, A)} = \rho(x_i, A) = \frac{w_i}{\sum_{j \in A} w_j}.$$

Thus $\Theta = \mathbb{R}_{++}^K$, with each θ corresponding to the vector of normalized weights $(w_1, \dots, w_K)$. Consider the map:

$$\bar{\varphi}(w_1, \dots, w_K) = \begin{bmatrix} \frac{\rho(x_1, X)}{\rho(x_0, X)} \\ \vdots \\ \frac{\rho(x_K, X)}{\rho(x_0, X)} \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_K \end{bmatrix},$$

noting this defines a smooth submodel.⁴⁵ Since (i) the Jacobian of $\bar{\varphi}$ is everywhere the identity, and (ii) some component of $\bar{\varphi}(\theta_n)$ trivially tends to zero or infinity whenever the same is true of θ_n , it follows from [Theorem 4](#) that the submodel $\bar{\varphi}$, and hence the full Luce model, is parametrically identified. ■

Our results can also be used to detect identification failures in more complex settings. To illustrate, we establish a novel result, that the dynamic, logit habit-formation model of [Turansick \(2025\)](#) is not parametrically identified from unconditional choice data.

Example 8. Suppose $X = \{x_0, \dots, x_N\}$, where x_0 denotes an outside option which is always available. A menu is any non-empty subset of X which contains x_0 . A stochastic choice rule represents the unconditional choice probabilities of the habit-formation logit model of [Turansick \(2025\)](#) if it takes the form:

$$\rho(x_i, A) = \frac{v_i \left(\sum_{j \in A \setminus i} v_j + v_i c_i \right)}{\sum_{j \in A} v_j \left(\sum_{k \in A \setminus j} v_k + v_j c_j \right)},$$

for any menu A , where $v_0 = c_0 \equiv 1$, and $v_i, c_i > 1$ for all $i = 1, \dots, N$. Here, the parameters v_i capture the desirability of consuming each alternative, while c_i speaks to how addicting, or habit-forming, consumption of x_i is in a given period.

Define a submodel as follows. For $i = 1, \dots, N$, let:

$$\bar{\varphi}^i(v_1, c_1, \dots, v_N, c_N) = \frac{\rho(x_0, \{x_0, x_i\})}{\rho(x_i, \{x_0, x_i\})} = \frac{1 + v_i}{v_i(1 + v_i c_i)},$$

⁴⁵Here, the map G takes each component of φ and normalizes it by $\varphi_{\rho(x_0, X)}$; $\bar{\varphi}$ is then obtained by selecting the $d = K$ components corresponding to normalized choice probabilities for $x_1, \dots, x_K$ from X .

and for $i = N+1, \dots, 2N$:

$$\bar{\varphi}^{N+i}(v_1, c_1, \dots, v_N, c_N) = \frac{\rho(x_0, \{x_0, x_i, x_{i+1}\})}{\rho(x_i, \{x_0, x_i, x_{i+1}\})} = \frac{1 + v_i + v_{i+1}}{v_i + v_i v_{i+1} + c_i v_i^2},$$

where $v_{N+1} \equiv v_1$, and $c_{N+1} \equiv c_1$.

We now show this submodel fails to satisfy condition (ii) of [Theorem 4](#), and hence fails to be parametrically identified. To do so, we first define the function:

$$c(v_i) = \frac{1 + v_i}{10v_i^2} - \frac{1}{v_i}.$$

Note that, for any $i = 1, \dots, N$, we have $\bar{\varphi}^i(\dots, v_i, c(v_i), \dots) = \frac{1}{10}$. Consider now the restriction of $\bar{\varphi}$ to the curve $(v, c(v), \dots, v, c(v))$. The last N components of $\bar{\varphi}$ are identical, and equal to:

$$\bar{\varphi}^{N+i}(v, c(v), \dots, v, c(v)) = \frac{1 + 2v}{v^2 + 10(1 + v)}.$$

As $v \rightarrow 1$ along this curve, $\bar{\varphi}^{N+i} \rightarrow \frac{1}{7}$. But this is precisely the value attained by $\bar{\varphi}^{N+i}$ at $v = 3$. Since the first N components of $\bar{\varphi}$ are constant in v , we have shown that for any sequence of parameters of the form $(v_n, c(v_n), \dots, v_n, c(v_n))$ with $v_n \rightarrow 1$, we obtain a violation of condition (ii) of [Theorem 4](#), and hence $\bar{\varphi}$ is not parametrically identified. It is then straightforward to use the insight from this submodel to construct a distinct pair of parameter vectors that yield identical choice probabilities in the full model.⁴⁶ ■

7 Beyond Random Utility

Thus far, we have considered primarily the random utility model. However, our core notion of Ryser swap characterizes the identified sets of many other stochastic choice models. In each of these additional settings, exact analogs of our main results obtain. We illustrate several such extensions below.

Random Choice

The simplest extension is to the unrestricted class of random choice models, possibly subject to limited observability. Here, X is a finite set of alternatives, and $\Sigma \subseteq 2^X \setminus \{\emptyset\}$ an arbitrary collection of menus on which choice frequencies are observed. Let $\mathcal{C}$ denote the collection of all choice functions on Σ .⁴⁷ A random choice model is a probability measure $\mu \in \Delta(\mathcal{C}_\Sigma)$

⁴⁶See [Online Appendix D.1](#) for an explicit example.

⁴⁷That is, $\mathcal{C}_\Sigma = \prod_{A \in \Sigma} A$.

such that, for all $a \in A \in \Sigma$:

$$\rho(a, A) = \sum_{c \in \mathcal{C}} \mu(c) \mathbb{1}_{\{a=c(A)\}}.$$

Here, a Ryser swap is a signed measure $R \in \mathbb{R}^{\mathcal{C}_\Sigma}$ of the form:

$$R = \mathbb{1}_{\{c'_1, c'_2\}} - \mathbb{1}_{\{c_1, c_2\}},$$

where $c'_1, c'_2 \in \mathcal{C}_\Sigma$ are the choice functions obtained by swapping the choices made by c_1 and c_2 on a single menu $A \in \Sigma$. Similarly, we define the Ryser subspace for the random choice model $\mathcal{R}^{rc} \subset \mathbb{R}^{\mathcal{C}_\Sigma}$ as the linear span of all such vectors. Once again, the subspace $\mathcal{R}^{rc}$ characterizes identifying restrictions.

Theorem 5. Fix $\Sigma \subseteq 2^X \setminus \{\emptyset\}$. Two distributions $\mu, \mu' \in \Delta(\mathcal{C}_\Sigma)$ generate identical choice probabilities on every menu in Σ if and only if $\mu - \mu' \in \mathcal{R}^{rc}$. In particular, for any $\mathcal{M} \subseteq \Delta(\mathcal{C}_\Sigma)$ the set of distributions generating identical choice probabilities to μ on every menu in Σ is precisely $(\mu + \mathcal{R}^{rc}) \cap \mathcal{M}$.

Exact analogues of [Theorem 2](#) and [Theorem 3](#) also obtain in this setting.⁴⁸ Notably, since $\mathcal{C}_\Sigma$ contains the rational choice functions on Σ , the random utility model can be viewed as a support restriction within this more general framework. Thus [Theorem 5](#), and the random choice analogue of [Theorem 2](#), allow our results to extend to the case of incompletely observed data, i.e. when $\Sigma \subsetneq 2^X \setminus \{\emptyset\}$.

Frame Dependent Choice

Our results also extend naturally to models of boundedly rational behavior. We illustrate this on a simple version of the frame-dependent random utility model of [Cheung and Masatlioglu \(2024\)](#).⁴⁹ Here, individuals face a fixed, finite choice set $X = \{x_1, \dots, x_K\}$, but receive varying suggestions from a black-box recommendation system that highlights some subset $A \subseteq X$, rendering them more desirable.

Here, the empiricist observes *frame-dependent* stochastic choice data $\rho : X \times 2^X \rightarrow [0, 1]$, where:

$$\sum_{x \in X} \rho(x, A) = 1,$$

⁴⁸In particular, the analogue of our notion of swap-progressivity coincides precisely with the progressivity condition of [Filiz-Ozbay and Masatlioglu \(2023\)](#); the analogue of our [Theorem 3](#) in this setting corresponds their Theorem 1.

⁴⁹We extend this analysis to the full model of [Cheung and Masatlioglu \(2024\)](#) in [Online Appendix B.2](#).

for each $A \subseteq X$, interpreted as the observed probability of the the agent choosing x from X when faced with the recommendation A .

Define $X^* = \{x_1^N, \dots, x_K^N, x_1^F, \dots, x_K^F\}$, where we regard x_i when framed (x_i^F) as distinct from x_i when it is not (x_i^N). Given a recommendation set $A \subseteq X$, the subject may be regarded as choosing from the virtual menu:

$$A^* = \{x^N : x \in X \setminus A\} \cup \{x^F : x \in A\} \subset X^*.$$

Since framing makes any alternative strictly more preferred, no alternative below the most desirable x^N can ever affect observed choice. Thus, let $\mathcal{L}^*$ denote the set of *truncated* strict preferences on X^* , consisting only of the ranking up to, and including, the best, non-framed alternative, and which satisfy $x^F \succ^* x^N$ for all $x \in X$.⁵⁰

The data are compatible with a frame-dependent random utility representation $\mu \in \Delta(\mathcal{L}^*)$ if:

$$\rho(x, A) = \sum_{\succ^* \in \mathcal{L}^*} \mu(\succ^*) \mathbb{1}_{\{\hat{x} \text{ is } \succ^*\text{-maximal in } A^*\}},$$

where $\hat{x} = x^N$ if $x \notin A$, and x^F if $x \in A$. Given two truncated preferences in $\succ_1^*, \succ_2^* \in \mathcal{L}^*$ we once again say they are compatible if they agree upon the set of k -best alternatives, though not necessarily their ranking. A Ryser swap for the frame-dependent model is simply a signed measure in $\mathbb{R}^{\mathcal{L}^*}$ which places unit mass on a k -compatible pair $\succ_1^*$ and $\succ_2^*$ and mass negative one on the pair of truncated preferences obtained by swapping their k -initial segments.⁵¹ Let $\mathcal{R}^{fd} \subset \mathbb{R}^{\mathcal{L}^*}$ denote the linear span of these vectors.

Theorem 6. Two distributions $\mu, \mu' \in \Delta(\mathcal{L}^*)$ are observationally equivalent if and only if $\mu - \mu' \in \mathcal{R}^{fd}$. In particular, for any $\mathcal{M} \subseteq \Delta(\mathcal{L}^*)$, its identified set is precisely $(\mu + \mathcal{R}^{fd}) \cap \mathcal{M}$.

As before, analogues of [Theorem 2](#) and [Theorem 3](#) also obtain in this setting. In particular, as was true for the random utility model, a swap-progressive *frame-dependent* representation exists whenever the data are rationalizable, and will itself always belong to $\Delta(\mathcal{L}^*)$.

Dynamic Discrete Choice

Our methodology is also applicable to the problem of identifying distributions over dynamic choice functions from Markovian conditional choice probabilities. Let X denote a finite set of alternatives, and T a finite time horizon. In period $t = 1$, the analyst observes

⁵⁰For example, an (unrestricted) preference $x_1^F \succ x_2^F \succ x_1^N \succ x_2^N$ would correspond to the truncated preference $x_1^F \succ^* x_2^F \succ^* x_1^N$.

⁵¹Note this is well-defined, as these swaps necessarily belong to $\mathcal{L}^*$.

unconditional choice probabilities $\rho_1(x)$. For each period $t = 2, \dots, T$, the analyst observes probabilities of time t choice, conditional the prior period's choice, $\rho_t(y | x)$. Together, we refer to $\{\rho_1, \dots, \rho_T\}$ as a system of conditional choice probabilities.

An analyst seeks to decompose a system of conditional choice probabilities into a measure over dynamic choice functions in $\mathcal{C}^{dc} = X^T$. A system of conditional choice probabilities is compatible with a dynamic discrete choice representation $\mu \in \Delta(\mathcal{C}^{dc})$ if:

$$\rho_1(x) = \sum_{c \in \mathcal{C}^{dc}} \mu(c) \mathbb{1}_{\{x=c(1)\}}$$

and

$$\rho_t(y | x) = \frac{\sum_{c \in \mathcal{C}^{dc}} \mu(c) \mathbb{1}_{\{x=c(t-1), y=c(t)\}}}{\sum_{c \in \mathcal{C}^{dc}} \mu(c) \mathbb{1}_{\{x=c(t-1)\}}}$$

for all $t = 2, \dots, T$ and $x, y \in X$.

We say two choice functions $c_1, c_2 \in \mathcal{C}^{dc}$ are t -compatible if c_1 and c_2 make the same choice at time $t-1$. A Ryser swap is a signed measure $R \in \mathbb{R}^{\mathcal{C}^{dc}}$ of the form $\mathbb{1}_{\{c'_1, c'_2\}} - \mathbb{1}_{\{c_1, c_2\}}$, where c'_1 and c'_2 are the choice functions obtained by swapping the first $t-1$ choices of some t -compatible pair of choice functions $c_1, c_2 \in \mathcal{C}^{dc}$. Let $\mathcal{R}^{dc} \subset \mathbb{R}^{\mathcal{C}^{dc}}$ denote the linear span of these measures.

Theorem 7. Two distributions $\mu, \mu' \in \Delta(\mathcal{C}^{dc})$ are observationally equivalent if and only if $\mu - \mu' \in \mathcal{R}^{dc}$. In particular, for any $\mathcal{M} \subseteq \Delta(\mathcal{C}^{dc})$, its identified set is precisely $(\mu + \mathcal{R}^{dc}) \cap \mathcal{M}$.

Finally, as in each previous setting, analogues of [Theorem 2](#) and [Theorem 3](#) also obtain. In particular, a swap-progressive rationalization exists and is unique for any rationalizable data, and will always be supported on $\mathcal{C}^{dc}$.

While we consider here dynamic discrete choice, our results also extend to more complex, finite dynamic programming models, including those involving state-dependent utility and menu variation. This allows for further application to models such as [Rust \(1987\)](#).

8 Related Literature

The first explicit example of identification failure in the random utility model is due to [Fishburn \(1998\)](#). His counterexample forms the basis of our notion of a (non-trivial) compatible pair. [Turansick \(2022\)](#) characterizes data sets which possess a unique random utility rationalization. We obtain this characterization as a straightforward consequence of the proof of our [Theorem 1](#).⁵²

⁵²See also [Doignon and Saito \(2023\)](#).

A number of papers consider restrictions on the random utility model for purposes of obtaining identification. [Manzini and Mariotti \(2018\)](#) show that in a two-state model, distributions over preferences can generically be recovered. [Suleymanov \(2024\)](#) recovers identification by requiring certain independence properties of the rationalization; [Honda \(2021\)](#) considers a random cravings model where identification obtains under monotonicity restrictions. By directly characterizing the identified sets of the general random utility model, we instead provide necessary and sufficient conditions for any restriction to be identifying.

