

Binary Choice Paths and Random Utility*

Rohan Dutta[†] Yusufcan Masatlioglu[‡] Xiaosheng Mu[§]

October 17, 2026

Abstract

A decision maker proceeds through a list, making binary choices between each new alternative and the current provisional choice. We study such binary choice path data, aggregated across a population with heterogeneous preferences, as in the random utility model. Two simple axioms characterize complete data rationalizable by this model. Complete data trivially point-identifies the preference distribution, but remarkably, so does some partial data. We derive a necessary condition and a separate sufficient condition for such partial data. These conditions allow us to construct a minimal set of lists that achieves point-identification, which features a vanishing share of all feasible lists as the number of alternatives grows. For partial data that satisfies the sufficient condition for identification, we show that rationalizability again admits a simple characterization.

1 Introduction

A decision maker (DM) faces the set of all feasible alternatives X , as a list. She proceeds through the list, making binary choices between each new alternative and the current

*We thank Jean-Marie Dufour, Larry Epstein, David Freeman, Zilin Jiang, Fedor Sandomirsky, and seminar audiences at the 7th Columbia Conference on Economic Theory, BRIC 2026, and SAET 2026 for helpful comments.

[†]Department of Economics, McGill University, E-mail: rohan.dutta@mcgill.ca.

[‡]Department of Economics, University of Maryland, E-mail: yusufcan@umd.edu.

[§]Department of Economics, Princeton University, E-mail: xmu@princeton.edu

provisional choice. Call the resulting sequence of choices a binary choice path. We study a novel form of data consisting of sets of lists, and for each, a distribution over binary choice paths. Call it random binary choice path data. Using this data, we study the random utility hypothesis (RUH), which holds that aggregate choice behavior reflects the decisions of heterogeneous preference-maximizing individuals. We ask whether RUH imposes testable restrictions on the observed distributions of binary choice paths and, if so, which features of the data are necessary or sufficient for consistency with the hypothesis. For data consistent with RUH, labeled rationalizable, we further ask whether the underlying distribution of preference orderings can be uniquely identified, and whether such *point-identification* requires observing the full dataset.

The classical random utility model (RUM) also investigates RUH, hence the primitives are exactly the same. But it considers a different form of data, that we call *standard*. Standard data consists of subsets of X , and for each a distribution over its elements. Such data poses considerable difficulty with identification of the distribution of preferences and characterization with partial data. This has led most applications of the random utility model to impose far more structure. Our study asks whether our new form of data can overcome these challenges. In this, we follow the broader agenda pursued in [Caplin and Dean \[2011\]](#) and [Caplin et al. \[2011\]](#), of studying non-standard data generated from carefully engineered choice environments to identify key model parameters. Despite its non-standard nature in the literature, similar data is collected in industry through a process called adaptive choice-based conjoint analysis. We discuss this in detail in section 6.

Rationalizable random binary choice path data exhibits two properties of a local nature, in that they relate to choice path probabilities from two lists that are different only by the swap of a single pair of adjacent entries. For two lists that differ only by the position of the first two alternatives, a DM with a fixed preference ordering would generate the exact same binary choice path. The probability of observing such a choice path from either list should therefore be the same. This is the first property. The second relates to lists that differ by a single pair of adjacent entries that are not the first two. DMs with preference profiles that generate a specific choice path in one of these lists are identical to those that generate one of two specific choice paths in the other. Therefore, the

probability of observing the former choice path given the first list must be the same as the sum of probabilities of the two latter choice paths from the second list. This is the second property. Remarkably, these two local properties are also sufficient for complete random binary choice path data to be rationalizable (Theorem 1 and 2). The result is in sharp contrast to the global nature of the characterization of RUM with standard choice data, that features the Block-Marschak (BM) polynomials.

Standard data, even when complete, can fail to point-identify the preference distribution, as shown in [Fishburn, 1998]. By contrast, rationalizable complete random binary choice path data point-identifies the preference distribution, trivially. For each list there exists a binary choice path that is generated by a single preference profile alone. The probability of such a choice path for that list reveals the probability of the corresponding preference profile. With data from all possible lists, point-identification is therefore trivial. More interestingly, partial data too can point-identify the preference distribution.

Beyond the direct revelation from a list described above, data from multiple lists can be combined methodically to indirectly reveal the probabilities of other preference profiles. This is what permits data from incomplete sets of lists to point-identify the preference distribution. We establish a necessary condition that such a collection of lists must satisfy (Theorem 3). The condition partitions the set of all lists in a particular manner and requires the collection to contain at least one list from each cell of that partition. The result implicitly sets a lower bound on the number of lists needed for point-identification. For N alternatives the bound is $N!/2^{\lfloor N/2 \rfloor}$.

Toward a sufficient condition for identification, we develop a different approach by building on the key insight that, if rationalizable, data from some lists fully reveal data from some other lists in two specific ways. First, data from a list pins down data from the adjacent list that simply interchanges the first pair of entries. Second, data from a pair of lists that are related in a particular way, together pin down data from a third list. We formalize these two properties as operations by which a set of lists can “generate” other lists. Given the property of direct revelation mentioned above, for any set of lists, its span under the two operations delivers a partial identification result. If a set of lists spans all lists, then data from such a *spanning set* achieves point-identification. This is our sufficiency result for identification (Theorem 4).

The problem of finding minimal sets of lists that span all, and therefore achieve point-identification, is of a combinatorial nature that gets remarkably complex as the number of alternatives increase. While the problem is trivial for $N = 2$ and we can fully characterize it for $N = 4$, for general N we instead construct a class of such minimal sets. Despite the complexity of the problem, the construction is both simple and easy to implement (Theorem 5). The size of these minimal sets matches the aforementioned lower bound implied by the necessary condition. In particular, it requires a vanishing share of all feasible lists as the number of alternatives gets large.

Characterization of RUM with partial standard data is notoriously difficult, as we discuss in the literature review. We too are interested in random choice path data that is partial, in that it features data from fewer than all possible lists. In general, rationalizability of such data can be reframed as the non-emptiness of the core of an appropriately defined cooperative game. While it can be difficult to verify the non-emptiness condition, we show that rationalizability of partial data from a spanning set is amenable to a simple characterization. This is our final result (Theorem 6). Table 1 summarizes how our random binary choice path data compares with standard data.

			Complete	Partial
Type of Data	Standard	Characterization	✓ (one global axiom)	?
		Identification	×	×
	Binary Choice Path	Characterization	✓ (two local axioms)	✓ (for spanning sets)
		Identification	✓ (immediate)	✓

Table 1: Comparison with Standard Data

Related Literature

We first discuss the counterparts to our results in the standard RUM framework, concerning characterization and identification. We then turn to the literature on sequential choice in decision theory, which aims to explain observed behavioral phenomena. Although our approach shares with the behavioral literature an emphasis on non-standard

choice data, our objective differs. Rather than modeling behavioral phenomena, we seek to characterize and understand the aggregate implications of rational decision-making.

Random Utility Model

CHARACTERIZATION: The Random Utility Model (RUM) was first introduced by [Block and Marschak \[1959\]](#), who posed a foundational question: under what conditions can observed choice probabilities be rationalized as arising from a distribution over deterministic utility maximizers? To address this question, they introduced the Block-Marschak (BM) polynomials (expressions based on inclusion-exclusion systems) that provide necessary conditions for stochastic choice data to be consistent with RUM. This contribution extended the classical revealed preference theory to probabilistic environments.¹ Subsequently, [Falmagne \[1978\]](#) showed that non-negativity of the BM polynomials is also sufficient for RUM.²

BM polynomials characterize the facets of the RUM polytope but are often treated as technical axioms because of their limited economic interpretability. In response, several authors have sought more intuitive formulations and interpretations; see, for example, [Billot and Thisse \[2005\]](#).

LIMITED DATA: Falmagne’s characterization assumes that choice frequencies are observable for all menus. By contrast, [Kono et al. \[2026\]](#) provides a characterization in which choice probabilities need not be observed for every choice set. Related work considers environments with even more limited data.³ One strand of this literature focuses on binary choice data. [Dridi \[1980\]](#) shows that every system of binary choice probabilities satisfying the triangular inequalities admits a random utility representation when the universal set contains no more than five alternatives; this result fails when the set contains six or more alternatives. [Fishburn \[1992\]](#) offers a comprehensive historical survey

¹[McFadden and Richter \[1990\]](#) later proposed an alternative characterization, the Axiom of Revealed Stochastic Preference (ARSP), which is both necessary and sufficient for RUM. However, ARSP requires checking a countably infinite number of conditions, even when the choice set is finite.

²This result was independently established by [Barberá and Pattanaik \[1986\]](#). [Cohen \[1980\]](#) further extended the sufficiency result to infinite choice sets by adopting finitely additive probability measures.

³See for instance, [Barseghyan et al. \[2021\]](#) and [Epstein and Patel \[2026\]](#).

of this literature. More broadly, characterizing random utility models under limited data remains a difficult problem and is still largely open.

IDENTIFICATION: A well-known limitation of RUM is the lack of uniqueness of its representation under standard choice data. While uniqueness holds when there are only three alternatives [Block and Marschak, 1959], it typically fails in larger domains [Barberá and Pattanaik, 1986, Fishburn, 1998]. The failure of uniqueness can be severe: two distinct RUM representations may have disjoint supports [Fishburn, 1998]. Turansick [2022] is the first to provide necessary and sufficient conditions for the existence of a unique RUM representation without imposing restrictions on the support.

Because uniqueness is often viewed as a desirable property, a parallel literature studies subclasses of preferences that restore it. For example, Gul and Pesendorfer [2006] shows that uniqueness obtains when utility functions are linear, though Lin [2020] demonstrates that this result is fragile: uniqueness fails once independence is weakened to the broader betweenness class. Apesteguia et al. [2017] imposes a single-crossing condition on the support of the random utility representation and shows that uniqueness is recovered. Filiz-Ozbay and Masatlioglu [2023] later establishes that the single-crossing condition yields uniqueness even beyond the rational choice paradigm.⁴

Sequential Choice

Our paper is also related to the literature on sequential choice in decision theory. Broadly speaking, this literature follows two approaches. The first treats the sequential procedure as internal to the decision maker and therefore unobservable to the analyst. The second assumes that the order of sequential choice is an explicit part of the decision problem and is observable.