When alternatives are ordered, a well-known class of restrictions rely on various single-crossing type conditions on the support of rationalizations to achieve uniqueness. This idea is originally due to [Apesteguia et al. \(2017\)](#); [Filiz-Ozbay and Masatlioglu \(2023\)](#) considered an extension to general stochastic choice models that guarantees existence while preserving uniqueness, at the cost of the unique rationalization possibly failing to belong to the random utility model, even when the data is rationalizable by such a distribution.⁵³

Our notion of swap-progressivity builds on these contributions. Applied to the random utility model, it generalizes single-crossing and provides unique predictions, even for data which fail the centrality axiom of [Apesteguia et al. \(2017\)](#).⁵⁴ In the context of general random choice, the analogue of our [Theorem 3](#) is precisely Theorem 1 of [Filiz-Ozbay and Masatlioglu \(2023\)](#). Moreover, swap-progressive representations exist, are unique, and belong to the models of interest for each of the environments considered in [Section 7](#).

In practice, it is common to specify random utility model parametrically, e.g. [Luce \(1959\)](#).⁵⁵ More recently, [Chambers et al. \(2025\)](#) considers an extension of the Luce model including both salience and utility in its parameterization which preserves identification. In general, Luce type structure allows for simple identification arguments. However, in more complex environments, [Theorem 4](#) provides a practical tool for establishing similar results.

While our paper focuses on the case of finite environments, there are numerous positive identification results when domains are infinite. [Gul and Pesendorfer \(2006\)](#) shows that the random expected utility model is identified, as is the random quasilinear utility model ([Williams, 1977](#); [Daly and Zachary, 1979](#); [Yang and Kopylov, 2023](#)).

Methodologically, our [Theorem 1](#) builds on work in the mathematical discipline of to-

⁵³[Yildiz \(2023\)](#) studies the structure of such ordered models and shows that a unique ordered representations is possible only when the types in the model form a lattice. On the other hand, [Petri \(2022\)](#) considers an extension of these ordered models to allow for multi-valued choice.

⁵⁴In particular, in this setting, swap progressivity has the same explanatory power as the baseline random utility model, while also being identified.

⁵⁵This model is observationally equivalent to the logit model of [McFadden \(1974\)](#). See also [Sandomirskiy and Tamuz \(2023\)](#) for a recent characterization.

mography (Ryser 1957; Fishburn et al. 1991; Kong and Herman 1999).⁵⁶ We show that both classic tomographic applications, as well as related questions about certain, restricted sets of measures with prescribed marginals, can be fruitfully encoded as flow problems on directed graphs. In this general setting, we provide a unified notion of Ryser swap, which specializes to the forms seen in Section 3 and Section 7, and show it characterizes the abstract analogue of our identified sets.

Similarly, to the extent of our knowledge, Theorem 4 does not appear in the economic or mathematical literature. Relative to classical results of Hadamard (Gordon 1972; Krantz and Parks 2002), we show that when the domain of a non-linear system is convex (or, more generally, contractible) and the system contains an equal number of equations and unknowns, the classical, difficult-to-verify conditions on the system’s range which characterize invertibility are always satisfied. Our proof relies on a fixed point theorem due to Smith (1934) which we believe has not found prior economic application.

9 Conclusion

When analysts observe only aggregate choice frequency data, many work-horse models of stochastic choice are unidentified. In such instances, explicit descriptions of the identified sets are of direct practical importance, as they allow analysts not only to understand the range of possible aggregate effects of a policy intervention, but also its distributional implications. Despite this, little has been known about the structure of these sets, even for many widely used models.

We characterize the identified sets of a wide range of stochastic choice models. Our key insight is that any transformation of a distribution over primitives that preserves all choice probabilities can be represented as a finite sequence of simple, mass-swapping operations that we term *Ryser swaps*. Using this, we give explicit characterizations of both the defining inequalities and extreme points of these models’ identified sets. In each of these settings, our results provide tight bounds on the population cross-sections compatible with observed choice frequencies. This yields valuable insights, e.g., for assessing the effects of policies which may differentially affect various subsets of a population.

We also provide specialized characterizations of identification for a number of common, structured classes of restriction, such as constraints on the support of the rationalizing distri-

⁵⁶Heckman and Pinto (2018) use Ryser’s theorem to show identification in a different context. A different strand of literature uses tensor-decomposition techniques to study identification; see Dardanoni et al. (2023) and Kops et al. (2026).

bution, or when choice frequencies are assumed to vary smoothly in some vector of parameters. These provide novel tools for practitioners to employ when evaluating the identification properties of various restrictions going forward.

Our results suggest several fruitful directions for further inquiry. One would be to understand how our methodological techniques generalize to the case of continuous consumption spaces. While we have emphasized that our results apply broadly across various economic domains of interest, another interesting question would be to determine precisely which models admit a notion of Ryser swap which characterizes their identified sets. We leave these questions for future research.

References

- AGARWAL, N. AND P. SOMAINI (2018): “Demand analysis using strategic reports: An application to a school choice mechanism,” *Econometrica*, 86, 391–444.
- AHUJA, R. K., T. L. MAGNANTI, J. B. ORLIN, ET AL. (1993): *Network flows: theory, algorithms, and applications*, vol. 1, Prentice hall Englewood Cliffs, NJ.
- ALLEN, R. (2022): “Injectivity and the law of demand,” *Economics Letters*, 215, 110496.
- ALLEN, R. AND J. REHBECK (2019): “Identification with additively separable heterogeneity,” *Econometrica*, 87, 1021–1054.
- APESTEGUIA, J. AND M. A. BALLESTER (2018): “Monotone stochastic choice models: The case of risk and time preferences,” *Journal of Political Economy*, 126, 74–106.
- (2023): “Random utility models with ordered types and domains,” *Journal of Economic Theory*, 211, 105674.
- APESTEGUIA, J., M. A. BALLESTER, AND J. LU (2017): “Single-Crossing Random Utility Models,” *Econometrica*, 85, 661–674.
- BARSEGHYAN, L., F. MOLINARI, AND M. THIRKETTLE (2021): “Discrete choice under risk with limited consideration,” *American Economic Review*, 111, 1972–2006.
- BECKERT, W. AND R. BLUNDELL (2004): “Invertibility of Nonparametric Stochastic Demand Functions,” Tech. rep., Birkbeck, Department of Economics, Mathematics & Statistics.

- BERRY, S., A. GANDHI, AND P. HAILE (2013): “Connected substitutes and invertibility of demand,” *Econometrica*, 81, 2087–2111.
- BERRY, S. T. AND P. A. HAILE (2024): “Nonparametric identification of differentiated products demand using micro data,” *Econometrica*, 92, 1135–1162.
- BLOCK, H. D. AND J. MARSCHAK (1959): “Random Orderings and Stochastic Theories of Response,” Tech. rep., Cowles Foundation for Research in Economics, Yale University.
- BREDON, G. E. (2013): *Topology and geometry*, vol. 139, Springer Science & Business Media.
- CHAMBERS, C. P., Y. MASATLIOGLU, P. NATENZON, AND C. RAYMOND (2025): “Weighted Linear Discrete Choice,” *American Economic Review*, 115, 1226–57.
- CHAMBERS, C. P., Y. MASATLIOGLU, AND C. TURANSICK (2024): “Correlated choice,” *Theoretical Economics*.
- CHAMBERS, C. P. AND C. TURANSICK (2025): “The Limits of Identification in Discrete Choice,” *Games and Economic Behavior*.
- CHEUNG, P. H. AND Y. MASATLIOGLU (2024): “Frame-Dependent Random Utility,” *Working Paper*.
- CHIAPPORI, P.-A., B. SALANIÉ, F. SALANIÉ, AND A. GANDHI (2019): “From aggregate betting data to individual risk preferences,” *Econometrica*, 87, 1–36.
- DALY, A. AND S. ZACHARY (1979): “Identifying and Measuring the Determinants of Mode Choice,” *Improved Multiple Choice Models. London: Teakfield*.
- DARDANONI, V., P. MANZINI, M. MARIOTTI, H. PETRI, AND C. J. TYSON (2023): “Mixture choice data: revealing preferences and cognition,” *Journal of Political Economy*, 131, 687–715.
- DAVIS-STOBER, C. P., J.-P. DOIGNON, S. FIORINI, F. GLINEUR, AND M. REGENWETTER (2018): “Extended formulations for order polytopes through network flows,” *Journal of mathematical psychology*, 87, 1–10.
- DEATON, A. (1980): “The Measurement of Welfare: Theory and Practical Guidelines,” LSMS Working Paper 7, World Bank, Washington, DC.
- DEB, R., Y. KITAMURA, J. K. QUAH, AND J. STOYE (2023): “Revealed price preference: theory and empirical analysis,” *The Review of Economic Studies*, 90, 707–743.

- DOIGNON, J.-P. AND K. SAITO (2023): “Adjacencies on random ordering polytopes and flow polytopes,” *Journal of Mathematical Psychology*, 114, 102768.
- DURAJ, J. (2018): “Dynamic Random Subjective Expected Utility,” Unpublished.
- EILENBERG, S. (1940): “On a theorem of PA Smith concerning fixed points for periodic transformations,” *Duke Math. J.*, 6, 428–437.
- FALMAGNE, J.-C. (1978): “A representation theorem for finite random scale systems,” *Journal of Mathematical Psychology*, 18, 52–72.
- FILIZ-OZBAY, E. AND Y. MASATLIOGLU (2023): “Progressive random choice,” *Journal of Political Economy*, 131, 716–750.
- FIORINI, S. (2004): “A Short Proof of a Theorem of Falmagne,” *Journal of Mathematical Psychology*, 48, 80–82.
- FISHBURN, P. C. (1998): “Stochastic Utility,” in *Handbook of Utility Theory*, ed. by S. Barbera, P. Hammond, and C. Seidl, Kluwer Dordrecht, 273–318.
- FISHBURN, P. C., J. C. LAGARIAS, J. A. REEDS, AND L. A. SHEPP (1991): “Sets uniquely determined by projections on axes II Discrete case,” *Discrete Mathematics*, 91, 149–159.
- FRICK, M., R. IJIMA, AND T. STRZALECKI (2019): “Dynamic Random Utility,” *Econometrica*, 87, 1941–2002.
- GALE, D. AND H. NIKAIDO (1965): “The Jacobian matrix and global univalence of mappings,” *Mathematische Annalen*, 159, 81–93.
- GORDON, W. B. (1972): “On the diffeomorphisms of Euclidean space,” *The American Mathematical Monthly*, 79, 755–759.
- GUL, F. AND W. PESENDORFER (2006): “Random expected utility,” *Econometrica*, 74, 121–146.
- HE, K., F. SANDOMIRSKIY, AND O. TAMUZ (Forthcoming): “Private private information,” *Journal of Political Economy*.
- HECKMAN, J. J. AND R. PINTO (2018): “Unordered monotonicity,” *Econometrica*, 86, 1–35.
- HONDA, E. (2021): “A Model of Random Cravings,” *Working Paper*.

- KALOUPTSIDI, M., Y. KITAMURA, L. LIMA, AND E. SOUZA-RODRIGUES (2021): “Counterfactual analysis for structural dynamic discrete choice models,” *NBER Working Paper Series*.
- KAMAT, V. AND S. NORRIS (2025): “Estimating Welfare Effects in a Non-Parametric Choice Model: The Case of School Vouchers,” *Review of Economic Studies*, rda064.
- KASHAEV, N., M. PLÁVALA, AND V. H. AGUIAR (2024): “Entangled vs. Separable Choice,” *arXiv preprint arXiv:2403.09045*.
- KITAMURA, Y. AND J. STOYE (2018): “Nonparametric Analysis of Random Utility Models,” *Econometrica*, 86, 1883–1909.
- KOMUNJER, I. (2012): “Global identification in nonlinear models with moment restrictions,” *Econometric Theory*, 28, 719–729.
- KONG, T. Y. AND G. T. HERMAN (1999): *Tomographic Equivalence and Switching Operations*, Boston, MA: Birkhäuser Boston, 59–84.
- KOPS, C., P. MANZINI, M. MARIOTTI, AND I. PASICHNICHENKO (2026): “Identifying Behavioral Types,” *mimeo*.
- KRANTZ, S. G. AND H. R. PARKS (2002): *The implicit function theorem: history, theory, and applications*, Springer Science & Business Media.
- LEE, J. M. (2012): *Introduction to smooth manifolds*, Springer.
- (2013): *Introduction to Smooth Manifolds*, vol. 218 of *Graduate Texts in Mathematics*, Springer New York, 3rd ed.
- LINDENLAUB, I. (2017): “Sorting multidimensional types: Theory and application,” *The Review of Economic Studies*, 84, 718–789.
- LU, J. (2016): “Random choice and private information,” *Econometrica*, 84, 1983–2027.
- LU, J., Y. LUO, K. SAITO, AND Y. XIN (2024): “Did Harold Zuercher Have Time-Separable Preferences?” *arXiv preprint arXiv:2406.07809*.
- LU, J. AND K. SAITO (2018): “Random Intertemporal Choice,” *Journal of Economic Theory*, 177, 780–815.
- LUCE, R. D. (1959): *Individual Choice Behavior*, John Wiley.

- MANSKI, C. F. (2007): “Partial identification of counterfactual choice probabilities,” *International Economic Review*, 48, 1393–1410.
- MANZINI, P. AND M. MARIOTTI (2018): “Dual random utility maximisation,” *Journal of Economic Theory*, 177, 162–182.
- MARSCHAK, J. (1953): “Economic Measurements for Policy and Prediction,” in *Studies in Econometric Method*, ed. by W. C. Hood and T. C. Koopmans, New York: John Wiley & Sons, Cowles Commission Monograph 14, 1–26, chapter 1.
- MCCLELLON, M. (2015): “Unique random utility representations,” *Unpublished manuscript, Harvard University OpenScholar*.
- McFADDEN, D. (1972): “Conditional logit analysis of qualitative choice behavior,” .
- (1974): “Conditional Logit Analysis of Qualitative Choice Behavior,” *Frontiers in Econometrics*.
- McFADDEN, D. L. (2005): “Revealed Stochastic Preference: A Synthesis,” *Economic Theory*, 26, 245–264.
- McKENZIE, G. W. (1983): *Measuring economic welfare: new methods*, Cambridge University Press.
- McKENZIE, L. W. (1960): “Matrices with Dominant Diagonals and Economic Theory,” in *Equilibrium, Trade, and Growth: Selected Papers of Lionel W. McKenzie*, ed. by T. Mitra and K. Nishimura, Oxford: Oxford University Press, 201–216.
- (1967): “Theorem and counter-example,” *International Economic Review*, 8, 279–285.
- PALAIS, R. S. (1959): “Natural operations on differential forms,” *Transactions of the American Mathematical Society*, 92, 125–141.
- PETRI, H. (2022): “Random (sequential) multivalued choice,” *Working Paper*.
- ROSEN, J. B. (1965): “Existence and uniqueness of equilibrium points for concave n-person games,” *Econometrica: Journal of the Econometric Society*, 520–534.
- RUST, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 57, 999–1033.

- RYSER, H. J. (1957): “Combinatorial properties of matrices of zeros and ones,” *Canadian Journal of Mathematics*, 9, 371–377.
- SANDOMIRSKIY, F. AND O. TAMUZ (2023): “Decomposable Stochastic Choice,” *arXiv preprint arXiv:2312.04827*.
- SIMSEK, A., A. OZDAGLAR, AND D. ACEMOGLU (2005): “On the uniqueness of solutions for nonlinear and mixed complementarity problems,” *Massachusetts Inst. Technol., [Online]*. Available: <https://economics.mit.edu/files/204>.
- SMITH, P. A. (1934): “A theorem on fixed points for periodic transformations,” *Annals of Mathematics*, 35, 572–578.
- STRZALECKI, T. (2025): “Stochastic choice theory,” *Cambridge Books*.
- SULEYMANOV, E. (2024): “Branching-independent random utility model,” *Journal of Economic Theory*, 105880.
- TEBALDI, P., A. TORGOVITSKY, AND H. YANG (2023): “Nonparametric estimates of demand in the california health insurance exchange,” *Econometrica*, 91, 107–146.
- TURANSICK, C. (2022): “Identification in the random utility model,” *Journal of Economic Theory*, 203, 105489.
- (2024): “Consumption dependent random utility,” *arXiv preprint arXiv:2412.05344*.
- (2025): “Intertemporal Aggregation of Choice Data with Consumption Dependent Preferences,” *arXiv preprint arXiv:2302.05806*.
- WANG, A. (2021): “A BLP Demand Model of Product-Level Market Shares with Complementarity,” Tech. rep., University of Warwick, Department of Economics.
- WILLIAMS, H. C. (1977): “On the formation of travel demand models and economic evaluation measures of user benefit,” *Environment and planning A*, 9, 285–344.
- WINKLER, G. (1988): “Extreme points of moment sets,” *Mathematics of Operations Research*, 13, 581–587.
- YANG, E. AND I. KOPYLOV (2023): “Random quasi-linear utility,” *Journal of Economic Theory*, 209, 105650.
- YILDIZ, K. (2023): “Foundations of self-progressive choice models,” *arXiv preprint arXiv:2212.13449*.

A Graphical Results

A.1 Definitions & Preliminaries

A **graph** is a pair $(\mathcal{N}, \mathcal{E})$ where $\mathcal{N}$ is a finite set of **nodes**, and $\mathcal{E}$ a finite collection of ordered pairs in $\mathcal{N} \times \mathcal{N}$, where we explicitly allow for repetition.⁵⁷ We interpret an ordered pair $(x, y) \in \mathcal{E}$ as directed from x (the ‘tail’) to y (the ‘head’). We will sometimes denote the ordered pair $(x, y) \in \mathcal{E}$ by $x \rightarrow y$.

For a given node $n \in \mathcal{N}$, we write $\mathcal{E}_n^\downarrow$ (resp. $\mathcal{E}_n^\uparrow$) for the sub-collection of all edges in $\mathcal{E}$ whose head (resp. tail) is n . A graph is said to be **acyclic** if there is no subset of edges with the property that any node appearing in some edge in the set does so precisely once as the head of some edge, and once as a tail. A **directed acyclic graph** (DAG) is a tuple $(\mathcal{N}, \mathcal{E}, s, t)$, where $(\mathcal{N}, \mathcal{E})$ is an acyclic graph, and s and t denote specialized nodes in $\mathcal{N}$, respectively the **source** and **sink**, such that $\mathcal{E}_s^\downarrow = \mathcal{E}_t^\uparrow = \emptyset$. We will assume that s and t are the unique nodes with these properties.

Flows & Path Decompositions

Given a directed acyclic graph, a **flow** is a function $f : \mathcal{E} \rightarrow \mathbb{R}_+$ such that (i) at every non-source/sink node n , total in-flow equals total out-flow, and (ii) the total out-flow from the source is normalized to unity:

$$(i) \sum_{e \in \mathcal{E}_n^\downarrow} f(e) = \sum_{e \in \mathcal{E}_n^\uparrow} f(e) \quad \text{and} \quad (ii) \sum_{e \in \mathcal{E}_s^\downarrow} f(e) = 1. \quad (4)$$

A function $f : \mathcal{E} \rightarrow \mathbb{R}_+$ is a **quasi-flow** if it satisfies only condition (i). A **path** P is a sequence of edges such that (i) the source s and sink t each appear in exactly one edge, and (ii) every other $n \in \mathcal{N}$ that appears in any edge appears exactly twice, once as a head and once as a tail. A path P is said to pass through a node n if n appears in some edge in P , and for any node $n \in \mathcal{N}$ appearing in P , we let $P_n^\downarrow$ (resp. $P_n^\uparrow$) denote the unique edge in P whose head (resp. tail) is n .⁵⁸ Let $\mathcal{P}$ denote the collection of all paths.⁵⁹ We say a non-negative

⁵⁷That is, $\mathcal{E}$ is a multi-set, hence our definition of graph explicitly allows for *multigraphs*.

⁵⁸Further, we use $s_k^\uparrow(P)$ (resp. $s_k^\downarrow(P)$) to denote the first k (resp. last $|P| - k$) edges in a path P . We will also sometimes use $s_n^\uparrow$ and $s_n^\downarrow$ to denote the initial/terminal segments of a path up to/starting from some $n \in \mathcal{N}$.

⁵⁹In particular, throughout the appendix we will use $\mathcal{P}$ to denote the path set of a graph, rather than the set of random choice rules as used in the body.

measure $\pi \in \mathbb{R}_+^{\mathcal{P}}$ defines a **path decomposition** of a quasi-flow f if, for all $e \in \mathcal{E}$:

$$f(e) = \sum_{\{P \in \mathcal{P} : e \in P\}} \pi(P).$$

A path decomposition belongs to $\Delta(\mathcal{P})$ if and only if it decomposes a flow. A map $f : \mathcal{E} \rightarrow \mathbb{R}_+$ admits at least one path decomposition if and only if it is a quasi-flow (e.g. [Ahuja et al. 1993](#)).

Compatibility & Ryser Swaps

Two paths P_1 and P_2 are **n -compatible** if they pass through some common node $n \in \mathcal{N}$. A pair of paths P'_1, P'_2 are **n -conjugate** to an n -compatible pair P_1, P_2 if they are the paths obtained by following P_1 until n , then P_2 thereafter, and vice-versa.⁶⁰ Two path decompositions $\pi, \pi' \in \Delta(\mathcal{P})$ differ by a **weighted Ryser swap** if:

$$\pi - \pi' = \varepsilon (\mathbb{1}_{P_1} + \mathbb{1}_{P_2} - \mathbb{1}_{P'_1} - \mathbb{1}_{P'_2}), \quad (5)$$

where P'_1 and P'_2 are conjugate to some compatible pair P_1, P_2 .⁶¹ Note that if π' and π differ by a weighted Ryser swap, they are path decompositions of the same quasi-flow. We define the **Ryser subspace** $\mathcal{R} \subseteq \mathbb{R}^{\mathcal{P}}$ as the span of all weighted Ryser swaps with R denoting a typical Ryser swap.

Rearrangements

Let $P_1, \dots, P_K$ be a sequence of paths, explicitly allowing repetition. For any $n \in \mathcal{N}$, let $\Pi_n \subseteq \{1, \dots, K\}$ denote the subset of indices corresponding to paths P_i passing through n . A **rearrangement** of $P_1, \dots, P_K$ is a choice of permutation $\sigma_n : \Pi_n \rightarrow \Pi_n$ for each $n \in \mathcal{N}$. Any rearrangement of $P_1, \dots, P_K$ recursively defines a new sequence of paths $P'_1, \dots, P'_K$ via:

$$P'_k : \quad n_0 \rightarrow \dots \rightarrow n_M$$

where $n_0 \equiv s$, $n_M \equiv t$, and:

$$n_{i+1} = \text{head} [P_{(\sigma_{n_i} \circ \dots \circ \sigma_{n_0})^{-1}(k)}]_{\downarrow}^{n_i},$$

for all $0 \leq i \leq M-1$. In other words, P'_k is formed by initially following $P_{\sigma_{n_0}^{-1}(k)}$ to its first non-source node, n_1 , then following the outgoing edge from n_1 belonging to $P_{(\sigma_{n_1} \circ \sigma_{n_0})^{-1}(k)}$

⁶⁰That is, P'_1 agrees with P_1 up to n , then agrees with P_2 thereafter, and vice-versa.

⁶¹Formally, we refer to any vector in $\mathbb{R}^{\mathcal{P}}$ of the form taken by the right-hand side of (5) to be a weighted Ryser swap.

to n_2 , and so forth. Two sequences are said to be **rearrangement equivalent** if they can be obtained from one another by a rearrangement.

The Random Utility Graph

Our main technical contribution is to characterize the sets of path decompositions of arbitrary quasi-flows, for general directed acyclic graphs. Each of our economic results then obtains simply by applying these abstract theorems to the particular family of graphs encoding a specific stochastic choice theory. As our exposition has focused on the random utility model, we introduce its graphical representation, due to [Fiorini \(2004\)](#), here. In [Online Appendix B](#), we collect similar constructions for the various other stochastic choice theories mentioned in the text.

The random utility graph is defined by $\mathcal{N} = 2^X$ and $\mathcal{E} = \{(A, B) \mid B = A \setminus \{a\}, a \in A\}$, where $s = X$ and $t = \emptyset$. Any rationalizable (i.e. consistent with the random utility model) random choice rule ρ induces a unique flow on this graph as follows:

$$f(A \rightarrow A \setminus \{a\}) = \sum_{A \subseteq B} (-1)^{|B \setminus A|} \rho(a, B). \quad (6)$$

The quantities $f(A \rightarrow A \setminus \{a\})$ are often referred to as *Block-Marschak polynomials*.⁶²

[Fiorini \(2004\)](#) shows this f indeed defines a flow, i.e. satisfies both conditions of (4) and, conversely, that every flow on this graph may be regarded as arising via (6) from some unique, rationalizable random choice rule. Moreover, there is a bijective correspondence between the path set $\mathcal{P}$ and strict preferences on X , by regarding a path:

$$X \rightarrow X \setminus \{x_1\} \rightarrow \cdots \rightarrow X \setminus \{x_1, \dots, x_{N-1}\} \rightarrow \emptyset$$

as a sequence of nested lower contour sets, which uniquely defines the preference $x_1 \succ \cdots \succ x_N$. As a consequence, we may identify $\Delta(\mathcal{P})$ with $\Delta(\mathcal{L})$, and, under this identification, [Fiorini \(2004\)](#) shows a distribution in $\Delta(\mathcal{P})$ rationalizes the data ρ if and only if it is a path decomposition of the flow (6). Thus the partial identification properties of the random utility model are precisely equivalent to the path decompositions of flows on $(\mathcal{N}, \mathcal{E}, s, t)$.

A.2 Graphical Results and Proofs

Lemma 1 (Zipper Lemma 1). Consider two path decompositions $\pi, \pi' \in \Delta(\mathcal{P})$ of a quasi-flow f . Let $P \in \text{supp}(\pi)$, $\{P_u\}_{u=1}^U \subseteq \text{supp}(\pi')$ and, for some fixed $1 \leq \bar{k} < |P|$, suppose

⁶²For the origin of (6), see [Block and Marschak \(1959\)](#).

$s_{\bar{k}}^\uparrow(P) = s_{\bar{k}}^\uparrow(P_u)$ for each u .⁶³ Then, if $\pi(P) \leq \sum_{u=1}^U \pi'(P_u)$, there exists a finite sequence of weighted Ryser swaps $(c_v R_v)_{v=1}^V$ and a finite set of paths $\{P_w\}_{w=1}^W$ satisfying $s_{\bar{k}+1}^\uparrow(P) = s_{\bar{k}+1}^\uparrow(P_w)$ for all w , such that:

$$\pi(P) \leq \sum_{w=1}^W \left(\pi'(P_w) + \sum_{v=1}^V c_v \langle R_v, e_{P_w} \rangle \right),$$

where e_{P_w} denotes the P_w -th standard Euclidean basis vector of $\mathbb{R}^{\mathcal{P}}$.

Proof. Let $(t_{\bar{k}+1}, h_{\bar{k}+1})$ denote the $(\bar{k} + 1)$ -st edge of the path P . First observe that each path in $\{P_u\}_{u=1}^U \subseteq \text{supp}(\pi')$ passes through node $t_{\bar{k}+1}$ since $t_{\bar{k}+1}$ is the head of the $\bar{k}$ -th edge in P and $s_{\bar{k}}^\uparrow(P) = s_{\bar{k}}^\uparrow(P_u)$ for each u . Now, since π and π' are path decompositions of the same quasi-flow, we have by hypothesis:

$$\sum_{P' \in \mathcal{P}} \pi'(P') \mathbb{1}_{\{(t_{\bar{k}+1}, h_{\bar{k}+1}) \in P'\}} = \sum_{P' \in \mathcal{P}} \pi(P') \mathbb{1}_{\{(t_{\bar{k}+1}, h_{\bar{k}+1}) \in P'\}} \geq \pi(P),$$

i.e. both π and π' put at least $\pi(P)$ mass on paths which pass through $(t_{\bar{k}+1}, h_{\bar{k}+1}) \in P$.