SEQUENTIAL CHOICE AS AN UNOBSERVABLE DECISION PROCEDURE: Early models in this tradition include Tversky [1969], Tversky [1972], and Rubinstein [1988]. While the specific heuristics are different, they all feature procedures where alternatives are compared or even eliminated in a sequential manner. In Tversky [1972], for instance, the

⁴Additional related results can be found in Dridi [1994], Manzini and Mariotti [2018], Lu and Saito [2018], Yang and Kopylov [2023], Petri [2023], Suleymanov [2024], Cattaneo et al. [2024], Chambers and Turansick [2024], Valkanova [2024], Caradonna and Turansick [2026].

DM selects aspects salient to a choice problem, sequentially one at a time, and eliminates alternatives that lack that aspect. These models show that behaviour generated by coherent and well-defined procedures may violate standard consistency axioms, while explaining systematic departures from standard rational choice. They constitute early careful studies of procedural rationality, as formalized in [Simon \[1976\]](#).

While earlier works emphasize underlying procedures, [Manzini and Mariotti \[2007\]](#) shows how such models of bounded rationality can be characterized by their observable choice implications. The DM in their model eliminates alternatives by the sequential application of two rationales, modeled as asymmetric binary relations. While the rationales themselves are unobservable, [Dutta and Horan \[2013\]](#) shows that they can be partially inferred from choice data. Subsequent work has analyzed numerous variants of the procedure.⁵ [Dutta \[2020\]](#) studies a multiple-rationale variant that generates stochastic choice data. [Masatlioglu and Nakajima \[2013\]](#) makes a connection to the important literature on consideration sets by modeling the formation of such sets as an unobservable dynamic search process, a sequential procedure.

CHOOSING A SINGLE ALTERNATIVE FROM A LIST: The study of choice from lists is motivated by many real-world decision environments. Consumers browse ranked search results on platforms such as Amazon, encounter ordered lists of advertisements, and navigate alternatives through web search engines. The seminal paper by [Rubinstein and Salant \[2006\]](#) formalizes decision problems in which alternatives are presented as lists.⁶

CHOICE PATH FROM A SET OF ALTERNATIVES: [Caplin and Dean \[2011\]](#) introduces choice process data to study satisficing behavior. Similar to our data, it records not only the final choice but also the sequence of intermediate decisions. This is key to revealing the dynamics of search and evaluation, enabling direct tests of models of sequential search and satisficing. Unlike our setting, in the theoretical analysis of [Caplin and Dean \[2011\]](#), alternatives are presented as unordered sets. The experimental study in [Caplin](#)

⁵See [Houy \[2008\]](#), [Au and Kawai \[2011\]](#), [Manzini and Mariotti \[2012\]](#), [Mandler et al. \[2012\]](#), [Apesteguia and Ballester \[2013\]](#), [Horan \[2016\]](#).

⁶This framework has since generated a large and influential literature (see also [Horan \[2010\]](#), [Caplin et al. \[2011\]](#), [Guney \[2014\]](#), [Yildiz \[2016\]](#), [Dimitrov et al. \[2016\]](#), [Aguiar et al. \[2016\]](#), [Kovach and Ülkü \[2020\]](#), [Ishii et al. \[2021\]](#), [Tserenjigmid \[2021\]](#), [Yegane \[2022\]](#), [Manzini et al. \[2024\]](#)).

et al. [2011], however, presents alternatives as explicit lists, and find that subjects tend to follow the list order during search. Geng [2016] further exploits choice process data to study the effects of default options on choices and attention allocation.

CHOOSING A SINGLE ALTERNATIVE FROM A LIST OF SETS: Papi [2012] introduces a novel data set to study the satisficing model of Simon [1955]. The decision maker sequentially explores a series of nested menus, a generalization of the standard list model in which alternatives are examined one at a time. Koshevoy and Savaglio [2023] relaxes the nested-menu assumption to study the implications of a no-regret condition. Cheung and Masatlioglu [2025] employ a related data set in which each alternative is assigned to a distinct category representing a different frame, so that the collection of sets forms a partition. They study frame-dependent utility using both deterministic and probabilistic data.

CHOICE PATH FROM A LIST: To the best of our knowledge, the only other paper that utilizes a data set similar to ours is Yang [2025]. In a deterministic setting, Yang [2025] establishes a connection between choice overload and the primacy effect. Like much of the existing literature, this paper uses the novel data structure to study a behavioral phenomenon that cannot be analyzed using standard choice data. Our approach departs from this behavioral focus. Rather than emphasizing behavioral anomalies, we use the novel data set to address questions within the rational choice paradigm, including meaningful characterizations under both full and limited data, as well as point-identification in settings with limited observability.

2 Notation

Let N be a given positive integer, and X be a finite set of N alternatives. A list is a sequence of distinct elements of X , represented as $L = [x_1 x_2 x_3 \dots x_\ell]$ for some $\ell \leq N$. We use $L(k)$ to denote the k -th entry of L , and use $|L|$ to denote its length ℓ . Let $\mathcal{L}_N$ be the set of all lists of length $\ell = N$. Let $>$ denote a strict preference ordering over X , and $\mathcal{P}$ denote the set of all such orderings.

A list $L \in \mathcal{L}_N$ presents our decision-maker (DM) with a sequence of binary choices. She starts with the first two alternatives on the list and chooses one. This is her provi-

sional choice. She then faces the third option on the list. The DM could stay with the provisional choice or switch to the third option, which becomes the new provisional choice. This continues iteratively till the end of the list. Call such a sequence of choices a *binary choice path*. At each pairwise choice, a DM with preference $\succ$ chooses the alternative that ranks higher under $\succ$.

Binary choice paths for a given list admit a more convenient representation as binary strings, which are finite sequences of 0s and 1s. For $\ell \leq N - 1$, let $s \in \{0, 1\}^\ell$ denote a binary string of length ℓ , where $s(k)$ corresponds to the k -th entry of s . Let S be the set of all binary strings of length $N - 1$.

We represent the choices of a DM with preference $\succ$ as a binary string of length $N - 1$. With some abuse of notation, we will use $\succ$ to denote a function that encodes the binary choice paths of a DM with preference $\succ$ facing different lists. Define

$$\succ: \mathcal{L}_N \rightarrow S$$

through the following iterative procedure. Fix some $L \in \mathcal{L}_N$. First, compare $L(1)$ and $L(2)$ and define $s(1)$ and the provisional choice c_1 as follows:

$$s(1) = \begin{cases} 0 & \text{if } L(1) \succ L(2) \\ 1 & \text{if } L(2) \succ L(1) \end{cases} \quad c_1 = \begin{cases} L(1) & \text{if } s(1) = 0 \\ L(2) & \text{if } s(1) = 1. \end{cases}$$

Intuitively, the string representation treats $L(1)$ as the initial provisional choice. When the DM gets to the second alternative $L(2)$ in the list, if $L(1)$ is preferred to $L(2)$, then $s(1)$ is zero (no change) and the provisional choice is still $L(1)$. Otherwise, $s(1)$ is 1, corresponding to a change in the provisional choice to the new alternative $L(2)$.

The above comparison can be naturally extended to later alternatives in the list. At step k for each $1 < k \leq N - 1$, the DM compares the current provisional choice c_{k-1} to the new alternative $L(k + 1)$. Define $s(k)$ and the new provisional choice c_k as

$$s(k) = \begin{cases} 0 & \text{if } c_{k-1} \succ L(k + 1) \\ 1 & \text{if } L(k + 1) \succ c_{k-1} \end{cases} \quad c_k = \begin{cases} c_{k-1} & \text{if } s(k) = 0 \\ L(k + 1) & \text{if } s(k) = 1. \end{cases}$$

The resulting binary string $s = (s(1)s(2) \dots s(N - 1))$ is defined to be $\succ(L)$. It encodes the binary choice path of a DM with preference $\succ$, where $s(k) = 0$ indicates retaining

the provisional choice at step k , and $s(k) = 1$ indicates switching to the newly presented alternative.

Table 2 below shows the binary strings that emerge from different combinations of lists and preferences with $X = \{a, b, c\}$. Consider, for instance, the list $L = [bac]$ (third row in the table). A DM with preference $a \succ_1 b \succ_1 c$, facing this list, first switches to a , and then stays with a , which generates the binary string $\succ_1(L) = (10)$. The same string is observed if the preference were $a \succ_2 c \succ_2 b$ instead. On the other hand, a DM with preference $c \succ_5 a \succ_5 b$ generates the binary string $\succ_5(L) = (11)$ (first switches to a , and then switches to c). If preference were $c \succ_6 b \succ_6 a$, the observed string would be $\succ_6(L) = (01)$.

	Preferences					
Lists	$a \succ_1 b \succ_1 c$	$a \succ_2 c \succ_2 b$	$b \succ_3 a \succ_3 c$	$b \succ_4 c \succ_4 a$	$c \succ_5 a \succ_5 b$	$c \succ_6 b \succ_6 a$
$[abc]$	00	00	10	10	01	11
$[acb]$	00	00	01	11	10	10
$[bac]$	10	10	00	00	11	01
$[bca]$	01	11	00	00	10	10
$[cab]$	10	10	11	01	00	00
$[cba]$	11	01	10	10	00	00

Table 2: Decision-Maker's (DM) Binary Choice Paths and Strings

A few remarks are due. First, although different lists lead to different binary strings for the same individual, the final choice is not affected by the list since our DMs are rational. Second, for a given list, we can uniquely recover the binary choice path from the binary string that it generates, thus no information is lost in working with binary strings. Finally, as the list varies, the same binary string in general corresponds to different binary choice paths.

Now consider a group of rational individuals who follow the choice procedure described above but have heterogeneous preferences distributed according to $\pi \in \Delta(\mathcal{P})$. Their aggregate choice behavior is described by the probability of observing each binary

string s under each list L . This probability is given by

$$\rho_\pi(s, L) = \sum_{>: s=>(L)} \pi(>).$$

Similarly, data in our analysis take the form of distributions over binary choice paths for different lists, and we can equivalently work with distributions over binary strings. Thus, in our model, the analyst observes a function $\rho(s, L)$ that specifies the probability of binary string s when the list is L . For any $L \in \mathcal{L}_N$, $\rho(\cdot, L)$ is a distribution over S .

Our main interest is in determining when these observed distributions can arise from the aggregate behavior of rational agents. Formally, we say ρ is rationalizable if there exists a distribution of preferences $\pi \in \Delta(\mathcal{P})$ such that $\rho = \rho_\pi$.

3 Characterization

In this section, we investigate the observable implications of our choice environment. We introduce some further notation for the subsequent exposition. For a sequence of lists $L_1, L_2, \dots, L_m$, no two of which share any alternative, let $L_1 L_2 \dots L_m$ be the list that is the concatenation of the m lists. We skip the bracket notation in concatenated lists for notational ease. For instance, we write $L_1 ab L_2$ instead of $L_1 [ab] L_2$ to mean the list that is the concatenation of L_1 , $[ab]$, and L_2 , in that order. Similarly, for a sequence of strings $s_1, s_2, \dots, s_m$, let $s_1 s_2 \dots s_m$ be the string that is their concatenation.

Suppose a DM with preference $>$, facing a list xyL , generates a binary choice path represented by the string $0s$ (formally, $>(xyL) = 0s$). This reveals that x is preferred to y under $>$. We claim that if instead the list is yxL , then the same preference $>$ will generate the string $1s$. This is because while the first entry of the string changes due to swapping the first two alternatives, the provisional choice after the first two alternatives is the same in either list xyL or yxL . As the rest of the two lists are identical, the remaining choice path is also the same.

Since the above claim holds for any preference, we must have $\rho(0s, xyL) = \rho(1s, yxL)$ as probabilities. This is the first behavioral implication of a rationalizable ρ :

First-Pair Consistency. For any list $xyL \in \mathcal{L}_N$ and any binary string s of length $N - 2$, it holds that

$$\rho(0s, xyL) = \rho(1s, yxL).$$

Intuitively, first-pair consistency says that the ordering of the first two alternatives on a list has no effect on what we can learn from the choice data generated by this list. However, swapping other pairs of adjacent alternatives in the list does have choice implications, which we investigate next.

Suppose a DM with preference $\succ$, facing a list $LxyL'$, generates the string $s10s'$, where L is non-empty, and the strings s and s' have length $|L| - 1$ and $|L'|$, respectively. This implies that when reaching the alternative x in the list $LxyL'$, x is preferred to the provisional choice at the end of L , which we denote by z for future reference. The preference $x \succ z$ corresponds to the 1 in the string $s10s'$. Moreover, the next binary choice reveals $x \succ y$, corresponding to the 0 in $s10s'$.

Given these properties, what string will the same preference $\succ$ generate when facing the list $LyxL'$, with x and y swapped? Since the first part of the list, L , is the same, it must generate the same first $|L| - 1$ entries in the string, namely s . The next binary choice in this list is new: the DM must now compare the provisional choice at the end of L , which is z , to the alternative y . There are two possibilities, either $z \succ y$, or $y \succ z$. If $z \succ y$, then z remains the provisional choice in Ly before it gets replaced by x in Lyx (recall $x \succ z$). Since the rest of the list is still L' , the list $LyxL'$ results in the string $s01s'$ in this case. If instead $y \succ z$, then y becomes the provisional choice in Ly before it gets replaced by x as well. In this case the resulting binary string will be $s11s'$.

To summarize, if a preference $\succ$ generates the string $s10s'$ in the list $LxyL'$, then it will generate either $s01s'$ or $s11s'$ in the list $LyxL'$ after swapping x and y . In fact, a similar analysis shows that the converse also holds, so that if $s01s'$ or $s11s'$ is generated in $LyxL'$, then $s10s'$ must be generated in $LxyL'$. As this holds for any preference $\succ$, we obtain the second behavioral implication of rationalizability:

Adjacent-Pair Consistency. For any list $LxyL' \in \mathcal{L}_N$ where L is non-empty, and for any binary strings s and s' of length $|L| - 1$ and $|L'|$, respectively, it holds that

$$\rho(s10s', LxyL') = \rho(s01s', LyxL') + \rho(s11s', LyxL').$$

First-pair consistency and adjacent-pair consistency together describe the choice implications of swapping an adjacent pair of alternatives in a list. We have argued that they are necessary conditions for the choice data ρ to be rationalizable. Remarkably, the following theorem shows that they are also sufficient.

Theorem 1. *Binary choice path data ρ is rationalizable as the aggregate choice behavior of a group of rational agents if and only if ρ satisfies first-pair consistency and adjacent-pair consistency.*

Proof. We have already established the only if direction. For the if direction, suppose ρ satisfies first-pair consistency and adjacent-pair consistency. Consider the following measure π over preferences: for any list $x_1x_2 \cdots x_n$, let

$$\pi(x_N > \cdots > x_2 > x_1) = \rho(1^{N-1}, x_1x_2 \cdots x_N),$$

where we use 1^{N-1} to denote the binary string with all 1s.

While we will show this later, at this point we are not guaranteed that π is a probability measure. Nonetheless, we can define the aggregate choice ρ_π in the same way, according to $\rho_\pi(s, L) = \sum_{>: s=>(L)} \pi(>)$. Our definition of π above guarantees that

$$\rho(1^{N-1}, x_1x_2 \cdots x_N) = \rho_\pi(1^{N-1}, x_1x_2 \cdots x_N), \quad \forall x_1x_2 \cdots x_N \in \mathcal{L}_N.$$

In what follows, we will use this property, together with the consistency axioms, to show that $\rho = \rho_\pi$ more generally.

To do this, we define $S \subseteq \mathbf{S}$ to be the set of all binary strings s such that $\rho(s, L) = \rho_\pi(s, L)$ for every $L \in \mathcal{L}_N$. Thus, $1^{N-1} \in S$. Moreover, because both ρ and ρ_π satisfy first-pair consistency, it is easy to see that if $1s \in S$, then $0s \in S$ as well. Similarly, adjacent-pair consistency of both ρ and ρ_π tells us that if $s01s', s11s' \in S$, then $s10s' \in S$ as well. The following technical lemma establishes that under these properties S must be the set of all strings.

Lemma 3.1. *Let $S \subseteq \mathbf{S}$ be a set of binary strings with length $N - 1$, such that:*

1. $1^{N-1} \in S$

2. $0s \in S$ whenever $1s \in S$
3. $s10s' \in S$ whenever $s01s', s11s' \in S$

Then $S = \mathbf{S}$.

This lemma is proved by inducting on N , with the base case $N = 2$ being trivial. Suppose the result holds for $N - 1$, then in the case of N , we can restrict attention to those strings that end with 1 and use the induction hypothesis to show that all these strings belong to S . The induction hypothesis applies because for those strings in S that end with 1, after removing the final digit of 1, they become a set of strings with length $N - 2$ that satisfy the three properties.

Thus, all strings of the form $s01$ and $s11$ belong to S , which by the third property of S implies that $s10$ also belongs to S . That is, if a string's last digit of 1 is in the second-to-last position, then it belongs to S . As a result, all strings of the form $s010$ and $s110$ belong to S , which again by the third property implies that $s100$ also belongs to S . Hence, if a string's last digit of 1 is in the third-to-last position, then it also belongs to S . Iterating this argument, we conclude that if a string has any digit of 1, it must belong to S . Finally, as the string $100 \cdots 0 \in S$, we know from the second property that $0^{N-1} \in S$. This shows every string is in S .

From the lemma, we know that $\rho = \rho_\pi$. It only remains to show that π is a probability measure over the set of preferences. For this we choose any list $L_0 \in \mathcal{L}_N$ and consider the sum $\sum_{s \in \mathbf{S}} \rho(s, L_0)$. On one hand, this sum is equal to 1 because ρ is the observed choice data. On the other hand, this sum is also equal to $\sum_{s \in \mathbf{S}} \rho_\pi(s, L_0)$, which evaluates to the total measure of π . Hence π has total measure one, completing the proof. $\square$

This theorem delivers a striking characterization of the random utility hypothesis in our novel environment. It reduces what is traditionally a global verification problem to just two local properties. This stands in sharp contrast to the characterization of RUM under standard data, which relies on the global structure encoded by the Block-Marschak (BM) polynomials. The theorem thus reveals that, in our environment, random utility admits a remarkably simple and transparent local foundation.

3.1 A Stronger Version of Theorem 1

It turns out that the first-pair consistency and adjacent-pair consistency conditions in Theorem 1 are not jointly independent, when we require them to hold for all possible lists and strings. One type of redundancy pertains to first-pair consistency alone: if we know $\rho(0s, xyL) = \rho(1s, yxL)$ and $\rho(1s, xyL) = \rho(0s, yxL)$, then summing across all possible strings s yields $1 = 1$ given that the observed choice probabilities $\{\rho(\cdot, xyL)\}$ and $\{\rho(\cdot, yxL)\}$ both sum to 1. In other words, under the assumption that $\rho(\cdot, xyL)$ and $\rho(\cdot, yxL)$ are both probability distributions, any single one of the above first-pair consistency conditions is implied by the rest.

Another type of redundancy relates to the interaction between first-pair consistency and adjacent-pair consistency. In its simplest form, recall that for any L and L' with non-empty L , adjacent-pair consistency requires the following to hold for any binary strings s and s' of length $|L| - 1$ and $|L'|$:

$$\rho(s10s', LxyL') = \rho(s01s', LyxL') + \rho(s11s', LyxL').$$

Suppose $\hat{L}$ is the list obtained by swapping the first two alternatives in L , while $\hat{s}$ is the string that differs from s only in the first digit. By first-pair consistency, we can rewrite the above displayed equation as

$$\rho(\hat{s}10s', \hat{L}xyL') = \rho(\hat{s}01s', \hat{L}yxL') + \rho(\hat{s}11s', \hat{L}yxL').$$

This leads to another, redundant adjacent-pair consistency condition.

In light of these redundancies, we seek to sharpen the conclusion of Theorem 1 by identifying a subset of the consistency conditions that are independent from one another, but remain sufficient for the given binary choice path data to be rationalizable. To do this, we will work with a general non-negative function $\rho(s, L)$, and state as an axiom the requirement that they represent choice probabilities. Carefully done, we can get around the first type of redundancy discussed above.

To remove the second type of redundancy, we introduce the following weaker version of adjacent-pair consistency:

Weak Adjacent-Pair Consistency. For any list $LxyL' \in \mathcal{L}_N$ where L is non-empty, and for any binary string s' of length $|L'|$, it holds that

$$\rho(1^{|L|-1}10s', LxyL') = \rho(1^{|L|-1}01s', LyxL') + \rho(1^{|L|-1}11s', LyxL').$$

In words, weak adjacent-pair consistency only imposes the original adjacent-pair consistency condition on strings that start with all 1s, up to the position of the two swapped alternatives. Economically speaking, it is only concerned with the effect of swapping two adjacent alternatives x and y in a list, on those choice probabilities that come from a certain subset of preferences. All these preferences agree with the order in the list up to x , ranking later alternatives (that precede x) higher than earlier ones, thus leading to $1^{|L|-1}$ as the common start to their binary strings.