Let $\{P_v\}_{v=1}^V$ denote those paths in $\text{supp}(\pi')$ that contain $(t_{\bar{k}+1}, h_{\bar{k}+1})$ as an edge. Since each P_v and each P_u passes through node $t_{\bar{k}+1}$, each pair (P_u, P_v) is compatible at $t_{\bar{k}+1}$. Further, as $s_{\bar{k}}^\uparrow(P) = s_{\bar{k}}^\uparrow(P_u)$ for each u and $(t_{\bar{k}+1}, h_{\bar{k}+1}) \in P_v \cap P$ for each v , at least one path in the $t_{\bar{k}+1}$ -conjugate of each (P_u, P_v) agrees with P on $s_{\bar{k}+1}^\uparrow$. Let θ be the minimum between (i) the total amount of mass π' puts on paths in $\{P_u\}_{u=1}^U$ and (ii) the total mass π' puts on paths in $\{P_v\}_{v=1}^V$. By the preceding discussion, $\theta \geq \pi(P)$. For each pair of paths (P_u, P_v) , let

$$c_{uv} = \left(\frac{\pi'(P_u)}{\sum_{u'=1}^U \pi'(P_{u'})} \right) \left(\frac{\pi'(P_v)}{\sum_{v'=1}^V \pi'(P_{v'})} \right) \theta.$$

By construction, we get that $\sum_{u=1}^U \sum_{v=1}^V c_{uv} = \theta$. Further, $\sum_{u=1}^U c_{uv} = \pi'(P_v) \frac{\theta}{\sum_{v'=1}^V \pi'(P_{v'})} \leq \pi'(P_v)$, and analogously for P_u . Now let R_{uv} be the Ryser swap which puts -1 mass on paths P_u and P_v and $+1$ mass on their $t_{\bar{k}+1}$ -conjugate. Further, by previous arguments, the signed measure resulting from $\sum_{u=1}^U \sum_{v=1}^V c_{uv} R_{uv}$ puts θ mass on paths which agree with P on $S_{\bar{k}+1}^\uparrow(P)$. Hence we obtain a finite sequence of weighted Ryser swaps which shifts θ mass onto paths $\{P_w\}_{w=1}^W$ satisfying $s_{\bar{k}+1}^\uparrow(P) = s_{\bar{k}+1}^\uparrow(P_w)$, which gives:

$$\sum_{w=1}^W \left(\pi'(P_w) + \sum_{v=1}^V c_v \langle R_v, e_{P_w} \rangle \right) \geq \theta \geq \pi(P)$$

⁶³We use $|P|$ to denote the number of edges in P .

as desired while maintaining non-negativity of the resulting measure. $\square$

Lemma 2 (Zipper Lemma 2). Suppose π and π' are path decompositions of some quasi-flow f . Fix a path P in the support of π . Then there exists a finite sequence of weighted Ryser swaps with positive coefficients $\{c_i R_i\}$, $c_i > 0$, such that $\pi' + \sum_i c_i R_i = \pi''$, $\pi''(P) \geq \pi(P)$, and π'' is a path-decomposition.

Proof. Fix P and denote its first edge (s, n_1) . Since π and π' are path decompositions of f ,

$$\sum_{P' \in \mathcal{P}} \pi(P) \mathbb{1}_{\{(s, n_1) \in P'\}} = \sum_{P' \in \mathcal{P}} \pi'(P') \mathbb{1}_{\{(s, n_1) \in P'\}} \geq \pi(P) > 0.$$

As a consequence, the set of paths in $\text{supp}(\pi')$ which include (s, n_1) is non-empty; enumerate it $\{P_u\}_{u=1}^U$. Then by the above we have $\pi(P) \leq \sum_{u=1}^U \pi'(P_u)$, and hence the requirements of Lemma 1 are satisfied. However, note that the enumerated set of paths $\{P_w\}_{w=1}^W$ whose existence Lemma 1 asserts *also* satisfy the requirements of Lemma 1 (for the same fixed choice of P), but this time for $s_2^\uparrow(P)$. Repeating this process $|P|$ times yields a path decomposition π'' differing from π by a sum of weighted Ryser swaps, and an enumerated set of paths $\{P_l\}_{l=1}^L$ belonging to the support of π'' such that (i) $s_{|P|}^\uparrow(P) = s_{|P|}^\uparrow(P_l)$ and (ii) $\pi(P) \leq \sum_{l=1}^L \pi''(P_l)$, but (i) implies $L = 1$ and $P_l = P$, hence (ii) reduces to $\pi''(P) \geq \pi(P)$ as desired. $\square$

Lemma 3. Both π and π' are path decompositions of the same quasi-flow f if and only if there exists a finite sequence of Ryser swaps with positive coefficients $\{c_i R_i\}$, $c_i > 0$, such that $\pi + \sum_i c_i R_i = \pi'$.

Proof. ($\Leftarrow$): Suppose $\pi + \sum_i c_i R_i = \pi'$, where $\{c_i R_i\}$ is a finite sequence of weighted Ryser swaps. If two path decompositions differ by a single weighted Ryser swap, they are necessarily path decompositions of the same quasi-flow. By induction, we obtain π' and π are path decompositions of the same quasi-flow f .

($\Rightarrow$): Toward the converse, we provide an explicit algorithm to construct π from π' via a sequence of weighted Ryser swaps.

1. Initialize by enumerating the set of paths in the support of π' via $i \in \{1, \dots, I\}$ and set $i = 1$. Set $\pi_1 = \pi$ and set $\pi'_1 = \pi'$.
2. π_i and π'_i are path decompositions of the quasi-flow f_i .⁶⁴ As such, by Lemma 2, we obtain a path decomposition π''_i of the same quasi-flow as π_i, π'_i , but satisfying $\pi''_i(P_i) \geq \pi_i(P_i)$.

⁶⁴Here, $f_1 = f$ and for all $i > 1$, $f_i(e) = \sum_{\{P \in \mathcal{P} : e \in P\}} \pi_i(P)$.

3. Set $i = i + 1$. Set $\pi'_{i+1}(P) = \pi''_i(P) - \pi(P)\mathbb{1}_{\{P=P_i\}}$, and $\pi_{i+1}(P) = \pi_i(P) - \pi(P)\mathbb{1}_{\{P=P_i\}}$.
4. If $\pi_i = \mathbf{0}$, terminate the algorithm. If not, return to step 2.

At each step, the algorithm applies only finitely many weighted Ryser swaps. As it terminates with $\pi_i = \mathbf{0}$, it entails a finite sequence of weighted Ryser swaps which transforms π' into π , as desired. $\square$

Proof of Theorem 1 and Proposition 1

Theorem 1 follows from applying the following abstract result to the random utility graph defined at the start of this appendix. Proposition 1 then follows as an immediate corollary.

Theorem 8. A pair $\pi, \pi' \in \Delta(\mathcal{P})$ are path decompositions of the same quasi-flow if and only if $\pi - \pi' \in \mathcal{R}$. In particular, for any $\pi \in \mathcal{M} \subseteq \Delta(\mathcal{P})$, the set of decompositions of the same flow as π corresponds to $(\pi + \mathcal{R}) \cap \mathcal{M}$.

Proof. The first claim follows directly from Lemma 3. For the latter, observe that $\{\sum_i c_i R_i | c_i \in \mathbb{R}\}$ defines $\mathcal{R}$. So $\pi + \mathcal{R}$ is the set $\{\pi + \sum_i c_i R_i | c_i \in \mathbb{R}_{++}\}$, as the set of (unweighted) Ryser swaps is closed under multiplication by -1 . Thus again by Lemma 3 the claim follows. $\square$

A.2.1 Graphical Support Restrictions

Theorem 9. Let $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ be finite sequences of paths, allowing for repetition. $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ are rearrangement equivalent if and only if the sums:

$$\sum_{k=1}^K \mathbb{1}_{P_k} \quad \text{and} \quad \sum_{k=1}^K \mathbb{1}_{P'_k}$$

differ by a finite sum of unweighted Ryser swaps.

Before proving Theorem 9, we require a preliminary lemma.

Lemma 4. For any directed acyclic graph $(\mathcal{N}, \mathcal{E}, s, t)$, there is an enumeration of the set of nodes from 1 to $|\mathcal{N}|$ such that, for each i , there is no $j \geq i$ such that $(n_j, n_i) \in \mathcal{E}$.

Proof. Let $d : \mathcal{N} \rightarrow \mathbb{Z}_+$ denote the function which assigns to each node $n \in \mathcal{N}$ the length (i.e. number of edges) of the longest sequence of edges in $\mathcal{E}$ of the form: $s \rightarrow \dots \rightarrow n$ connecting s to n . As $(\mathcal{N}, \mathcal{E}, s, t)$ is finite and acyclic, d is well-defined.

Let $\succeq_d$ denote the weak order on $\mathcal{N}$ represented by d , and $\succeq^*$ any linear order extension of its asymmetric component. Let n_i denote the unique node such that $|\{n \in \mathcal{N} : n_i \succeq^* n\}| = i$.

It is straightforward to verify this gives an ordering $n_1, \dots, n_{|\mathcal{N}|}$. We claim this ordering has the desired monotonicity properties.

To see this, suppose for sake of contradiction that $(n_j, n_i) \in \mathcal{E}$, where $j \geq i$. Since $(\mathcal{N}, \mathcal{E}, s, t)$ is acyclic, $j > i$. For any sequence of edges $s \rightarrow \dots \rightarrow n_j$, by appending the edge $n_j \rightarrow n_i$ we obtain a sequence terminating at n_i , hence $d(n_i) > d(n_j)$ and $n_i \succ_d n_j$, and therefore $n_i \succ^* n_j$, which implies $i > j$, a contradiction. $\square$

We now move to our proof of [Theorem 9](#).

Proof. To begin, suppose that $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ are rearrangement equivalent. Choose some enumeration of the nodes of our graph satisfying the conditions of [Lemma 4](#). For a path P that passes through node n , let $s_n^\uparrow(P)$ denote the subset of edges in P that precede node n (and thus ends with an edge of form (m, n)). Similarly, for a path P which passes through node n , let $s_n^\downarrow(P)$ denote the subset of edges in P that come after node n (and thus begins with an edge of form (n, m)). We now propose an algorithm that takes a rearrangement and constructs Ryser swaps that induce the rearrangement.

1. Initialize at $j = 1$, and define the signed measure $\mu_0 = \sum_{k=1}^K \mathbb{1}_{P_k}$, and sequence of paths $P_k^0 \equiv P_k$, for all $k = 1, \dots, K$
2. For node n_j , let $\Pi_j \subseteq \{1, \dots, K\}$ consist of the indices of the paths in $P_1^{j-1}, \dots, P_K^{j-1}$ which pass through n_j . For any $i \in \Pi_j$, let $R_i^j = \mathbb{1}_{\{\tilde{P}_i, \tilde{P}_{\sigma_{n_j}^{-1}(i)}\}} - \mathbb{1}_{\{P_i, P_{\sigma_{n_j}^{-1}(i)}\}}$, where $\tilde{P}_i = s_{n_j}^\uparrow(P_i) s_{n_j}^{\downarrow}(P_{\sigma_{n_j}^{-1}(i)})$ and $\tilde{P}_{\sigma_{n_j}^{-1}(i)} = s_{n_j}^\uparrow(P_{\sigma_{n_j}^{-1}(i)}) s_{n_j}^{\downarrow}(P_i)$. Note every R_i^j , by definition of a rearrangement, is a valid Ryser swap. Define $\mu_j = \mu_{j-1} + \sum_{i \in \Pi_j} R_i^j$. This creates a new, $\mathbb{Z}_+$ -valued signed measure in $\mathbb{R}^{\mathcal{P}}$ which we may identify with a sequence of paths $P_1^j, \dots, P_K^j$, where:

$$P_k^j = \begin{cases} P_k^{j-1} & \text{if } k \notin \Pi_j \\ s_{n_j}^\uparrow(P_k^j) s_{n_j}^{\downarrow}(P_{\sigma_{n_j}^{-1}(k)}^j) & \text{else.} \end{cases}$$

3. Set $j = j + 1$. If $n_j = t$, terminate the algorithm. If $n_j \neq t$, return to step 2.

By construction, the algorithm terminates with a set of paths $P_1^{|\mathcal{N}|}, \dots, P_K^{|\mathcal{N}|}$ such that, if $P_k^{|\mathcal{N}|}$ is written as $s \rightarrow n_1 \rightarrow \dots \rightarrow t$, we have $n_{i+1} = \text{head}[P_{(\sigma_{n_i} \circ \dots \circ \sigma_s)^{-1}(k)}]_{\downarrow}^{n_i}$ and hence $P_1^{|\mathcal{N}|}, \dots, P_K^{|\mathcal{N}|}$ is precisely the sequence $P'_1, \dots, P'_K$. Therefore, $\sum_{k=1}^K \mathbb{1}_{P'_k} = \sum_{k=1}^K \mathbb{1}_{P_k} + \sum_{j=1}^{|\mathcal{N}|} \sum_{i \in \Pi_j} R_i^j$ as desired.

Conversely, suppose $\sum_{k=1}^K \mathbb{1}_{P'_k} - \sum_{k=1}^K \mathbb{1}_{P_k} = \sum_{l=1}^L R_l$. We now show their underlying sequences $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ are rearrangement equivalent by induction on L . Suppose

first that $L = 1$. By relabeling, without loss we may suppose the sequences $P_1, \dots, P_K$ to $P'_1, \dots, P'_K$ are identical but for two paths in each, P_i, P_j and P'_i, P'_j , and R_1 must correspond to swapping the outgoing terminal segments $s_{\downarrow}^{n^*}(P_i)$ and $s_{\downarrow}^{n^*}(P_j)$ to yield $P'_i = s_{n^*}^{\uparrow}(P_i)s_{\downarrow}^{n^*}(P_j)$ and $P'_j = s_{n^*}^{\uparrow}(P_j)s_{\downarrow}^{n^*}(P_i)$. In particular, $s_{n^*}^{\uparrow}(P'_i) = s_{n^*}^{\uparrow}(P_i)$ and $s_{n^*}^{\uparrow}(P'_j) = s_{n^*}^{\uparrow}(P_j)$. Then define a rearrangement where σ_s fixes every index except i, j, i', j' and maps $i \leftrightarrow i'$ and $j \leftrightarrow j'$, σ_{n^*} fixes every element in Π_{n^*} except i, j and maps $i \leftrightarrow j$, and σ_n is the identity for all $n \in \mathcal{N} \setminus \{s, n^*\}$. By construction, this defines a rearrangement taking $P_1, \dots, P_K$ to $P'_1, \dots, P'_K$.

Now, suppose that for $L \leq \bar{L}$ the claim holds, and let $\sum_{k=1}^K \mathbb{1}_{P'_k}$ and $\sum_{k=1}^K \mathbb{1}_{P_k}$ differ by a sum of $\bar{L} + 1$ weighted Ryser swaps. Consider first the case in which there exists some R_l , without loss $R_{\bar{L}+1}$, such that $\sum_{k=1}^K \mathbb{1}_{P_k} + R_{\bar{L}+1} \geq 0$. Then $\sum_{k=1}^K \mathbb{1}_{P_k} + R_{\bar{L}+1}$ differs from $\sum_{k=1}^K \mathbb{1}_{P'_k}$ by a sum of $\bar{L}$ unweighted Ryser swaps, and hence identifying $\sum_{k=1}^K \mathbb{1}_{P_k} + R_{\bar{L}+1}$ with a sequence $\tilde{P}_1, \dots, \tilde{P}_K$, by hypothesis this sequence is rearrangement equivalent to $P'_1, \dots, P'_K$. Moreover, since $1 \leq \bar{L}$, we obtain that similarly $P_1, \dots, P_K$ is rearrangement equivalent to $\tilde{P}_1, \dots, \tilde{P}_K$. Denote these rearrangements by $\{\sigma_n^1\}_{n \in \mathcal{N}}$ and $\{\sigma_n^2\}_{n \in \mathcal{N}}$ respectively. Then defining, for each $n \in \mathcal{N}$, $\sigma_n = \sigma_n^1 \circ \sigma_n^2$ defines a rearrangement under which $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ are equivalent.