The following result shows that weak adjacent-pair consistency together with first-pair consistency still characterizes rationalizability. Moreover, this weakening removes all the redundancies from the original combination of axioms. As a result, these conditions are the minimum amount that one needs to check in order to verify whether a random binary choice path dataset is rationalizable by our model.

Theorem 2. Given a non-negative function $\rho(s, L)$ defined for all $s \in \mathbf{S}$ and all $L \in \mathcal{L}_N$. Then ρ coincides with ρ_π for some preference distribution π , if and only if it satisfies first-pair consistency, weak adjacent-pair consistency, and $\sum_s \rho(s, L_0) = 1$ for some $L_0 \in \mathcal{L}_N$. In addition, given any $L_0 \in \mathcal{L}_N$, these axioms are jointly independent.

The proof is similar to, but more involved than the one we gave for Theorem 1, so we leave it to the appendix.

4 Identification with Complete and Partial Data

We turn to the identification problem. In particular, for random binary choice path data that is rationalizable, we wish to identify distributions over preference profiles that can generate such data through our model.

It is well known that with standard stochastic choice data, the random utility model cannot be point-identified once there are more than 3 alternatives. By contrast, we now show that rationalizable random binary choice path data permits point-identification, so

long as the data is *complete* – that is, when binary choice path data from *every* list in $\mathcal{L}_N$ is observed. Later in this section we will extend the analysis to *partial* binary choice path data, when only *some* lists are available.

Consider the following 4-alternative example:

		Preferences																							
Lists		abcd	abdc	acbd	acdb	adbc	adcb	bacd	badc	bcad	bcda	bdac	bdca	cabd	cadb	cbad	cbda	cdab	cdba	dabc	dacb	dbac	dbca	dcab	dcba
	[abcd]	000	000	000	000	000	000	100	100	100	100	100	100	010	010	110	110	010	110	001	001	101	101	011	111
	[acdb]	000	000	000	000	000	000	001	001	101	101	011	111	100	100	100	100	100	100	010	010	010	110	110	110
	[acdb]	000	000	000	000	000	000	001	001	101	101	011	111	100	100	100	100	100	100	010	010	010	110	110	110

Table 3: Preferences and Strings for Two Lists

Table 3 shows the strings generated by every possible preference profile for the two lists $[abcd]$ and $[acdb]$. Table 4 is an example of binary choice path data from these lists.

		Strings							
Lists		000	001	010	011	100	101	110	111
[abcd]		0.270	0.100	0.120	0.030	0.075	0.090	0.180	0.135
[acdb]		0.270	0.030	0.140	0.010	0.300	0.015	0.215	0.020

Table 4: Example of Random Binary Choice Path Data

Notice in Table 3 the list $[abcd]$ generates the string 111 for a unique preference profile, $d > c > b > a$. This is true more generally, in that any list L generates the string 1^{N-1} for a unique preference profile that ranks every alternative higher than its predecessors in L . Choice data from L is therefore sufficient to pin down the probability of this preference profile. Table 4, for example, identifies the probability of the profile $d > c > b > a$ as 0.135, and that of $b > d > c > a$ as 0.02. The following result is then immediate:

Observation 1. *Complete random binary choice path data that is rationalizable point-identifies the preference distribution.*

This positive identification result, however, has an onerous data requirement. Complete data requires all of the $N!$ lists. We can halve the requirement by taking advantage

of first-pair consistency – in the example, observed choice data from the list $[abcd]$ pins down unobserved data from $[bacd]$, so data from the latter list is repetitive. Even then, the number of lists required grows at the same rate as the number of preference profiles with N . Can we do better?

Return to Table 3. As discussed, data from the list $[abcd]$ pins down the probability of the preference profile $d > c > b > a$. It also pins down the probability of the profile $d > c > a > b$, as the unique preference that generates the string 011 in the list $[abcd]$ (equivalently, this preference uniquely generates 111 in the list $[bacd]$, whose data is given by first-pair consistency). Similarly, the list $[acdb]$ by itself pins down the probability of the profiles $b > d > c > a$ and $b > d > a > c$.

But we can say more if we combine the data from the two lists $[abcd]$ and $[acdb]$. Notice that for the list $[acdb]$, the string 110 is generated by three preference profiles, $d > c > b > a$, $d > c > a > b$ and $d > b > c > a$. Since we already know the probabilities of the first two, we can subtract them from the probability of 110 under $[acdb]$ to back out the probability of the third. This indirect method of identification suggests a way to reduce the number of lists needed to identify the preference distribution. Since each preference profile that is identified could be used to carry out further rounds of indirect identification with data from other lists, there is hope for a small set of lists to provide sufficient data that jointly identify the entire preference distribution.

Formally, let $\mathcal{L} \subseteq \mathcal{L}_N$ be a non-empty set of lists from which the analyst observes random binary choice path data. Thus, the (partial) data takes the form of distributions $\rho(\cdot, L)$ over binary strings $\mathbf{S}$, for every $L \in \mathcal{L}$. Say a set $\mathcal{L}$ of lists is **identification-sufficient**, if for any two preference distributions π and π' , the equality $\rho_\pi(\cdot, L) = \rho_{\pi'}(\cdot, L)$ holds for every list $L \in \mathcal{L}$ if and only if $\pi = \pi'$.

In the rest of this section, we will provide a detailed analysis of the properties of identification-sufficient sets. The analysis will culminate in the construction of a cardinality-minimal identification-sufficient set, which uses a vanishing share of all feasible lists as N grows large. Therefore, the data requirement for point-identification in our setting can be much less than what Observation 1 suggests.

4.1 Orbital Selection as a Necessary Condition

We begin by partitioning $\mathcal{L}_N$ in a specific way, which will help us understand identification-sufficient sets.

Orbital. *An orbital is a collection of lists that specifies which unordered pair of alternatives occupies position-pairs $L(i)$ and $L(i + 1)$ for each odd $i < N$.*

For example, with $N = 6$, one orbital assigns $\{a, b\}$ to positions $L(1), L(2)$, $\{c, d\}$ to $L(3), L(4)$ and $\{e, f\}$ to $L(5), L(6)$. This orbital therefore consists of 8 lists, $[abcdef]$, $[abdcfe]$, $[abdcef]$, $[abdcfe]$, $[bacdef]$, $[bacdfe]$, $[badcef]$ and $[badcfe]$. There are 90 such distinct orbitals for $N = 6$.

Orbital Selection. *An orbital selection is a set of lists, one from each orbital.*

The following result provides a general necessary condition for identification. In particular, it implies a lower bound on the cardinality of any identification-sufficient set of lists. The number of lists needed to uniquely identify the preference distribution must be at least as large as the number of distinct orbitals, which is $N!/2^{\lfloor N/2 \rfloor}$.

Theorem 3. *An identification-sufficient set of lists must contain an orbital selection.*

Proof. We present the proof for $N = 6$ and leave its generalization to the appendix. When $N = 6$, consider the 8 lists $[abcdef]$, $[abdcfe]$, $[abdcef]$, $[abdcfe]$, $[bacdef]$, $[bacdfe]$, $[badcef]$, $[badcfe]$ in a generic orbital. To show that an identification-sufficient set of lists must contain one of these 8 lists, define the following four preferences:

$$P_1 : e > f > c > d > a > b$$

$$P_2 : e > f > c > d > b > a$$

$$P_3 : e > f > d > c > a > b$$

$$P_4 : e > f > d > c > b > a$$

Note that in any list that does not begin with ab or ba , the preferences P_1 and P_2 lead to the same sequence of provisional choices. This is because P_1 and P_2 only differ in

their ranking of the two worst alternatives a and b , which is only relevant in lists where these alternatives are the first two. Similarly, in any list that does not begin with ab or ba , the preferences P_3 and P_4 also result in the same binary choice path data.

Similarly, consider a list that begins with ab or ba but does not continue with cd or dc . Since the preferences P_1 and P_3 only differ in the ranking between c and d , which are worst after a and b , they must lead to the same sequence of provisional choices in this list (note that in such a list, c and d are never compared to one another). The preferences P_2 and P_4 also lead to the same choice data in such a list.

Combining the two observations above, we know that for any list outside of the 8 in the given orbital, either P_1 and P_2 as well as P_3 and P_4 lead to the same choices, or P_1 and P_3 as well as P_2 and P_4 lead to the same choices. Therefore, if we start with any full support distribution of preferences π , increase the probabilities of P_1 and P_4 by a small ϵ and decrease the probabilities of P_2 and P_3 by ϵ , this change will not have any impact on the choice data in any list outside of the orbital.⁷ As a result, a set of lists disjoint from the orbital is not sufficient to identify the underlying π , which proves the result. See the appendix for a generalization of this argument to any N . $\square$

4.2 Spanning Set as a Sufficient Condition

Toward a sufficient condition for identification, we recall Observation 1, which tells us that access to data from all feasible lists solves the identification problem. Suppose we could use the choice data from one or more lists to fully infer the data from another list. Then a set of lists is identification-sufficient if all lists in $\mathcal{L}_N$ can be inferred in this way. In fact, we have already mentioned one way of inferring data of one list from another, using first-pair consistency:

Lemma 4.1. *Given random binary choice path data from xyL , rationalizable under some preference distribution, we can infer the corresponding data from yxL under the same preference distribution.*

The proof is straightforward, as first-pair consistency gives $\rho(1s, yxL) = \rho(0s, xyL)$ and $\rho(0s, yxL) = \rho(1s, xyL)$ whenever ρ is rationalizable. This lemma naturally defines

⁷This construction adapts the classic non-identification argument in Fishburn [1998] to our setting.

an operation on the space of lists, wherein data from one list can be used to “generate” that from another. We call this operation **switch**, such that

$$\text{switch}(xyL) = yxL.$$

It turns out that another simple operation can be used to make inference:

Lemma 4.2. *Given random binary choice path data from $LyxzL'$ and $LxzyL'$ with non-empty L , rationalizable under some preference distribution, we can infer the corresponding data from $LxyzL'$ under the same preference distribution.*

Proof. Similar to the previous lemma, the main idea of the proof is to show that if ρ is rationalizable, then for any binary string s_1 , the probability $\rho(s_1, LxyzL')$ can be written as a certain linear combination of the known probabilities $\rho(\cdot, LyxzL')$ and $\rho(\cdot, LxzyL')$.

To do this, we write s_1 as the concatenation of s with length $|L| - 1$, s^* with length 3, and s' with length $|L'|$, respectively. Thus, in the three relevant lists $LxyzL'$, $LyxzL'$ and $LxzyL'$, the string s captures the provisional choices in L , the string s^* describes how x, y, z compare to the best alternative in L and to each other, and s' represents how the alternatives in L' compare to the best alternative before them.

We now consider several different cases for the 3-digit string s^* . If $s^* = 000$ or 001 , we have the following identities:

$$\rho(s000s', LxyzL') = \rho(s000s', LyxzL')$$

$$\rho(s001s', LxyzL') = \rho(s001s', LyxzL')$$

These hold because the first two digits of s^* being 0 corresponds to preferences in which x and y are worse than the best alternative in L . This is an event that is not affected by swapping the positions of x and y .

If instead $s^* = 100$ or 101 , then by adjacent-pair consistency applied to the xy pair, we have

$$\rho(s100s', LxyzL') = \rho(s010s', LyxzL') + \rho(s110s', LyxzL')$$

$$\rho(s101s', LxyzL') = \rho(s011s', LyxzL') + \rho(s111s', LyxzL')$$

So the probabilities on the left-hand-side are determined. Similarly, if $s^* = 010$ or 110 , then adjacent-pair consistency applied to the yz pair gives

$$\rho(s010s', LxyzL') = \rho(s001s', LxzyL') + \rho(s011s', LxzyL')$$

$$\rho(s110s', LxyzL') = \rho(s101s', LxzyL') + \rho(s111s', LxzyL').$$

The remaining possibilities are $s^* = 011$ or 111 . For these cases, we note that if we start with the list $LxzyL'$, then adjacent-pair consistency applied to the zy pair gives

$$\rho(s010s', LxzyL') = \rho(s001s', LxzyL') + \rho(s011s', LxzyL')$$

$$\rho(s110s', LxzyL') = \rho(s101s', LxzyL') + \rho(s111s', LxzyL')$$

In these two identities, the left-hand side probabilities are known, and so are the first terms on the right-hand side (by the preceding analysis). It follows that $\rho(s011s', LxzyL')$ and $\rho(s111s', LxzyL')$ are also uniquely determined, completing the proof. $\square$

Lemma 4.2 defines another operation that takes specific pairs of lists to “generate” a third. We call this operation **merge**, such that for any L and L' with non-empty L ,⁸

$$\text{merge}(LyxzL', LxzyL') = LxyzL'.$$

For a set of lists, $\mathcal{L}$, we define the **span** of $\mathcal{L}$, denoted $sp(\mathcal{L})$, to be the smallest set of lists containing $\mathcal{L}$ that is closed under *switch* and *merge*. Thus, rationalizable random binary choice path data from $\mathcal{L}$ in fact determines the data from all of $sp(\mathcal{L})$. Observation 1 then implies the following sufficient condition for point-identification of the preference distribution:

Theorem 4. *If $sp(\mathcal{L}) = \mathcal{L}_N$, then $\mathcal{L}$ is an identification-sufficient set of lists.*

Deriving the span of a set of lists is computationally simple and transparent, so Theorem 4 provides a useful method to show that a given set of lists achieves point-identification. Call a set $\mathcal{L}$ of lists that spans $\mathcal{L}_N$ a **spanning set**. If in addition $\mathcal{L}$ has the smallest cardinality among all spanning sets, then call it a **basis**. By Theorem 3 and 4,

⁸In case L is empty, $xyzL'$ is already generated by $yxzL'$ by the previous switch operation.

a basis must contain an orbital selection.⁹ Thus, the number of lists in a basis is at least $N!/2^{\lfloor N/2 \rfloor}$. In what follows, we will show that this lower bound can be attained.

4.3 Identification-Sufficient Sets for Small N

Notice that in the above lower bound for the number of lists in a basis, the lower bound for an odd number $N + 1$ is exactly $N + 1$ times the lower bound for the even number N . The following simple lemma reduces the construction of bases that match the lower bound to the even N case.

Lemma 4.3. *If a spanning set for N alternatives consists of k lists, then there exists a spanning set for $N + 1$ alternatives with $(N + 1)k$ lists.*

Proof. Given the $N + 1$ alternatives, for any subset of size N , we can find k lists that span all possible lists in these alternatives. These lists can be extended to lists over all $N + 1$ alternatives, by simply appending the remaining alternative to the end of the lists. The k extended lists end with the same alternative, and span every list that ends with that alternative because switch and merge can be applied to the first N positions without affecting the last position. The union of such k lists across different subsets of size N then produces a desired spanning set for $N + 1$ alternatives. $\square$

When $N = 2$, each singleton list is trivially a basis. When $N = 3$, Lemma 4.3 tells us that any 3 lists that end with the three distinct alternatives form a basis. The converse is also true, since for $N = 3$ each orbital simply corresponds to those lists that have the same final alternative.

When $N = 4$, the bases are fully characterized in the following result, with proof in the appendix:

Proposition 1. *With 4 alternatives, a set of lists is a basis if and only if it is an orbital selection, necessarily having 6 lists, such that each of the 4 alternatives appears as the last alternative in one of the 6 lists.*

⁹This statement can be proved more simply by noting that from any two lists outside of a given orbital, neither switch nor merge would generate a list in the orbital.

Example 1. For $N = 4$ with alternatives $\{a, b, c, d\}$, consider the following orbital selection

$$\mathcal{L} = \{abdc, acbd, dacb, bcda, dbac, cdab\}.$$

The six lists lie in the six distinct orbitals (one list per orbital), and each alternative appears last in at least one list. By Proposition 1, $\mathcal{L}$ is a basis. We illustrate this directly.

We first apply the switch operation to each list to yield

$$\begin{aligned} \text{switch}(abdc) &= badc, & \text{switch}(acbd) &= cabd, & \text{switch}(dacb) &= adcb, \\ \text{switch}(bcda) &= cbda, & \text{switch}(dbac) &= bdac, & \text{switch}(cdab) &= dcab, \end{aligned}$$

bringing the span to 12 lists. Applying the merge operation gives 8 new lists:

$$\begin{aligned} \text{merge}(adcb, acbd) &= acdb, & \text{merge}(acbd, abdc) &= abcd, \\ \text{merge}(cabd, cbda) &= cbad, & \text{merge}(dbac, dacb) &= dabc, \\ \text{merge}(bcda, bdac) &= bdca, & \text{merge}(cdab, cabd) &= cadb, \\ \text{merge}(cbda, cdab) &= cdba, & \text{merge}(abdc, adcb) &= adbc, \end{aligned}$$

A final round of switch on the newly derived lists,

$$\text{switch}(cdba) = dcba, \quad \text{switch}(cbad) = bcad, \quad \text{switch}(bdca) = dbca, \quad \text{switch}(abcd) = bacd,$$

brings the span to all 24 lists, so $\text{sp}(\mathcal{L}) = \mathcal{L}_4$. □

This example illustrates the usefulness of the switch and merge operations. Instead of collecting data on all 24 lists for $N = 4$, a carefully chosen set of just 6 lists is enough to recover everything, by Theorem 4.

By Lemma 4.3, we can further construct a class of basis sets for $N = 5$ alternatives. However, for $N = 6$ and higher the combinatorial nature of the problem of finding a basis becomes remarkably complex. While we do not have a full characterization of all bases, we will present a simple construction that delivers a basis for any N .

4.4 A Minimal Identification-Sufficient Set for Any N

Let $0 \notin X$ be an auxiliary alternative that will simplify the exposition. We arrange the elements of $X \cup \{0\}$ on a circle with a fixed clockwise order. For distinct $a, b, c \in X \cup \{0\}$,

we say that the property $C(a, b, c)$ holds if, going clockwise from a , one encounters b before c . For $u \neq v$, let $d(u, v) \in \{1, \dots, N\}$ be the clockwise distance from u to v , and $\text{arc}(u, v)$ the set of elements strictly between u and v going clockwise. So $C(a, b, c)$ holds if and only if $b \in \text{arc}(a, c)$, which is equivalent to $c \in \text{arc}(b, a)$ or $a \in \text{arc}(c, b)$. For any $Z \subseteq X$, let $\mathcal{L}(Z)$ denote the set of all lists consisting exactly of all alternatives in Z .

For every subset Y of X , consider the following set of lists over the alternatives in Y :

$$\mathcal{C}(Y) = \{L \in \mathcal{L}(Y) \mid C(L(2k), L(2k+1), L(2k+2)) \text{ for } k = 0, \dots, \lfloor |Y|/2 \rfloor - 1\},$$

with $L(0)$ defined to be the auxiliary alternative 0. We will show that the set $\mathcal{C}(X)$ is a spanning set with cardinality that matches the aforementioned lower bound $N!/2^{\lfloor N/2 \rfloor}$.¹⁰

Figure 1 illustrates an instance of this construction for $N = 4$. Note that the specific set of lists this construction delivers depends on the specific clockwise ordering of the alternatives in $X \cup \{0\}$, but we will suppress this dependency for notational ease.

To analyze the properties of $\mathcal{C}(X)$, we first show that it is an orbital selection:

Lemma 4.4. *The set $\mathcal{C}(X)$ is an orbital selection that contains $N!/2^{\lfloor N/2 \rfloor}$ lists.*

Proof. Fix an orbital. It consists of unordered pairs $\{p_{2k+1}, p_{2k+2}\}$ for $k = 0, \dots, \lfloor N/2 \rfloor - 1$. Choosing a list from this orbital amounts to specifying an order for each pair. It turns out that our construction of $\mathcal{C}(X)$ pins down this order for each pair, one by one. Starting with $\{p_1, p_2\}$, we note that either $0, p_1, p_2$ are in clockwise order, or $0, p_2, p_1$ are in clockwise order (but not both). Thus, the property $C(L(0), L(1), L(2))$ uniquely determines which of p_1 and p_2 should occupy the $L(1)$ position, and which should be $L(2)$. But, once $L(2)$ is determined, the property $C(L(2), L(3), L(4))$ similarly determines which of p_3 and p_4 should be $L(3)$ and which should be $L(4)$. Iterating this argument, we can conclude that exactly one list in each orbital belongs to $\mathcal{C}(X)$. $\square$

We next show that $\mathcal{C}(X)$ is in fact a smallest spanning set/basis:

Theorem 5. *For any N and any ordering of the alternatives, the set $\mathcal{C}(X)$ spans all lists. Thus it is an identification-sufficient set with minimal cardinality.*

¹⁰We thank Zilin Jiang for suggesting this construction.