On the other hand, suppose for every R_l , the sum $\sum_{k=1}^K \mathbb{1}_{P_k} + R_l \not\geq 0$. Define $\mathcal{Q}$ as the set of paths which receive negative mass under $\sum_{k=1}^K \mathbb{1}_{P_k} + R_{\bar{L}+1}$. Then $\sum_{k=1}^K \mathbb{1}_{P_k} + R_{\bar{L}+1} + \sum_{P \in \mathcal{Q}} \mathbb{1}_P$ and $\sum_{k=1}^K \mathbb{1}_{P_k} + \sum_{P \in \mathcal{Q}} \mathbb{1}_P$ differ by a single Ryser swap, and the former differs from $\sum_{k=1}^K \mathbb{1}_{P'_k} + \sum_{P \in \mathcal{Q}} \mathbb{1}_P$ by $\bar{L}$ Ryser swaps. By the preceding argument, the sequence representations of these last two expressions are rearrangement equivalent. We may regard the path(s) in $\mathcal{Q}$ as the last path(s) in each sequence; if this rearrangement does not fix the paths corresponding to these last indices, then the path(s) in $\mathcal{Q}$ appearing in the non-prime sequence correspond to some P'_i or P'_i and P'_j . In such a case, define a new rearrangement that is the identity at every node except s , and which at s maps (respectively) $i \leftrightarrow K+1$ or $i \leftrightarrow K+1$ and $j \leftrightarrow K+2$, and fixes every other index. Composing this new rearrangement with the former then gives a rearrangement between the augmented sequences which fixes the terms in the sequence corresponding to $\mathcal{Q}$, and hence restricts to a rearrangement between $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ as desired. $\square$

Given a directed acyclic graph, its set of paths $\mathcal{P}$, and some subset $\mathcal{P}'$, let $\mathbb{R}_{\mathcal{P}'}^{\mathcal{P}}$ denote the set of vectors in $\mathbb{R}^{\mathcal{P}}$ which take on value zero in the dimensions indexed by paths $P \in \mathcal{P} \setminus \mathcal{P}'$. By minor abuse of notation, let $\Delta(\mathcal{P}')$ denote the set of path decompositions whose support is a subset of $\mathcal{P}' \subseteq \mathcal{P}$.

Theorem 10. Fix a graph $(\mathcal{N}, \mathcal{E}, s, t)$. The following are equivalent.

1. For each $\pi \in \Delta(\mathcal{P}')$, there is no path decomposition $\pi' \neq \pi$ with $\pi' \in \Delta(\mathcal{P}')$ that induces the same flow as π .
2. For all $\pi \in \Delta(\mathcal{P}')$,

$$(\pi + \mathcal{R}) \cap \Delta(\mathcal{P}') = \{\pi\}.$$

3. $\mathcal{P}'$ contains no distinct, rearrangement-equivalent sequences of paths.
4. For every finite sequence of *unweighted* Ryser swaps $\{R_i\}_{i=1}^K$ (allowing for repetition):

$$\sum_{i=1}^K R_i \in \mathbb{R}_{\mathcal{P}'}^{\mathcal{P}} \iff \sum_{i=1}^K R_i = \mathbf{0}.$$

Proof. As $\Delta(\mathcal{P}') \subseteq \Delta(\mathcal{P})$, the equivalence between (1) and (2) is an immediate consequence of [Theorem 8](#). We now show (3) implies (1). By [Lemma 9](#), (1) holds if and only the restriction $\mathcal{M} = \Delta(\mathcal{P}') \cap \mathbb{Q}^{\mathcal{P}}$ is identifying. Suppose, for purposes of contraposition, that $\Delta(\mathcal{P}') \cap \mathbb{Q}^{\mathcal{P}}$ is not identifying. Then there exist distinct $\pi, \pi' \in \Delta(\mathcal{P}') \cap \mathbb{Q}^{\mathcal{P}}$ which decompose the same flow. By [Lemma 3](#), π and π' differ by a finite sum of Ryser swaps, $\pi - \pi' = \sum_{i=1}^K R_i$. Since both π and π' take only rational values, by multiplying by a sufficiently large positive integer, we obtain $\mathbb{Z}_+$ -valued signed measures $\bar{\pi}, \bar{\pi}'$. We may identify these signed measures with sequences of paths $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$, where $\bar{\pi} = \sum_{k=1}^K \mathbb{1}_{P_k}$ and $\bar{\pi}' = \sum_{k=1}^K \mathbb{1}_{P'_k}$. As these signed measures differ by a sum of Ryser swaps (by repeatedly summing copies of $\sum_i R_i$), by [Theorem 9](#), π_N and π'_N being rearrangement-equivalent. Thus $\neg(1)$ implies $\neg(3)$.

We now show that (4) implies (3). Once more, by means of contraposition, suppose $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ are distinct, rearrangement-equivalent sequences of paths. Then by [Theorem 9](#), $\sum_{i=1}^N R_i = \sum_{k=1}^K \mathbb{1}_{P_k} - \mathbb{1}_{P'_k} \neq \mathbf{0}$ is a finite sum of unweighted Ryser swaps. Furthermore, since $P_1, \dots, P_K$ and $P'_1, \dots, P'_K$ contain only paths in $\mathcal{P}'$, the sum $\sum_{i=1}^K R_i$ is also supported on $\mathcal{P}'$. Thus $\sum_{i=1}^k \mathbb{1}_{P_i} - \sum_{i=1}^k \mathbb{1}_{P'_i} = \sum_{i=1}^K R_i \neq \mathbf{0}$ and yet $\sum_{i=1}^k \mathbb{1}_{P_i} - \sum_{i=1}^k \mathbb{1}_{P'_i} = \sum_{i=1}^K R_i \in \mathbb{R}_{\mathcal{P}'}^{\mathcal{P}}$. Thus $\neg(3)$ implies $\neg(4)$.

Finally, we show (1) implies (4). To this end, suppose (4) is false. Thus there exists some finite sequence of unweighted Ryser swaps such that $\sum_{i=1}^K R_i \neq \mathbf{0}$ with $\sum_{i=1}^K R_i \in \mathbb{R}_{\mathcal{P}'}^{\mathcal{P}}$. Since each Ryser swap has two entries equal to one, two entries equal to negative one, and all other entries equal to zero, $\sum_{i=1}^K R_i$ puts equal mass on its negative entries and its positive entries. We can then renormalize and treat the mass on the negative entries as a path decomposition and the mass on the positive entries as a path decomposition. By construction, these differ by a finite sum of (potentially weighted) Ryser swaps. By [Lemma 3](#), they induce the same flow and, by construction, they have distinct supports and are thus

distinct. It immediately follows that condition (1) of [Theorem 10](#) fails. Thus together we have shown that $\neg(1) \implies \neg(3) \implies \neg(4) \implies \neg(1)$. $\square$

Corollary 2. For any $\mathcal{P}' \subseteq \mathcal{P}$, the flows $\{\mathbb{1}_{\{e \in P\}}\}_{P \in \mathcal{P}'}$ are linearly independent if and only if $\mathcal{P}'$ contains no distinct, rearrangement-equivalent sequences of paths.

Proof. ($\Leftarrow$): Suppose, by way of contraposition, that these flows are linearly dependent. Thus there exist scalars $\{w_P\}_{P \in \mathcal{P}'}$, not all zero, such that $\sum_{P \in \mathcal{P}'} w_P \mathbb{1}_{\{e \in P\}} = 0$. Define mass functions $\pi_+(P) = \max\{0, w_P\}$ and let $\pi_-(P) = \max\{0, -w_P\}$ in $\mathbb{R}^{\mathcal{P}}$. As every path leaves the source from a unique edge, and π_+ and π_- put the same total mass on each edge, π_+ and π_- have equal total mass. Further, as $\sum_{P \in \mathcal{P}'} [\pi_+(P) \mathbb{1}_{\{e \in P\}} - \pi_-(P) \mathbb{1}_{\{e \in P\}}] = 0$, i.e. the zero quasi-flow, we obtain that π_+ and π_- are path decompositions of the same quasi-flow, which is supported only on the edge set of paths in $\mathcal{P}'$. By [Theorem 10](#), $\mathcal{P}'$ then contains a pair of distinct, rearrangement-equivalent sequences, as desired.

($\Rightarrow$) : Conversely, suppose $\{\mathbb{1}_{\{e \in P\}}\}_{P \in \mathcal{P}'}$ are linearly independent. Thus these vectors span a simplex, and any path decomposition is necessarily unique. Then, by equivalence of (1) and (3) in [Theorem 10](#), $\mathcal{P}'$ contains no pair of distinct, rearrangement-equivalent sequences. $\square$

Theorem 11. Fix a graph $(\mathcal{N}, \mathcal{E}, s, t)$. As a slight abuse of notation, let $\Delta(\mathcal{P}')$ denote the set of path decompositions whose support is a subset of $\mathcal{P}' \subseteq \mathcal{P}$. The following are equivalent for $\pi \in \Delta(\mathcal{P}')$.

1. π is an extreme point of $\Delta(\mathcal{P}') \cap (\pi + \mathcal{R})$.
2. The set $\{\mathbb{1}_{\{e \in P\}} : \pi(P) > 0\}_{P \in \mathcal{P}'}$ is linearly independent.
3. $\text{supp}(\pi)$ contains no distinct, rearrangement-equivalent sequences of paths.
4. For every finite sequence of *unweighted* Ryser swaps $\{R_i\}_{i=1}^K$ (allowing for repetition):

$$\sum_{i=1}^K R_i \in \mathbb{R}_{\text{supp}(\pi)}^{\mathcal{P}} \iff \sum_{i=1}^K R_i = \mathbf{0}.$$

Proof. The set $\Delta(\mathcal{P}') \cap (\pi + \mathcal{R})$ defines a moment set in the sense of [Winkler \(1988\)](#), i.e. it is a set of probability measures which satisfy some auxiliary, finite system of (linear) equality constraints. By Proposition 2.1.a of [Winkler \(1988\)](#), π is an extreme point of $\Delta(\mathcal{P}') \cap (\pi + \mathcal{R})$ if and only if the set $\{\mathbb{1}_{\{e \in P\}} : \pi(P) > 0\}_{P \in \mathcal{P}'}$ is linearly independent. This shows the equivalence of (1) and (2). The equivalence between (2), (3), and (4) follows directly from [Corollary 2](#) and [Theorem 10](#). $\square$

Proof of Theorem 2 and Corollary 1

Our Theorem 2 corresponds to the equivalence between (1) and (3) in Theorem 11 applied to the random utility graph. Similarly, Corollary 1 follows from the equivalence between conditions (1) and (3) in Theorem 10 applied to the random utility graph.

A.2.2 Graphical Ordered Results

Fix a graph and let π be a path decomposition of quasi-flow f . Suppose that $\succeq$ is a linear order over the edges of the fixed graph. We say that π is **swap-progressive** (with respect to $\succeq$) if $\text{supp}(\pi)$ can be ordered $P_1, \dots, P_N$ such that, whenever P_i and P_j are compatible at some node n , then $i > j$ implies that $(n, m_i) \succeq (n, m_j)$ where (n, m_i) and (n, m_j) are the edges of P_i and P_j , respectively, leaving node n .

Theorem 12. Fix a graph, a quasi-flow function f , and a linear order $\succeq$ over the edges of our graph. There is some swap-progressive path decomposition of f and it is unique.

Proof. We first show uniqueness. Suppose, toward a contradiction, that the quasi-flow f admits two distinct swap-progressive path decompositions, π_1 and π_2 . Consider the zero-sum signed measure $\pi_1 - \pi_2$. Since π_1 and π_2 decompose the same quasi-flow, $\pi_1 - \pi_2 = \sum_{i=1}^N c_i R_i$ where $c_i R_i \in \mathcal{R}$ for all $i = 1, \dots, N$. Define the signed measures $\pi_+, \pi_- \in \mathbb{R}^{\mathcal{P}}$ via $\pi_+(P) = \max\{0, \pi_1(P) - \pi_2(P)\}$ and $\pi_-(P) = \max\{0, \pi_2(P) - \pi_1(P)\}$. As $\pi_1, \pi_2 \geq \mathbf{0}$, $\text{supp}(\pi_+) \subseteq \text{supp}(\pi_1)$ and similarly $\text{supp}(\pi_-) \subseteq \text{supp}(\pi_2)$. By restriction of the ordering of the supports of π_1 and π_2 , the supports of π_+ and π_- still satisfy the definition of swap-progressivity and moreover:

$$\sum_{P \in \text{supp}(\pi_+)} \pi_+(P) \mathbb{1}_{\{e \in P\}} = \sum_{P' \in \text{supp}(\pi_-)} \pi_-(P') \mathbb{1}_{\{e \in P'\}}$$

for every $e \in \mathcal{E}$ and thus are path decompositions of the same quasi-flow.

Let P_+ be the highest ordered path in the support of π_+ . As π_- equals π_+ modulo a finite sequence of weighted Ryser swaps, there is some path P_- in the support of π_- and some number $k > 0$ such that $s_k^\uparrow(P_-) = s_k^\uparrow(P_+)$. Let P_- be the highest ordered path in the support of π_- such that $s_k^\uparrow(P_-) = s_k^\uparrow(P_+)$ for some $k > 0$. Further, $s_k^\downarrow(P_-) \neq s_k^\downarrow(P_+)$ for the corresponding k by disjointedness of the supports of π_+ and π_- . Let e_+ and e_- denote the first edges at which P_+ and P_- disagree. Since π_- can be reached from π_+ via a sequence of weighted Ryser swaps, we get that there is some path, P'_+ , in the support of π_+ such that $e_- \in P'_+$. By P_+ being the highest ordered path in the support of π_+ , we get that $e_+ \succeq e_-$ in our exogenous ordering. Once again, by π_+ being π_- plus a sequence of Ryser swaps, we

get that there is some path in the support of π_- , P'_- , which has $e_+ \in P'_-$. There are two cases.

1. $s_{k'}^\uparrow(P'_-) = s_{k'}^\uparrow(P_+)$ for some $k' > 0$.

In this case, we have $P_- > P'_-$ by P_- being the highest ordered path for which $s_k^\uparrow(P_-) = s_k^\uparrow(P_+)$ for some $k > 0$. However, we have that $e_+ \succeq e_-$, which gives us $P'_- > P_-$ as both e_+ and e_- originate from the same node. This is a contradiction to swap-progressiveness.

2. $s_k^\uparrow(P'_-) \neq s_k^\uparrow(P_+)$ for all $k > 0$.