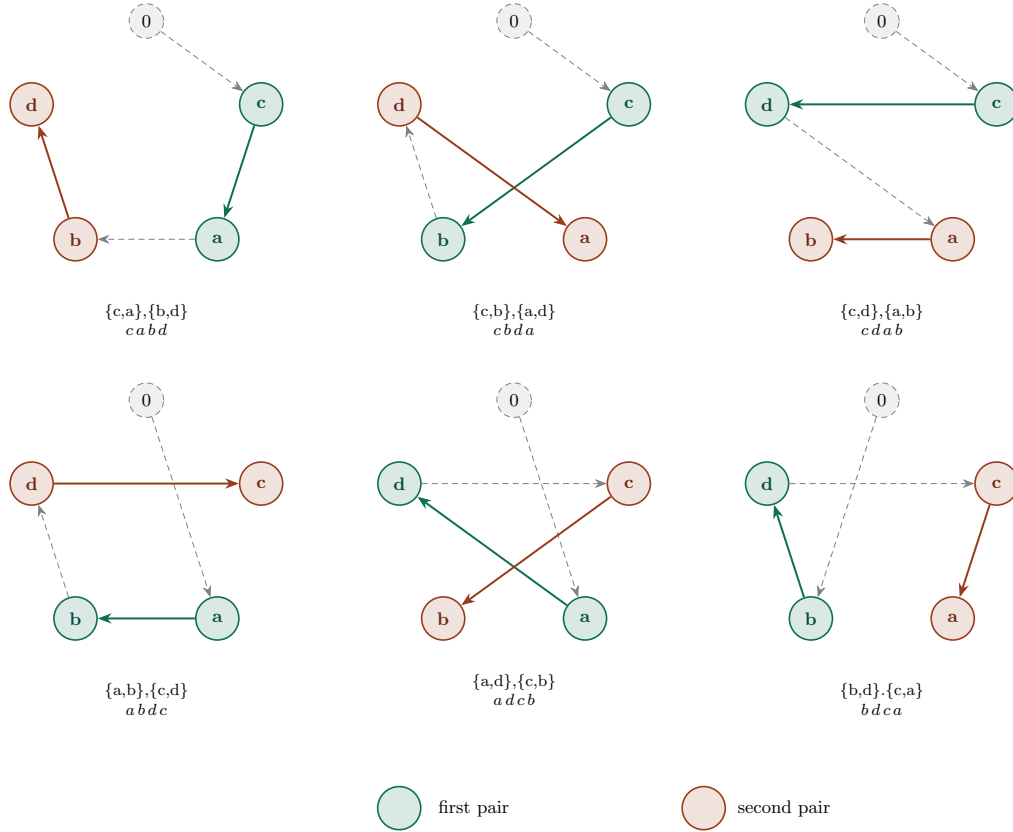


Figure 1: Construction for $N = 4$. Each panel illustrates the selection rule for a different orbital. Points $0, c, a, b, d$ sit at fixed positions on a circle, clockwise in that order, with 0 the auxiliary starting reference. Each panel splits $\{a, b, c, d\}$ into a first and a second unordered pair to define its orbital. The first pair are shown in green, and the second pair in orange. A solid colored arrow gives the order the construction picks for that pair. It points from the element reached first, going clockwise from the current reference, to that reached second. The dashed gray arrows carry the reference forward (from 0 to the start of the first pair, then from the end of the first to the start of the second). The list below each panel is the resulting permutation, read off by following the solid arrows in order.

Proof. We first observe that when N is an odd number, the definition of $\mathcal{C}(X)$ does not put any constraint on the last alternative in the list. Thus, the construction $\mathcal{C}(X)$ in the odd N case can be obtained from the construction in the even $N - 1$ case, in the same way as the proof of Lemma 4.3 describes. Because of this, to show $\mathcal{C}(X)$ is a spanning set it suffices to focus on even N .

We prove the result by induction on N . For the base case $N = 2$, suppose $X = \{x, y\}$ with $0, x, y$ arranged clockwise. Then $\mathcal{C}(X)$ consists of a single list $[xy]$, which spans the other possible list $[yx]$ by switch.

Let $N \geq 4$ be even, and assume the spanning result for $N - 2$. In the case of N alternatives, define T to be the span of $\mathcal{C}(X)$. For every ordered pair (u, v) in X with $u \neq v$, let $Y_{uv} = X \setminus \{u, v\}$ and further define

$$T_{uv} = \{L \in \mathcal{L}(Y_{uv}) \mid Luv \in T\}.$$

Observe that the set of lists T_{uv} is closed under switch and merge. This is because in the above definition, T is closed under these operations, and any operation on L directly translate into the same operation on Luv .

Our goal is to show that T_{uv} is the entire set of lists $\mathcal{L}(Y_{uv})$ on the $N - 2$ alternatives except u, v . If this is true for every pair (u, v) , then the result would follow that T is also the entire set of lists $\mathcal{L}(X)$ on all N alternatives. By the induction hypothesis, to prove $T_{uv} = \mathcal{L}(Y_{uv})$, it suffices (because T_{uv} is closed under switch and merge) to prove that it contains $\mathcal{C}(Y_{uv})$, where we maintain the clockwise ordering of the alternatives. This is stated in the following lemma:

Lemma 4.5. *For every (u, v) , we have $\mathcal{C}(Y_{uv}) \subseteq T_{uv}$ and thus $T_{uv} = \mathcal{L}(Y_{uv})$.*

The lemma is proved by induction on $d(u, v)$, the clockwise distance from u to v . Choose an arbitrary list $\hat{L} \in \mathcal{C}(Y_{uv})$, and write it in the form of Lx with x being the last alternative in $\hat{L}$. Since $x \neq u, v$, either $x \in \text{arc}(v, u)$ or $x \in \text{arc}(u, v)$. Below we investigate the two cases respectively.

If $x \in \text{arc}(v, u)$, then x, u, v are in clockwise order. Thus the list $\hat{L}uv = Lxuv$ satisfies the property $C(L(2k), L(2k+1), L(2k+2))$ for $k = N/2 - 1$. But since $\hat{L} \in \mathcal{C}(Y_{uv})$, it also satisfies the clockwise property for smaller k . As a result, $\hat{L}uv \in \mathcal{C}(X) \subseteq T$, proving that

$\hat{L} \in T_{uv}$. Note that when $d(u, v) = 1$, x must belong to $\text{arc}(v, u)$. So this analysis also resolves the base case of the (inner) induction.

Suppose instead that $x \in \text{arc}(u, v)$. Then by a similar argument, we have $\hat{L}vu = Lxvu \in \mathcal{C}(X) \subseteq T$. Moreover, x strictly between u and v clockwise gives $d(x, v) < d(u, v)$, and so by the induction hypothesis, $T_{xv} = \mathcal{L}(Y_{xv})$ which in particular contains the list Lu . That is to say, $Luxv \in T$. Now observe that

$$\text{merge}(Luxv, Lxvu) = Lxuv = \hat{L}uv.$$

So we are again able to deduce $\hat{L} \in T_{uv}$, which completes the proof of the lemma and the theorem. $\square$

We conclude this section on identification by noting that in our construction, the ratio of the number of lists in the basis to the number of all lists ($N!$), is $1/2^{\lfloor N/2 \rfloor}$. Therefore, as N grows large, we only need a vanishing share of all feasible lists to achieve point-identification of the underlying preference distribution.

5 Rationalizability with Partial Data

We return to the question of which observed data can be explained by our model, but now ask this question on partial data. Throughout the section, assume that the analyst observes the choice probabilities $\rho^{\mathcal{L}}(s, L)$ for all lists L in a non-empty subset $\mathcal{L} \subseteq \mathcal{L}_N$.

As illustrated by Table 2 and 3, each list L partitions the set of preferences according to the binary choice paths (strings) that they generate. Every string corresponds to a cell of the partition, such that each preference in the cell generates that string for the given list. Denote this partition of $\mathcal{P}$ induced by the list L as $\mathcal{B}^L$. That is,

$$s_1, s_2 \in B_i \in \mathcal{B}^L \quad \text{if and only if} \quad s_1(L) = s_2(L).$$

For a cell B in $\mathcal{B}^L$ induced by the list L , denote the corresponding string as $s^L(B)$.

Suppose the set of preferences that generate a string s for one list L is identical to those that generate a (possibly different) string s' for a different list L' . Then, in any rationalizable random binary choice path data, the probability of observing these strings given their corresponding lists must be the same. Similarly, if the set of preferences

that generate a string for one list is the complement of the set that generates a (possibly different) string for a different list, then the probabilities of observing these strings given their corresponding lists must add up to one. For any subset of preferences $B \subseteq \mathcal{P}$, let B^c denote its complement. The above discussion motivates the following definition:

Consistency. *Partial random binary choice path data $\rho^{\mathcal{L}}$ is consistent if for any $L, L' \in \mathcal{L}$,*

$$(i) : B \in \mathcal{B}^L \cap \mathcal{B}^{L'} \Rightarrow \rho^{\mathcal{L}}(\mathbf{s}^L(B), L) = \rho^{\mathcal{L}}(\mathbf{s}^{L'}(B), L'),$$

$$(ii) : B \in \mathcal{B}^L, B^c \in \mathcal{B}^{L'} \Rightarrow \rho^{\mathcal{L}}(\mathbf{s}^L(B), L) = 1 - \rho^{\mathcal{L}}(\mathbf{s}^{L'}(B^c), L').$$

Given consistent partial random binary choice path data, we can define the following cooperative game, which we call the *consistent data game*.

Consistent Data Game. *Given consistent partial random binary choice path data $\rho^{\mathcal{L}}$, let $v : 2^{\mathcal{P}} \rightarrow [0, 1]$ be as follows,*

$$\begin{aligned} v(B) &= \rho^{\mathcal{L}}(\mathbf{s}^L(B), L) && \text{if } B \in \mathcal{B}^L \text{ for some } L \in \mathcal{L} \\ &= 1 - \rho^{\mathcal{L}}(\mathbf{s}^L(B^c), L) && \text{if } B^c \in \mathcal{B}^L \text{ for some } L \in \mathcal{L} \\ &= 1 && \text{if } B = \mathcal{P}, \\ &= 0 && \text{otherwise.} \end{aligned}$$

The following result shows that the rationalizability of partial data can be reframed as the non-emptiness of the core of this cooperative game:

Proposition 2. *Partial random binary choice path data is rationalizable if and only if it is consistent and the corresponding consistent data game has a non-empty core.*

The proof is left to the appendix. Note, however, that the simple statement of this result hides the complexity of verifying the non-emptiness of the core. Indeed, a similar characterization is possible for standard data, but as noted by the previous literature, the difficulty lies in checking the Bondareva-Shapley balancedness conditions for the cooperative game.

Despite this general difficulty, there is a more straightforward characterization for a large class of partial data. Recall that data from any spanning set of lists point-identifies

the preference distribution, if such data is rationalizable. Below we provide a simple rationalizability test for exactly such data.

Recall Lemma 4.1 and 4.2, which described ways to generate implied data on the span of a set of lists that would respect rationalizability. Starting from any rationalizable partial data $\rho^{\mathcal{L}}$ where $\mathcal{L}$ is a spanning set, we can infer the complete data $\rho(\cdot, L)$, for every list L . But these inferences can be made even without knowing whether $\rho^{\mathcal{L}}$ is rationalizable. In this more general case, different paths for generating lists outside $\mathcal{L}$ may disagree on the implied probabilities, or the implied probabilities may not even be probabilities. Nevertheless, for any partial data $\rho^{\mathcal{L}}$, we can write down a candidate complete data function $\rho(\cdot, L)$ that corresponds to some path of generating the list L from those in $\mathcal{L}$ using switch and merge.

The complete data ρ then yields a candidate preference distribution π as in Observation 1, such that for each preference $\succ$, $\pi(\succ)$ is given by the ρ -probability of observing the string 1^{N-1} in the list that “aligns with” $\succ$ (any later alternative is preferred to earlier ones). Once again, a priori π is not guaranteed to be a probability distribution, or even non-negative. But it is the only possible preference distribution that explains the partial data $\rho^{\mathcal{L}}$, whose rationalizability thus reduces to checking whether ρ_{π} coincides with $\rho^{\mathcal{L}}$.

Theorem 6. *Suppose $\mathcal{L}$ spans all lists. Partial random binary choice path data $\rho^{\mathcal{L}}$ is rationalizable if and only if the candidate preference distribution π defined above is non-negative, and satisfies $\rho_{\pi}(s, L) = \rho^{\mathcal{L}}(s, L)$ for all $L \in \mathcal{L}$ and $s \in S$.*

Proof. The only if direction is straightforward given our previous discussions. The if direction is also easy, except that we need to show π is a probability distribution given the assumptions. This follows from the familiar argument that for any $L \in \mathcal{L}$, we have

$$\sum_{\succ \in \mathcal{P}} \pi(\succ) = \sum_{s \in S} \rho_{\pi}(s, L) = \sum_{s \in S} \rho^{\mathcal{L}}(s, L) = 1.$$

□

6 Discussion

We are not aware of a dataset in the literature with the exact form as ours. However, nearby specifications establish the feasibility of obtaining such data. Experiment II in Li

[2021] implements precisely our choice procedure, with a binary choice in each round, previous choice carried forward, exogenous order of alternatives, and entire choice path recorded. However, in that experiment each subject faced a different list and therefore aggregate data was unavailable. Geng [2016] also collects binary choice path data in the absence of preference heterogeneity, to test for status quo bias. The alternatives are monetary payoffs that can be objectively ranked, but are presented as mathematical expressions that require attention and reasoning to perceive. Differential attention and reasoning skills generate the choice path heterogeneity. Our results deliver an exact test for rationality with such data, in the more general setting that involves unobservable preference heterogeneity.

Similar data is also collected in industry, and in large volume. Choice-based conjoint (CBC) analysis is a widely used tool in market research for pricing and product design. An advanced variant of this tool, labeled Adaptive CBC, collects aggregate choice path data (see Johnson and Orme [2007] and Cunningham et al. [2010]). It deviates from our setting in a couple of ways. First, at an earlier step the DM effectively narrows down the set of alternatives they will eventually face. Second, at every step the DM selects from three alternatives (including the previous winner) instead of two. A major use of the tool is precisely to identify the underlying preference distribution, but compared to our general RUH, more structure is imposed to achieve identification.

A few other features of our novel data, beyond the scope of the present paper, deserve mention. First, choice settings that generate such data are fair, in that every agent gets to choose from the same set of alternatives. Second, the choice environment is simple. At any stage the DM chooses from a pair of alternatives, which is arguably easier than a choice from more alternatives. Further, the list is fixed, making it easy to describe, implement, and understand. Compare this to a choice setting that mimics a sorting algorithm, wherein subsequent choices depend on current ones in ways far more complex than merely “winner survives”. The complexity of a choice environment likely has a strong bearing on the integrity of the resulting data and is therefore an important metric for comparing different data sets.

Our results show that random binary choice path data has better testability and identification properties compared to standard random choice data. It is possible that there

exists other data that does even better on one or both of these counts. We are unaware of such work, but hope it follows.

References

Victor H Aguiar, Maria Jose Boccardi, and Mark Dean. Satisficing and stochastic choice. *Journal of Economic Theory*, 166:445–482, 2016. [7](#)

Jose Apesteguia and Miguel A Ballester. Choice by sequential procedures. *Games and Economic Behavior*, 77(1):90–99, 2013. [7](#)

Jose Apesteguia, Miguel A Ballester, and Jay Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, 2017. [6](#)

P.H. Au and K. Kawai. Sequentially rationalizable choice with transitive rationales. *Games and Economic Behaviour*, 73 (2):608–614, 2011. [7](#)

Salvador Barberá and Prasanta K Pattanaik. Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, pages 707–715, 1986. [5](#), [6](#)

Levon Barseghyan, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048, 2021. doi: 10.3982/ECTA17448. [5](#)

Antoine Billot and Jacques-François Thisse. Stochastic rationality and möbius inverse. *International Journal of Economic Theory*, 1(3):211–217, 2005. [5](#)

Henry David Block and Jacob Marschak. Random orderings and stochastic theories of response. 1959. [5](#), [6](#)

Andrew Caplin and Mark Dean. Search, choice, and revealed preference. *Theoretical Economics*, 6(1):19–48, 2011. [2](#), [7](#)

Andrew Caplin, Mark Dean, and Daniel Martin. Search and satisficing. *The American Economic Review*, 101(7):2899–2922, 2011. [2](#), [7](#)

- Peter P. Caradonna and Christopher Turansick. Identification in stochastic choice, 2026. URL <https://arxiv.org/abs/2602.19950>. 6
- Matias D. Cattaneo, Paul Cheung, Xinwei Ma, and Yusufcan Masatlioglu. Attention overload, 2024. URL <https://arxiv.org/abs/2110.10650>. 6
- Christopher P. Chambers and Christopher Turansick. The limits of identification in discrete choice, 2024. URL <https://arxiv.org/abs/2403.13773>. 6
- Paul HY Cheung and Yusufcan Masatlioglu. Frame-dependent random utility. *arXiv preprint arXiv:2502.00209*, 2025. 8
- Michael A Cohen. Random utility systems—the infinite case. *Journal of Mathematical Psychology*, 22(1):1–23, 1980. 5
- Charles E Cunningham, Ken Deal, and Yvonne Chen. Adaptive choice-based conjoint analysis: a new patient-centered approach to the assessment of health service preferences. *The Patient: Patient-Centered Outcomes Research*, 3(4):257–273, 2010. 31
- Dinko Dimitrov, Saptarshi Mukherjee, and Nozomu Muto. ‘divide-and-choose’ in list-based decision problems. *Theory and Decision*, 81(1):17–31, 2016. 7
- T. Dridi. Sur les distributions binaires associées à des distributions ordinales. *Mathématiques et Sciences humaines*, 69:15–31, 1980. 5
- T. Dridi. Distributions binaires unimodales. *Discrete Mathematics*, 126(1-3):373–378, 1994. 6
- Rohan Dutta. Gradual pairwise comparison and stochastic choice. *Theoretical Economics*, 15(4):1335–1364, 2020. doi: <https://doi.org/10.3982/TE3647>. URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/TE3647>. 7
- Rohan Dutta and Sean Horan. Inferring rationales from choice: Identification for rational shortlist methods. *mimeo*, 2013. 7
- Larry G. Epstein and Kaushil Patel. Identifying heterogeneous decision rules from choices when menus are unobserved. Working paper, McGill University, 2026. URL <https://arxiv.org/abs/2405.09500>. 5

- Jean-Claude Falmagne. A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1):52–72, 1978. [5](#)
- Emel Filiz-Ozbay and Yusufcan Masatlioglu. Progressive random choice. *Journal of Political Economy*, 131(3):716–750, 2023. [6](#)
- Peter C Fishburn. Induced binary probabilities and the linear ordering polytope: A status report. *Mathematical Social Sciences*, 23(1):67–80, 1992. [5](#)
- Peter C Fishburn. Stochastic utility. In Salvador Barbera, Peter Hammond, and Christian Seidl, editors, *Handbook of Utility Theory*, pages 273–318. Kluwer Dordrecht, 1998. [3](#), [6](#), [20](#)
- Sen Geng. Decision time, consideration time, and status quo bias. *Economic Inquiry*, 54(1):433–449, 2016. [8](#), [31](#)
- Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006. [6](#)
- Begum Guney. A theory of iterative choice in lists. *Journal of Mathematical Economics*, 53:26–32, 2014. [7](#)
- Sean Horan. Sequential search and choice from lists. *Working Paper*, 2010. [7](#)
- Sean Horan. A simple model of two-stage choice. *Journal of Economic Theory*, 162:372–406, 2016. [7](#)
- Nicolas Houy. Choice functions with states of mind. *Theory and Decision*, 65(1):1–26, 2008. [7](#)
- Yuhta Ishii, Matthew Kovach, and Levent Ülkü. A model of stochastic choice from lists. *Journal of Mathematical Economics*, 96:102509, 2021. [7](#)
- Richard M Johnson and Bryan K Orme. A new approach to adaptive cbc. In *Sawtooth software conference*, page 85, 2007. [31](#)
- Haruki Kono, Kota Saito, and Alec Sandroni. Random utility with unobservable alternatives. *American Economic Review*, 116(8):2850–2882, 2026. [5](#)