Let e'_+ and e'_- denote the edges where P'_- and P_+ first disagree. We now have $e'_+ \in P_+ \cap P_-$ (by $s_k^\uparrow(P_-) = s_k^\uparrow(P_+)$ for some $k > 0$ and e'_+ being the first edge of P_+) and $e'_- \in P'_-$. Once again, applying the Ryser swap theorem, we get that there is some path in the support of π_+ , P'_+ , such that $e'_- \in P'_+$. This, plus P_+ being the highest ordered path, gives us that $e'_+ \succeq e'_-$. However, $e_+ \succeq e_-$ means that $P'_- > P_-$ and $e'_+ \succeq e'_-$ means that $P_- > P'_-$. This is a contradiction to swap-progressiveness.

By contradiction in our two cases, we get that there cannot be two different swap-progressive path decompositions for the same flow function, and so we are done.

We now show that every quasi-flow function can be induced by a swap-progressive path decomposition. We present an algorithm which takes as input a quasi-flow and returns the unique swap-progressive path decomposition.

1. Initialize at $k = 0$. Set $f_0(e) = f(e)$ and $\pi_0(P) = 0$ for all $P \in \mathcal{P}$.
2. Now, let $k = k + 1$.
 - (a) Set $l = 1$, $n_1 = s$, and $\tilde{P}_k = \emptyset$. Let e be the highest ranked edge according to $\succeq$ at node n_l such that $f_k(e) > 0$. Set $\tilde{P}_k = \tilde{P}_k \cup \{e\}$ and $m = n_{l+1}$ where m is the head of e .
 - (b) Set $l = l + 1$. By property (i) of (4), at every interior node, we have that there is some edge e whose tail is n_l such that $f_k(e) > 0$.⁶⁵ Let e be the highest ranked edge according to $\succeq$ whose tail is n_l satisfying $f_k(e) > 0$. Set $\tilde{P}_k = \tilde{P}_k \cup \{e\}$ and $m = n_{l+1}$ where m is the head of e .
 - (c) If $n_{l+1} = t$, proceed to the Step 3. Otherwise, return to 2.(b).
3. We have now defined a path $P_k = \tilde{P}_k$. Set $\pi_k(P_k) = \min_{e \in P_k} f_k(e)$.

⁶⁵We note that the algorithm starts with property (i) of (4) holding. At every iteration of the algorithm, we subtract out a vector proportional to a flow along a single path, thus preserving property (i).

4. For $e \in P_k$, define $f_{k+1}(e) = f_k(e) - \pi_k(P_k)$. For all other e , define $f_{k+1}(e) = f_k(e)$.
5. If $f_{k+1}(e) = 0$ for all $e \in \mathcal{E}$, terminate, else return to step 2.

When this algorithm terminates, it produces a swap-progressive path decomposition. To see that it yields a path decomposition, note that at each iteration (i.e. for each k), we subtract $\min_{e \in P_k} f_k(e)$ from each $f_k(e)$ for $e \in P_k$. This means that firstly, $f_k \geq \mathbf{0}$ at each step of the algorithm. Second, at each iteration, we are zeroing the mass on some edge which previously had strictly positive weight. As the underlying graph is finite, we obtain that after finitely many steps, every edge will satisfy $f_k(e) = 0$.

We now argue this path decomposition is swap-progressive. Note that in each iteration of the algorithm, the path we construct always chooses the highest ranked available (i.e. with strictly positive weight) edge at each node it visits. Further, as discussed in the previous paragraph, the set of edges with strictly positive weights properly decreases with each iteration. It follows that, for $k < l$, whenever P_k and P_l visit the same node in the graph, P_k leaves on a weakly higher edge than P_l according to $\succeq$. This corresponds precisely to swap-progressivity. $\square$

Proof of Theorem 3

Proof. In the random utility graph, each edge is of the form $(A, A \setminus \{x\})$. Assign edges of this form to the alternative x . Given the linear order over alternatives $\succeq_X$, choose any linear order $\succeq_{\mathcal{E}}$ over edges in the random utility graph such that $x \succeq_X y$ implies that $(A, A \setminus \{x\}) \succeq_{\mathcal{E}} (B, B \setminus \{y\})$. Part (i) from Theorem 3 now follows immediately from applying Theorem 12 to the random utility graph using the linear order $\succeq_{\mathcal{E}}$.⁶⁶

We now prove claim (ii). First, we show that a single-crossing representation is also a swap-progressive representation. Suppose that $\succeq_X$ is the linear order over alternatives for both the single-crossing and swap-progressive representations. Suppose towards a contradiction that $\succ_i$ and $\succ_j$, both in the support of the single-crossing representation, are non-trivially compatible, $i > j$, and that $x_j^{k+1} \succeq x_i^{k+1}$. Note that $x_j^{k+1} \succ_j x_i^{k+1}$. This, the single-crossing property, $x_j^{k+1} \succeq x_i^{k+1}$, and $i > j$ tells us that $x_j^{k+1} \succ_i x_i^{k+1}$. However, we have that $x_i^{k+1} \succ_i x_j^{k+1}$, which is a contradiction. Thus any single-crossing representation is also swap-progressive.

Now, it follows from Filiz-Ozbay and Masatlioglu (2023) that a progressive representation is a random utility representation if and only if it is single-crossing. Thus, by the preceding paragraph, if a swap-progressive representation is single-crossing, it is also progressive. Since

⁶⁶Note that we could have used any order over edges for this construction. This translates to part (i) from Theorem 3 holding even in the case of a menu dependent order over alternatives.

every swap-progressive representation is a random utility representation, it follows that every swap-progressive representation which is also progressive must be a random utility representation and thus single-crossing. $\square$

A.3 Parametric Invertibility Results

A.3.1 Topological Preliminaries

Let $\mathcal{X}, \mathcal{Y}$ be topological spaces. A continuous map $p : \mathcal{X} \rightarrow \mathcal{Y}$ is a **covering map** if, for every $y \in \mathcal{Y}$ there exists a connected open neighborhood V such that $p^{-1}(V)$ is a disjoint union of open sets $\{U_i\}_{i \in \mathcal{I}}$ in $\mathcal{X}$ each of which p maps homeomorphically onto V . When $\mathcal{Y}$ is connected, the cardinality of the set $\mathcal{I}$ does not depend on the choice of $y \in \mathcal{Y}$, and $|\mathcal{I}|$ is referred to as the number of **sheets** of the covering p . A homeomorphism $D : \mathcal{X} \rightarrow \mathcal{X}$ is called a **deck transformation** if $(p \circ D)(x) = p(x)$ for all $x \in \mathcal{X}$. The set of all deck transformations for a covering map form a group under composition.

Fix $x_0 \in \mathcal{X}$ and $y_0 \in \mathcal{Y}$. The structure of a covering map is intimately tied to the fundamental groups $\pi_1(\mathcal{X}, x_0)$ and $\pi_1(\mathcal{Y}, y_0)$.⁶⁷ We gather several relevant results below.

Lemma 5 (Bredon 2013, Lemma III.4.4). Suppose D is a deck transformation for p such that $D(x) = x$ for some $x \in \mathcal{X}$. Then D is the identity.⁶⁸

Lemma 6 (Bredon 2013, Corr. III.5.3). If $\pi_1(\mathcal{X}, x_0)$ is trivial, then the cardinality of $\pi_1(\mathcal{Y}, y_0)$ equals the number of sheets of p .

Lemma 7 (Bredon 2013, Corr. III.6.10). If $\pi_1(\mathcal{X}, x_0)$ is trivial, the group of deck transformations of p is isomorphic to $\pi_1(\mathcal{Y}, y_0)$.

We say that a continuous map between topological spaces is **proper** if the preimage of any compact set is compact. The following shows condition (ii) of Theorem 4 gives a simple characterization of properness.

Lemma 8 (Lee 2012 Prop. 2.17). The map $\bar{\varphi}$ satisfies condition (ii) of Theorem 4 if and only if it is proper.

The final ingredient we require is a fixed-point theorem for periodic functions with prime period.⁶⁹

⁶⁷For the basic definitions and theory of fundamental groups, see, e.g. Bredon (2013) Chapter III.

⁶⁸The statement here is consequence of Lemma III.4.4 in Bredon (2013) where, in his notation, $W = X$ and $f = p$. Alternatively, our statement here appears as an unlabeled remark at the top of p. 148.

⁶⁹While Smith (1934) does not explicitly state the requirement that n be prime, this requirement is implicit in his proof. See also Eilenberg (1940), Footnote 2.

Theorem (Smith 1934). Suppose $\mathcal{X} \subseteq \mathbb{R}^n$ is contractible, and $h : \mathcal{X} \rightarrow \mathcal{X}$ a homeomorphism such that the n -fold composition $(h \circ \dots \circ h) = \text{id}$ for prime $n \in \mathbb{N}$. Then h has a fixed point.

A.3.2 Proof of Theorem 4

Proof. Necessity of (i) is standard, whereas (ii) follows from Lemma 8. We now consider sufficiency. Condition (i) implies that $\bar{\varphi}$ is a local diffeomorphism (Lee 2013, Prop. 4.8). In particular, this implies $\bar{\varphi}$ is an open map, hence $\bar{\varphi}(\Theta) \subseteq \mathbb{R}^d$ is open and therefore a smooth manifold. Furthermore, Lemma 8 implies $\bar{\varphi}$ is proper. Finally, as $\mathcal{X}$ is contractible, it is connected, hence so too is $\bar{\varphi}(\Theta)$. Then Lee (2013) Prop. 4.46 implies $\bar{\varphi} : \Theta \rightarrow \bar{\varphi}(\Theta)$ is a smooth covering map.

Fix some vector $y_0 \in \bar{\varphi}(\Theta)$. As $\bar{\varphi}$ is a covering map, the pre-image $\bar{\varphi}^{-1}(y_0)$ is discrete; since $\bar{\varphi}$ is proper, $\bar{\varphi}^{-1}(y_0)$ is finite, and by the connectedness of $\bar{\varphi}(\Theta)$, the preimage of every $y \in \bar{\varphi}(\Theta)$ has the same cardinality. Now, as Θ contractible, it is simply connected, thus Lemma 6 implies $\pi_1(\bar{\varphi}(\Theta), y_0)$ is finite, and by Lemma 7, $\pi_1(\bar{\varphi}(\Theta), y_0)$ acts on Θ via deck transformations such that the identity $\mathbb{1} \in \pi_1(\bar{\varphi}(\Theta), y_0)$ is the only element mapped to the identity deck transform $\Theta \rightarrow \Theta$.⁷⁰ Suppose now, for sake of contradiction, that $\bar{\varphi}(\Theta)$ is not simply connected. Then there exists some $\alpha \in \pi_1(\bar{\varphi}(\Theta), y_0)$, where $\alpha \neq \mathbb{1}$. Since $\pi_1(\bar{\varphi}(\Theta), y_0)$ is finite, α must be of finite order, hence for some $n > 1$, we have $\alpha^n = \mathbb{1}$, and $\alpha^m \neq \mathbb{1}$ for all $m < n$. Without loss of generality, we may suppose n is prime.⁷¹ Then under the action of $\pi_1(\bar{\varphi}(\Theta), y_0)$ on Θ , α corresponds to some non-identity deck transformation $D_\alpha : \Theta \rightarrow \Theta$ of prime period n . But by Smith's Theorem, D_α must have a fixed point, thus Lemma 5 implies D_α is the identity, a contradiction. Thus $\pi_1(\bar{\varphi}(\Theta), y_0) = \{\mathbb{1}\}$ and Lemma 6 implies the preimage $\bar{\varphi}^{-1}(y_0)$, and hence $\bar{\varphi}^{-1}(y)$ for every $y \in \bar{\varphi}(\Theta)$, is singleton; in particular this means $\bar{\varphi}$ is injective. By Lee (2013) Prop. 4.33.b, $\bar{\varphi}$ is then a diffeomorphism, as desired. $\square$

⁷⁰This action is simply the isomorphism whose existence is asserted by Lemma 7. Since the kernel of any group isomorphism is trivial, the claim follows.

⁷¹If n is not prime, let m be any prime factor of n and consider $\alpha^{\frac{n}{m}}$.

Online Appendix for “Identification in Stochastic Choice”

Peter Caradonna* and Christopher Turansick†

February 23, 2026

B Beyond Random Utility

B.1 Random Choice Models

The proof of [Theorem 5](#), as well as the corresponding analogues of [Theorem 2](#) and [Theorem 3](#), follow from [Theorem 8](#), [Theorem 11](#), and [Theorem 12](#) applied to the following graphical representation of random choice rules.

Let $\Sigma \subseteq 2^X \setminus \{\emptyset\}$ denote the collection of observable menus, and index these $A_1, \dots, A_{|\Sigma|}$. Define a graph with $\mathcal{N} = \Sigma \cup \{\emptyset\}$, and for every $1 \leq i \leq |\Sigma|$ and for each $x \in A_i$ there is a unique edge $A_i \rightarrow A_{i+1}$ in $\mathcal{E}$, where we define $A_{|\Sigma|+1} = \emptyset$.¹ For the edge $e \in \mathcal{E}$ of the form $A_i \rightarrow A_{i+1}$ associated with the alternative $x \in A_i$, we define $f(e) = \rho(x, A_i)$. Since $\rho \geq 0$ and $\sum_{x \in A_i} \rho(x, A_i) = 1$, by [\(4\)](#), f is a flow.

In the random choice graph, each path corresponds to a sequence of the form $(x_1, A_1), (x_2, A_2), \dots, (x_{|\Sigma|}, A_{|\Sigma|})$. This path exactly corresponds to the choice function which chooses x_i from A_i . Thus there is a one-to-one correspondence between $\mathcal{P}$ and $\mathcal{C}_\Sigma$, and hence between path decompositions of f and random choice representations. Moreover, a function $f : \mathcal{E} \rightarrow \mathbb{R}_+$ admits a path decomposition if and only if it is a (quasi-)flow, hence path decompositions on this graph precisely correspond to the identified sets of the random choice model.

We now proceed with the proof of [Theorem 5](#).

*Division of the Humanities and Social Sciences, Caltech. Email: ppc@caltech.edu.

†Department of Decision Sciences, Bocconi University. Email: christopher.turansick@unibocconi.it.

¹Thus $\mathcal{E}$ is a multiset.

Proof. As an application of [Lemma 3](#) to the graphical representation described above, we have observational equivalence between two distributions over choice functions, ν and μ , if and only if $\nu = \mu + \sum_{i=1}^n c_i R'_i$ where R_i do not correspond to the Ryser swaps in [Theorem 5](#) but the Ryser swaps from [Lemma 3](#). For the purposes of this proof, call the Ryser swaps from [Theorem 5](#) pseudo-Ryser swaps. We now show that every pseudo-Ryser swap can be generated by a series of Ryser swaps and every Ryser swap can be generated by a series of pseudo-Ryser swaps.

Consider an unweighted Ryser swap. Given two choice functions (c_0, c_1) it creates choice functions (c_2, c_3) where c_2 agrees with c_0 strictly above A_i (according to the indexing of sets used in our graphical representation) and agrees with c_1 weakly below A_i . Further, c_3 agrees with c_1 strictly above A and agrees with c_0 weakly below A_i . Now our aim is to construct a pseudo-Ryser swap out of Ryser swaps. We consider a Ryser swap between c_0 and c_1 at choice set A_i . Start with two choice functions (c_0, c_1) . Ryser swap these choice functions at choice set A_i . This induces choice functions (c_2, c_3) . Consider the choice set A_{i+1} . Now Ryser swap (c_2, c_3) at choice set A_{i+1} . The resulting choice functions (c_4, c_5) satisfy the following.

1. c_4 agrees with c_0 on every set $B \neq A_i$ and agrees with c_1 at A_i
2. c_5 agrees with c_1 on every set $B \neq A_i$ and agrees with c_0 at A_i

Thus (c_4, c_5) are the resultant choice functions of a pseudo-Ryser swap on (c_0, c_1) .

Now we show that a Ryser swap can be constructed from pseudo-Ryser swaps. A Ryser swap swaps the choices between c_0 and c_1 at every choice set A_j satisfying $j \geq i$ for some choice set A_i . Simply apply a pseudo-Ryser swap at each of these choice sets. This sequence of pseudo-Ryser swaps results in a Ryser swap. The prior equivalence shows that the span of Ryser swaps and the span of pseudo-Ryser swaps is the same. As such, by [Lemma 3](#), we get that [Theorem 5](#) holds. $\square$

B.2 Frame Dependent Random Utility

As mentioned in [Section 7](#), there is a more general version of the frame-dependent random utility model. Instead of an alternative being framed and not framed, there are now multiple frames. We label these frames as $\{1, \dots, M\}$. In addition, there is a second version of the model where the frames are labeled as $\{0, 1, \dots, M\}$ and the zero frame corresponds to an alternative not being available. For now, we will focus on the version of the model without the zero frame.

Let x^m denote alternative x under frame m . Mirroring our notation from [Section 7](#), define $X^* = \{x_1^1, \dots, x_K^1, \dots, x_1^M, \dots, x_K^M\}$ denoting the set of alternative-frame pairs. A

menu corresponds to some subset $A \subsetneq X^*$ satisfying, for each $i \in \{1, \dots, K\}$, there is exactly one m such that $x_i^m \in A$. Thus a menu contains every alternative x_i with each alternative under some alternative-dependent frame m .

The frame-dependent random utility model considers strict preferences over the set X^* under one additional assumption. The model restricts to preferences $\succ$ such that, if $m > m'$, then $x_i^m \succ x_i^{m'}$. Given a preference $\succ$, observe that if, for $m > 1$ and arbitrary $i \neq j$, $x_j^1 \succ x_i^m$, then there is no menu A in which x_i^m is maximal. With this in mind, we now define a truncated preference $\succ^*$. Given a preference $\succ$, let $\succ^*$ be the ranking which agrees with $\succ$ up to, and including, the best alternative under frame 1. That is, $\succ^*$ is formed by taking $\succ$ and removing all alternatives x_i^m which rank below the highest alternative of the form x_j^1 . Let $\mathcal{L}^*$ denote the set of truncated strict preferences.

The data are compatible with the frame-dependent random utility model if there exists some distribution $\mu \in \Delta(\mathcal{L}^*)$ such that:

$$\rho(x, A) = \sum_{\succ^* \in \mathcal{L}^*} \mu(\mathcal{L}^*) \mathbb{1}_{\{x \text{ is } \succ^* \text{-maximal in } A\}}$$

for all menus A and each $x \in A$. Here, x corresponds to some alternative x_i^m and menu A satisfies the assumption we have made in this section. As before, we say that two truncated preferences $\succ_1^*, \succ_2^* \in \mathcal{L}^*$ are k -compatible if they agree on their k -best alternatives, though not necessarily their ranking. A Ryser swap for the frame-dependent model is simply a signed measure in $\mathbb{R}^{\mathcal{L}^*}$ which places unit mass on a k -compatible pair $\succ_1^*$ and $\succ_2^*$ and mass negative one on the pair of truncated preferences obtained by swapping their k -initial segments. Let $\mathcal{R}^{fd} \subset \mathbb{R}^{\mathcal{L}^*}$ denote the linear span of these vectors. [Theorem 6](#) still holds under this formulation of the frame-dependent random utility model and $\mathcal{R}^{fd}$.

The graphical representation of the frame dependent random utility model is due to [Cheung and Masatlioglu \(2024\)](#) although we make a slight alteration to fit exactly into our framework. Let $\mathcal{X}$ denote the collection of all menus $A \subsetneq X^*$ satisfying, for each $i \in \{1, \dots, K\}$, there is exactly one m such that $x_i^m \in A$. Consider the directed acyclic graph where $\mathcal{N} = \mathcal{X} \cup \{t\}$ and the edge set is described as follows.

- For each $A \in \mathcal{X}$ and each $x_i^m \in A$ with $i > 1$, there is an edge from A to $(A \setminus \{x_i^m\}) \cup \{x_i^{m-1}\}$, $A \rightarrow (A \setminus \{x_i^m\}) \cup \{x_i^{m-1}\}$
- For each $A \in \mathcal{X}$ and each $x_i^1 \in A$, there is an edge from A to t , $A \rightarrow t$

Here, the source node corresponds to the node A^* where A^* is the menu given by $\{x_1^M, \dots, x_K^M\}$. In order to describe our flow function, first observe that each menu A can be represented as a vector in $\bar{A} \in \mathbb{R}^X$. Each dimension of $\bar{A}$ is indexed by some element $x \in X$. The dimension

of $\bar{A}$ corresponding to alternative x is given by the frame associated with x in menu A . Thus, if $x_i^m \in A$ then the i -th dimension of $\bar{A}$ is m . Under this observation, the vectors $\bar{A}$ can be partially ordered by the greater than or equal to relationship. As an abuse of notation, we rank menus $A \geq B$ if $\bar{A} \geq \bar{B}$. Given the frame-dependent random choice rule $\rho(x, A)$, we can implicitly define a new object:

$$\rho(x, A) = \sum_{A \leq B: \bar{A}_x = \bar{B}_x} y(x, B),$$

where $\bar{A}_x$ denotes the entry in vector $\bar{A}$ in the dimension indexed by x . [Cheung and Masatlioglu \(2024\)](#) provides a closed form expression for $y(x, A)$, but, to keep things simple, we only present the recursive definition here. For our graphical representation, we use $y(x, A)$ as the flow assigned to the edge leaving node A which corresponds to alternative x . [Cheung and Masatlioglu \(2024\)](#) shows that, whenever the frame-dependent random choice rule has a frame-dependent random utility representation, this constitutes a flow as in [Equation 4](#).

Our interpretation of the frame-dependent random utility graph relies on our notion of a truncated preference. Consider the path given by $A^* \rightarrow B \rightarrow \dots \rightarrow t$. Observe that B differs from A^* by a single element and only by the frame of that element. Specifically, A^* has x_1^M while B has x_1^{M-1} . This generalizes to successive nodes in our path. We now rewrite our path with superscripts on each $\rightarrow$ corresponding to the difference between these two successive nodes:

$$A^* \rightarrow^{x_1^M} B \rightarrow^{x_i^m} \dots \rightarrow^{x_j^1} t.$$

This path is then associated with the truncated preference $x_1^M \succ^* x_i^m \succ^* \dots \succ^* x_j^1$. Just as in the previous cases, a path decomposition putting mass on a path corresponds to a probability distribution putting mass on its corresponding truncated preference. [Theorem 6](#) is then recovered by applying [Theorem 8](#) to the frame-dependent random utility graph.

As we mentioned earlier, there is a second version of the frame-dependent random utility model which allows for variation in which alternatives are available. This can be encoded as a zero frame which places sufficiently negative utility on an alternative in order to make it never be chosen. The graphical representation which [Cheung and Masatlioglu \(2024\)](#) provides for this version of the model satisfies the assumptions we make on the graphs we consider in [Theorem 8](#). As such, our results can be applied directly out of the box to the frame-dependent random utility model with availability variation. In this case, types correspond to full, strict preferences (i.e. not truncated) over X^* .

B.3 Dynamic Discrete Choice

In this section, we extend the setup we described in [Section 7](#) to allow each period's menu to depend on the previous period's choice. There is a sequence $t = 1, \dots, T$ of discrete time periods. In each, an agent chooses from some menu $\emptyset \subsetneq A_t \subseteq X$. We assume A_1 is exogenously given and, for all $2 \leq t \leq T - 1$, that the menu A_t faced by the agent at time t is determined by:

$$A_t(x_{t-1}) \equiv g_{t-1}(x_{t-1}),$$

where $x_{t-1} \in X$ denotes the agent's choice at $t - 1$, and $g_{t-1} : X \rightarrow 2^X \setminus \{\emptyset\}$ is a fixed evolution function, describing how the agent's choice at $t - 1$ determines their choice set at time t .

We consider an empiricist who observes a system of conditional choice probabilities. Formally, this consists of (i) a distribution of time-one choice probabilities $\rho_1 \in \Delta(X)$ supported on A_1 , and (ii) a family of functions $\rho_t : X \times X \rightarrow [0, 1]$, for $2 \leq t \leq T$, such that:

$$\sum_{a \in A_t(b)} \rho_t(a|b) = 1,$$

whenever $\rho_{t-1}(b|\cdot)$ is not everywhere zero, and $\rho_t(a|b) = 0$ for all $x \in X$ otherwise. Given such data, we seek to describe the set of compatible distributions over choice *histories*.

To represent this problem as one of describing flow functions, we define $\mathcal{N}$ recursively, as the set of all pairs in (t, x) such that either (i) $t = 1$ and $x \in A_1$, or (ii) $2 \leq t \leq T$, and $x \in A_t(x')$ for some $(t - 1, x') \in \mathcal{N}$, as well as an abstract source and sink, s and t . Similarly, we let $\mathcal{E}$ consist of all directed edges of the form $(t - 1, x') \rightarrow (t, x)$, where $x \in A_t(x')$, as well as edges from the source s to each pair in $\{1\} \times A_1$, and edges from each pair in $\mathcal{N}$ with $t = T$ to the sink t .

Given such a graph, we recursively define the flow from the data ρ_t . For the edge connecting the source node to $(1, x)$, we assign $\rho_1(x)$ as the flow assigned to that edge. For edges connecting $(t - 1, x')$ to (t, x) , we normalize $\rho_t(x|x')$ by multiplying by the total in-flow into node $(t - 1, x')$ and assign this new value as the flow for this edge. Finally, for the edge connecting (T, x) to the sink, we simply assign the total in-flow into node (T, x) as the flow. It is straightforward to verify that our construction defines a flow. In fact, our construction recovers the absolute frequency of conditional choices in each period. Each path in our constructed graph corresponds to a full history of choices and a flow function of this graph assigns frequencies to each full history.

[Theorem 7](#) can be recovered by applying [Lemma 3](#) to this graphical representation under the assumption that the choice set is fixed across time periods and $g_{t-1}(\cdot)$ is equal to our fixed

choice set. A version of [Theorem 7](#) can be recovered for the more general setup described here. We simply need to restrict ourselves to considering swaps between histories which are possible given $g_{t-1}(\cdot)$.

B.4 Correlated Random Utility

In addition to the three types of models we propose in [Section 7](#), our techniques can also be used to study identification in the correlated random utility model of [Chambers et al. \(2024\)](#) as well as the consumption dependent random utility model of [Turansick \(2024\)](#). Suppose that an analyst has access to a random joint choice rule, $\rho : X^2 \times (2^X \setminus \{\emptyset\})^2 \rightarrow [0, 1]$, which satisfies, for all non-empty $A, B \subseteq X$,

$$\sum_{x \in A} \sum_{y \in B} \rho(x, y, A, B) = 1.$$

The **correlated random utility model** asks that there is distribution $\mu \in \Delta(\mathcal{L}^2)$ such that

$$\rho(x, y, A, B) = \sum_{\succ \in \mathcal{L}} \sum_{\succ' \in \mathcal{L}} \mu(\succ, \succ') \mathbb{1}_{\{x \succ A \setminus x, y \succ' B \setminus y\}}.$$

The correlated random utility model represents choice in two periods where choices in the two periods may be correlated through the distribution over preference pairs, but, conditional on this draw, the agent's choice in the first period has no impact on their choice in the second period. On the other hand, the **consumption dependent random utility model** asks that there is a distribution $\mu \in \Delta(\mathcal{L})$ as well as distributions for each alternative preference pair, $(x, \succ)$, $\nu_{(x, \succ)} \in \Delta(\mathcal{L})$ such that

$$\rho(x, y, A, B) = \sum_{\succ \in \mathcal{L}} \mu(\succ) \mathbb{1}_{\{x \succ A \setminus x\}} \sum_{\succ' \in \mathcal{L}} \nu_{(x, \succ)}(\succ') \mathbb{1}_{\{y \succ' B \setminus y\}}.$$

The consumption dependent random utility model represents choice in two periods where an agent's preference, and thus choice, in the second period depends both on their preference and choice in the first period. In each of these models, first period choices are well defined independently of the second period's choice set. As such, we use $\rho_1(x, A) = \sum_{y \in B} \rho(x, y, A, B)$ for any choice of non-empty B .

Both of these models share the same graphical representation, but, while every path on this graph corresponds to some type in the consumption dependent random utility model, the correlated random utility model constitutes a support restriction on this graph. The graphical representation for these two models consists of a series of graphs, one representing

choice in the first period and the rest, one for each edge of the first period graph, representing choice in the second period. These series of graphs can then be made into a single graph by replacing each edge of the first period graph with the corresponding second period graph.

To begin, the first period graph is exactly the random utility graph with flows coming from ρ_1 and Equation 6. Each second period graph has the same node and edge set as the random utility graph. That is $\mathcal{N} = 2^X$ and $\mathcal{E} = \{(A, B) \mid B = A \setminus \{a\}, a \in A\}$, where $s = X$ and $t = \emptyset$. The second period graphs differ from the first period graphs in their quasi-flow function. For the second period graph associated with the edge $(A \rightarrow A \setminus \{a\})$ in the first period graph, the random joint choice rule ρ induces a unique flow as follows:

$$f(B \rightarrow B \setminus \{b\}) = \sum_{A \subseteq A'} \sum_{B \subseteq B'} (-1)^{|A' \setminus A| + |B' \setminus B|} \rho(a, b, A', B'). \quad (7)$$

The function f given in Equation 7 defines a quasi-flow as shown in Chambers et al. (2024).

In the consumption dependent random utility model, a type corresponds to a preference in the first period and $|X|$ preferences in the second period. Each of these $|X|$ preferences correspond to the type's preference in the second period conditional on choosing alternative x in the first period. As is the case with the random utility graph, in the first period graph, as well as each of the second period graphs, each path is bijectively associated with a preference. Recall that this bijection is given by regarding a path:

$$X \rightarrow X \setminus \{x_1\} \rightarrow \cdots \rightarrow X \setminus \{x_1, \dots, x_{N-1}\} \rightarrow \emptyset$$

as a sequence of nested lower contour sets, which uniquely defines the preference $x_1 \succ \cdots \succ x_N$. With this bijection in mind, the path on the first period graph defines the preference which dictates the agent's first period choice. Given this path and edge $(A, A \setminus \{a\})$ in this path, a path on the second period graph associated with edge $(A, A \setminus \{a\})$ on the first period graph corresponds to the agent's second period preference conditional on choosing a in the first period. The consumption dependent random utility model puts no restriction on which paths are allowed in the second period graph while the correlated random utility model asks that, for a given type, all paths on the second period graphs are the same.

This graphical construction is most easily understood via a first period graph and a series of second period graphs. We can directly apply our result about path decomposition to each of these graphs to discuss uniqueness in the correlated and consumption dependent random utility models. As mentioned earlier, we can alternatively replace each edge in the first period graph with its corresponding second period graph to get a single graph which acts as a graphical representation for these two models. In this case, we need only apply our

graphical results to this single graph.

While our focus in this section was to discuss how to extend the graphical representation of the random utility model in order to capture multidimensional analogues of the random utility model, this process generalizes to other models which can be represented by path decompositions of quasi-flows on directed acyclic graphs. This allows for discussion of both correlation of types/paths across dimension, as in the correlated random utility model, as well as dependence of later types/paths on earlier types-choice/path-edge pairs, as in the consumption dependent random utility model.

B.5 Maximization of Other Types of Orders

In the main body of the paper, we focused on agents who maximize linear orders. However, an analyst may be wary of this assumption and instead only ask that agents maximize a weak order, an interval order, or a semiorder. Under differing data assumptions than what we use for the random utility model, [Davis-Stober et al. \(2018\)](#) provides graphical representation for models which consider a distribution over agents maximizing each of these three types of orders. [Doignon and Saito \(2023\)](#) also studies the graphical representation of these models in order to characterize adjacency of extreme points in their corresponding polytopes. The techniques we develop here can be applied to discuss uniqueness in each of these models.

C Further Results

C.1 Technical Results

Lemma 9. Let $\mathcal{R}$ denote the subspace of $\mathbb{R}^{\mathcal{P}}$ spanned by vectors of the form:

$$\mathbb{1}_{\{P, P'\}} - \mathbb{1}_{\{P'', P'''\}}$$

where (P, P') are compatible and (P'', P''') are their corresponding conjugate, and by minor abuse of notation let $\Delta(\mathcal{P}')$ denote the face of $\Delta(\mathcal{P}) \subseteq \mathbb{R}^{\mathcal{P}}$ spanned by the abstract simplex $\emptyset \subsetneq \mathcal{P}' \subseteq \mathcal{P}$. Suppose there exists $t \in \mathbb{R}^{\mathcal{P}}$ such that:

$$\dim \left[(\mathcal{R} + t) \cap \Delta(\mathcal{P}') \right] > 0.$$

Then there exists $t' \in \mathbb{R}^{\mathcal{P}}$ such that $(\mathcal{R} + t') \cap \Delta(\mathcal{P}')$ contains a pair of distinct points in $\mathbb{Q}^{\mathcal{P}}$.

Proof. Let $\mathcal{V}$ denote the subspace of $\mathbb{R}^{\mathcal{P}}$ given by:

$$\text{span} \{e^P - e^{P'}\}_{P, P' \in \mathcal{P}'},$$

where e^i denotes the i th standard Euclidean basis vector. By hypothesis,

$$K = \dim \mathcal{R} \cap \mathcal{V} > 0.$$

From their definitions, both $\mathcal{R}$ and $\mathcal{V}$ admit bases $\{q_{\mathcal{R}}^i\}_{i=1}^{\dim(\mathcal{R})}$ and $\{q_{\mathcal{V}}^j\}_{j=1}^{\dim(\mathcal{V})}$ which belong to $\mathbb{Q}^{\mathcal{P}}$. Define the $|\mathcal{P}| \times \dim(\mathcal{R})$ and $|\mathcal{P}| \times \dim(\mathcal{V})$ matrices:

$$Q_{\mathcal{R}} = \begin{bmatrix} q_{\mathcal{R}}^1 & \cdots & q_{\mathcal{R}}^{\dim(\mathcal{R})} \end{bmatrix}$$

and

$$Q_{\mathcal{V}} = \begin{bmatrix} q_{\mathcal{V}}^1 & \cdots & q_{\mathcal{V}}^{\dim(\mathcal{V})} \end{bmatrix},$$

respectively, and let:

$$Q = \begin{bmatrix} Q_{\mathcal{R}} & -Q_{\mathcal{V}} \end{bmatrix}$$

Since each of the above matrices consists exclusively of rational elements, their respective column spaces trivially admit a bases in $\mathbb{Q}^{\mathcal{P}}$. This implies that, by Gaussian elimination, the annihilators of their column spaces admit a bases in $\mathbb{Q}^{\mathcal{P}}$; in particular the annihilator of the column space of Q admits a rational basis $\{r^K\}_{k=1}^K$, where each vector $r^k = [r_{\mathcal{R}}^k \mid r_{\mathcal{V}}^k]$.

Let $\bar{q}^k = Q_{\mathcal{R}} r_{\mathcal{R}}^k (= Q_{\mathcal{V}} r_{\mathcal{V}}^k)$. By construction, $\{\bar{q}^k\}_k \subset \mathbb{Q}^{\mathcal{P}}$ and, by construction $\{\bar{q}^k\}_k$ form a basis for $\mathcal{R} \cap \mathcal{V}$. Since $\dim(\mathcal{R} \cap \mathcal{V}) > 0$, there exists some non-zero $\bar{q}^k \in \mathcal{R} \cap \mathcal{V} \cap \mathbb{Q}^{\mathcal{P}}$. Then letting:

$$t' = \frac{1}{|\mathcal{P}'|} \mathbb{1}_{\mathcal{P}'},$$

we have t' and $t' + \alpha \bar{q}^k$, for small enough $\alpha \in \mathbb{Q}_{++}$, are distinct rational vectors in $(\mathcal{R} + t') \cap \Delta(\mathcal{P}')$ as desired. $\square$

C.2 Order-Based Restrictions and Extreme Points

We begin by noting that every swap-progressive path decomposition of a quasi-flow is an extreme point of the set of path decompositions for the given quasi-flow.

Proposition 2. Consider a graph with quasi-flow function f . Each swap-progressive path decomposition of f is an extreme point of the set of path decompositions of f .

Proof. Suppose we have a swap-progressive path decomposition π for some quasi-flow func-

tion f . Note that any path decomposition with the same support is also a swap-progressive path decomposition as swap-progressivity is a property of the support. By [Theorem 12](#), we know that a swap-progressive path decomposition is unique. This means that no other path decomposition whose support is contained within $\text{supp}(\pi)$ decomposes the same quasi-flow. Thus, the set of flows $\{\mathbb{1}_{\{e \in P\}}\}_{P \in \text{supp}(\pi)}$ is linearly independent. By [Corollary 2](#) and [Theorem 11](#), we get that π is an extreme point of the set of path decompositions inducing quasi-flow f . $\square$

Now we note that our notion of swap-progressivity, for both graphs and random utility, is a special case of a more general notion of an ordered restriction. Consider a graph $(\mathcal{N}, \mathcal{E}, s, t)$ and the set of paths $\mathcal{P}$. Consider some linear order over $\mathcal{P}$, $\succeq_{\mathcal{P}}$. For ease, $\succeq_{\mathcal{P}}$ is equivalent to an enumeration of $\mathcal{P}$ where $P_i \succeq_{\mathcal{P}} P_j \iff i \leq j$. Given an ordering over edges $\succeq_{\mathcal{E}}$, we can induce some ordering over paths $\succeq_{\mathcal{P}}$. We do this by starting at node s and saying that a path $P \succeq_{\mathcal{P}} P'$ if $(s, n) \succeq_{\mathcal{E}} (s, n')$ where (s, n) and (s, n') are the first edges of P and P' respectively. Generally, this will not yet leave us with a full linear order over paths so we iterate on this construction. Now move onto any node n for which each edge of the form (m, n) has already been considered. At node n , each path P which passes through node n will already be partially ordered. Further, by construction, incomparability by this partial order is an equivalence relation among those paths which pass through node n . For each of these equivalence classes, say that $P \succeq_{\mathcal{P}} P'$ if $(n, m) \succeq_{\mathcal{E}} (n, m')$ where (n, m) and (n, m') are edges in P and P' respectively. By iterating this process through each node, we are left with a linear order over paths, $\succeq_{\mathcal{P}}$. Recall that our construction of a swap-progressive proceeds as follows.

1. Set $i = 1$.
2. Put as much mass on P_i as possible (i.e. the minimum of $f(e)$ among edges in P_i). Call this amount $\pi(P_i)$.
3. Set $f(e) = f(e)$ if $e \notin P_i$. Set $f(e) = f(e) - \pi(P_i)$ if $e \in P_i$.
4. Set $i = i + 1$. If $i > |\mathcal{P}|$, terminate the construction. Otherwise, return to step 2.

Notably, the construction works for any linear order over paths $\succeq_{\mathcal{P}}$ and, given a quasi-flow function f , the induced path decomposition of the quasi-flow function will be unique given $\succeq_{\mathcal{P}}$. We call a path decomposition induced by such a construction an **ordered path decomposition**. Single-crossing, swap-progressive, and progressive representations are ordered path decompositions of the correctly chosen graph, but there are also ordered path decompositions which are not single-crossing, swap-progressive, or progressive. While ordered path

decompositions share many of the nice identification properties of swap-progressivity, ordered path decompositions are not sufficiently general enough to recover every extreme point of the set of path decompositions for every quasi-flow.

Proposition 3. There exist graphs $(\mathcal{N}, \mathcal{E}, s, t)$, flows f , and extreme points of the set of path decompositions of f such that any ordered decomposition of f does not induce the extreme point. Further, the random utility graph has this property.

Proof. Consider the graph in Figure 2 of [Chambers and Turansick \(2025\)](#) and the uniform distribution over the eight paths considered in Example 4.1 of the same paper, i.e. putting $\frac{1}{8}$ mass on each of the eight paths. Chambers & Turansick show the indicators of these eight paths form a linearly independent set of flows. As such, by [Winkler \(1988\)](#), the uniform distribution is an extreme point of the set of edge decompositions inducing the corresponding flow. Now, fix any ordered path decomposition π . By definition (i.e. the construction in the paragraph preceding [Proposition 3](#)), the highest ranked path in $\text{supp}(\pi)$ must contain some edge unique to it among paths in $\text{supp}(\pi)$. However, as every path in Example 4.1 of [Chambers and Turansick \(2025\)](#) shares each of its edges with at least one other path in the example, the uniform distribution over these paths cannot arise as an ordered path decomposition, for any choice of order $\succeq_{\mathcal{E}}$; as the graph in this example is obtained as a support restriction of the random utility graph, we are done. $\square$

D Examples and Computations Omitted From Text

D.1 Failure of Identification in [Example 8](#)

Let $X = \{x_0, x_1, x_2\}$ and consider any random choice rule ρ which satisfies:

$$(i) \quad \rho(x_1, x_0x_1) = \rho(x_2, x_0x_2) = \frac{1}{11} \quad \text{and} \quad (ii) \quad \rho(x_1, x_0x_1x_2) = \rho(x_2, x_0x_1x_2).$$

(Recall here we only consider menus which include the outside option $x_0 \in X$, hence this completely specifies ρ up to the choice frequency of x_0 from X .) Consider again the submodel s defined in [Example 8](#). For $i = 1, 2$ we have:

$$s^i(v_1, c_1, v_2, c_2) = \frac{1 + v_i}{v_i(1 + v_i c_i)} = 10,$$

or, simplifying:

$$v_i^2 c_i = 10$$

for all $i = 1, 2$. Similarly, by hypothesis we have:

$$s^{i+2}(v_1, c_1, v_2, c_2) = \frac{1 + v_i + v_{i+1}}{v_i + v_i v_{i+1} + c_i v_i^2} = K \quad (8)$$

for each $i = 1, 2$ and some $K > 0$, hence simplifying yields:

$$v_1 = v_2.$$

Thus, there are only two pieces of information we have not determined yet: the value of v_1 (without loss), and the value of K in (8), which pins down the choice probability of the outside option from the menu X . From (8), we obtain:

$$\frac{1 + 2v_1}{v_1 + v_1^2 + 10} = K,$$

or:

$$K(10 + v_1 + v_1^2) = 1 + 2v_1.$$

When, e.g., $K = 0.3$ this has multiple solutions, i.e. $v_1 = \frac{5}{3}$ and $v_1 = 4$. ■

D.2 A Non-Single Crossing Example

Let $X = \{a, b, c, d\}$, and consider the random choice rule ρ , which chooses uniformly from any menu. For any linear ordering $\succeq$ of alternatives, ρ fails to satisfy the centrality axiom of [Apesteguia et al. \(2017\)](#), and hence does not admit a single-crossing representation. However, ρ is clearly consistent with the random utility model: for example it is rationalized by the uniform distribution on $\mathcal{L}$.

By [Theorem 3](#), the random choice rule ρ admits a swap-progressive rationalization for any order $\succeq$. To see how swap-progressivity pins down a unique representation, first note that the uniform distribution on $\mathcal{L}$ places positive (and equal) mass on six distinct, non-overlapping collections of four preferences, each consisting of a 2-compatible pair and its 2-conjugate.² Given $\succeq$, the unique swap-progressive representation places (equal) mass on only two of the four preferences in each collection, namely those satisfying:

$$(i) \quad \{x^1 \succeq x^2 \text{ and } x^3 \succeq x^4\} \quad \text{or} \quad (ii) \quad \{x^2 \succeq x^1 \text{ and } x^4 \succeq x^3\}.$$

In other words, on each collection, the swap-progressive representation uses the order $\succeq$ to

²These collections each consist of, for any $\{x, y\} \subset X$, the four preferences whose initial strings are xy or yx .

select one of the two compatible pairs to receive full mass.³ Concretely, if $a \succeq b \succeq c \succeq d$, the unique swap-progressive representation places equal mass on:

$$\succsim_1: dcba \quad \succsim_2: dbca \quad \succsim_3: dabd \quad \succsim_4: cdab \quad \succsim_5: cbda \quad \succsim_6: cadb$$

$$\succsim_7: bdac \quad \succsim_8: bcad \quad \succsim_9: badc \quad \succsim_{10}: adbc \quad \succsim_{11}: acbd \quad \succsim_{12}: abcd.$$

Thus, even when a single-crossing representation may fail to exist, swap-progressivity provides a natural criteria for selecting rationalizations in a manner that incorporates the natural ordering of the environment. ■

³See also [Example 6](#).