- Gleb Koshevoy and Ernesto Savaglio. On rational choice from lists of sets. *Journal of Mathematical Economics*, 109, 2023. [8](#)
- Matthew Kovach and Levent Ülkü. Satisficing with a variable threshold. *Journal of Mathematical Economics*, 87:67–76, 2020. [7](#)
- Yi Li. The abc mechanism: an incentive compatible payoff mechanism for elicitation of outcome and probability transformations: Y. li. *Experimental Economics*, 24(3):1019–1046, 2021. [30](#)
- Yi-Hsuan Lin. Random non-expected utility: Non-uniqueness. *arXiv preprint arXiv:2009.04173*, 2020. [6](#)
- Jay Lu and Kota Saito. Random intertemporal choice. *Journal of Economic Theory*, 177:780–815, 2018. [6](#)
- Michael Mandler, Paola Manzini, and Marco Mariotti. A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, 147(1):71–92, 2012. [7](#)
- P. Manzini and M. Mariotti. Sequentially rationalizable choice. *American Economic Review*, 97(5):1824–1839, 2007. [7](#)
- Paola Manzini and Marco Mariotti. Choice by lexicographic semiorders. *Theoretical Economics*, 7:1–23, 2012. [7](#)
- Paola Manzini and Marco Mariotti. Dual random utility maximisation. *Journal of Economic Theory*, 177:162–182, 2018. [6](#)
- Paola Manzini, Marco Mariotti, and Levent Ülkü. A model of approval with an application to list design. *Journal of Economic Theory*, 217:105821, 2024. [7](#)
- Yusufcan Masatlioglu and Daisuke Nakajima. Choice by iterative search. *Theoretical Economics*, 8(3):701–728, 2013. [7](#)
- Daniel McFadden and Marcel K Richter. Stochastic rationality and revealed stochastic preference. In *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, pages 161–186. Westview Press: Boulder, CO, 1990. [5](#)

- Mauro Papi. Satisficing choice procedures. *Journal of Economic Behavior & Organization*, 84(1):451–462, 2012. 8
- Henrik Petri. Binary single-crossing random utility models. *Games and Economic Behavior*, 138:311–320, 2023. 6
- Ariel Rubinstein. Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, 46(1):145–153, 1988. 6
- Ariel Rubinstein and Yuval Salant. A model of choice from lists. *Theoretical Economics*, 1(1):3–17, 2006. 7
- Herbert A Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, pages 99–118, 1955. 8
- Herbert A. Simon. *From substantive to procedural rationality*, pages 65–86. Springer US, Boston, MA, 1976. 7
- Elchin Suleymanov. Branching-independent random utility model. *Journal of Economic Theory*, 220:105880, 2024. ISSN 0022-0531. doi: <https://doi.org/10.1016/j.jet.2024.105880>. 6
- Gerelt Tserenjigmid. The order-dependent Luce model. *Management Science*, 67(11):6915–6933, 2021. 7
- Christopher Turansick. Identification in the random utility model. *Journal of Economic Theory*, 203:105489, 2022. 6
- A. Tversky. Intransitivity of preferences. *Psychological Review*, 76:31–48, 1969. 6
- Amos Tversky. Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281, 1972. 6
- Kremena Valkanova. Revealed preference domains from random choice. *Games and Economic Behavior*, 147:288–304, 2024. 6

Erya Yang and Igor Kopylov. Random quasi-linear utility. *Journal of Economic Theory*, 209:105650, 2023. ISSN 0022-0531. [6](#)

Jiaqi Yang. Choice overload via the primacy effect, 2025. [8](#)

Ece Yegane. Stochastic choice with limited memory. *Journal of Economic Theory*, 205:105548, 2022. [7](#)

Kemal Yildiz. List-rationalizable choice. *Theoretical Economics*, 11(2):587–599, 2016. [7](#)

7 Appendix

7.1 Proof of Theorem 2

Sufficiency: We already know that the axioms are necessary for rationalizability. To show they are sufficient, we follow the proof of Theorem 1 in the main text to define a measure π by

$$\pi(x_N > \cdots > x_2 > x_1) = \rho(1^{N-1}, x_1 x_2 \cdots x_N).$$

We again define $S \subseteq \mathbf{S}$ to be the set of all binary strings s such that $\rho(s, L) = \rho_\pi(s, L)$ for every $L \in \mathcal{L}_N$. Thus, $1^{N-1} \in S$. First-pair consistency of ρ and ρ_π tells us that if $1s \in S$, then $0s \in S$ as well. Weak adjacent-pair consistency ensures that if $1^k 01s', 1^k 11s' \in S$, then $1^k 10s' \in S$ as well. We then need the following stronger lemma to show that under these properties S must contain all strings.

Lemma 7.1. *Let $S \subseteq \mathbf{S}$ be a set of binary strings with length $N - 1$, such that:*

1. $1^{N-1} \in S$
2. $0s \in S$ whenever $1s \in S$
3. $1^k 10s' \in S$ whenever $1^k 01s', 1^k 11s' \in S$, for any integer k with $0 \leq k \leq N - 3$ and any string s' with length $N - 3 - k$.

Then $S = \mathbf{S}$.

The proof of this lemma is by strong induction, where we assume that the lemma holds for $2, 3, \dots, N - 1$. In the case of N , we can restrict attention to those strings that end with 1 and use the induction hypothesis for $N - 1$ to show that all these strings belong to S . The induction hypothesis applies because for those strings in S that end with 1, after removing the final digit of 1, they become a set of strings with length $N - 2$ that satisfy the three properties.

Thus, the strings $1^{N-3}01$ and $1^{N-3}11$ belong to S , which by the third property of S implies that $1^{N-3}10$ also belongs to S . We can then restrict attention to those strings that end with 10 and use the induction hypothesis for $N - 2$ to show that all these strings belong to S . That is, if a string's last digit of 1 is the second-to-last position, then it belongs to S . As a result, the strings $1^{N-4}010$ and $1^{N-4}110$ belong to S , which again by the third property implies that $1^{N-4}100$ also belongs to S . A similar argument then shows that all strings ending with 100 belong to S .

Hence, if a string's last digit of 1 is in the third-to-last position, then it also belongs to S . Iterating this argument, we conclude that if a string has any digit of 1, it must belong to S . Finally, as the string $100 \dots 0 \in S$, we know from the second property that $0^{N-1} \in S$. This shows every string is in S .

From the lemma, we know that $\rho = \rho_\pi$. The fact that π must be a probability measure follows from the same argument as in the main text, thanks to the axiom $\sum_s \rho(s, L_0) = 1$.

Independence of the axioms: Fix a list L_0 , and note that first-pair consistency, weak adjacent-pair consistency and the add-up conditions $\sum_s \rho(s, L_0) = 1$ are all linear constraints on the variables $\{\rho(s, L)\}$. We will show that these constraints are linearly independent by a dimension-counting argument.

The total number of variables involved is $|S| \times |\mathcal{L}_N|$, which is $2^{N-1} \cdot N!$. First-pair consistency equates pairs of these variables, so it entails $2^{N-2} \cdot N!$ constraints. Weak adjacent-pair consistency has bite for a string of the form $1^k 10s'$. For each k with $0 \leq k \leq N - 3$, there are 2^{N-3-k} possible choices for the string s' . So the number of strings of the form $1^k 10s'$ is $\sum_{0 \leq k \leq N-3} 2^{N-3-k} = 2^{N-2} - 1$. Multiplied by the number of lists, we obtain that the total number of weak adjacent-pair consistency constraints is $(2^{N-2} - 1) \cdot N!$. Finally we have the singleton constraint $\sum_s \rho(s, L_0) = 1$.

The three types of constraints above add up to a total number of $(2^{N-1} - 1) \cdot N! + 1$ constraints, in $2^{N-1} \cdot N!$ variables. Therefore, the affine dimension of those ρ that satisfy all the constraints is *at least*

$$2^{N-1} \cdot N! - ((2^{N-1} - 1) \cdot N! + 1) = N! - 1.$$

But the dimension is also *at most* $N! - 1$, because we already showed that such ρ coincides with some ρ_π , which as a linear mapping of π cannot exceed the latter's affine dimension $N! - 1$ (since π is a probability distribution over $N!$ preferences). Thus equality must hold, which in particular implies that the constraints we considered above are linearly independent.

7.2 Proof of Theorem 3 for General N

The argument in the main text can be generalized to any even N as follows. Suppose $N = 2n$, and take any orbital where x_1 and x_2 occupy the first two positions in the list, x_3 and x_4 occupy the next two positions, $\dots$, and x_{2n-1} and x_{2n} occupy the final two positions. Consider the following collection of 2^{n-1} preferences $P_{j_1, j_2, \dots, j_{n-1}}$, where each j_k in the subscript belongs to $\{0, 1\}$. In each of these preferences, x_{2n-1} and x_{2n} are the most and second-most preferred alternatives, followed by an ordering between x_{2n-3} and x_{2n-2} , followed by an ordering between x_{2n-5} and x_{2n-4} , so on and so forth, ending with x_1 and x_2 (or the other way around) as the two least preferred alternatives. For $1 \leq k \leq n - 1$, the ordering between x_{2k-1} and x_{2k} is determined by the subscript j_k , where $j_k = 0$ indicates $x_{2k-1} > x_{2k}$ and $j_k = 1$ indicates $x_{2k} > x_{2k-1}$.

Within these preferences, each pair $P_{0, j_2, \dots, j_{n-1}}$ and $P_{1, j_2, \dots, j_{n-1}}$ only differs from each other in the ranking between x_1 and x_2 , which are the two worst alternatives under these preferences. Thus, in any list that does not begin with $x_1 x_2$ or $x_2 x_1$, this pair of preferences leads to the same choice data. Likewise, each pair $P_{j_1, 0, \dots, j_{n-1}}$ and $P_{j_1, 1, \dots, j_{n-1}}$ only differs in the ranking between x_3 and x_4 , which are the two worst alternatives after x_1 and x_2 . In any list that begins with $x_1 x_2$ or $x_2 x_1$, but does not continue to $x_3 x_4$ or $x_4 x_3$, this pair of preferences lead to the same choice data because x_3 and x_4 are never compared to each other along the binary choice path.

More generally, for any $1 \leq k \leq n - 1$, suppose P and $\hat{P}$ belong to the above 2^{n-1} preferences and differ only in their k -th subscript j_k . Then they lead to the same choice data in any list that agrees with the orbital in the first $2k - 2$ positions but differs in the next two positions (that is, any list beginning with x_1x_2 or x_2x_1 , followed by x_3x_4 or $x_4x_3, \dots$, followed by $x_{2k-3}x_{2k-2}$ or $x_{2k-2}x_{2k-3}$, but not followed by $x_{2k-1}x_{2k}$ or $x_{2k}x_{2k-1}$). Since any list outside of the orbital has a unique k where it first differs from the orbital in the $2k - 1$ and $2k$ -th positions, this list partitions the 2^{n-1} preferences into pairs that lead to the same choice data in this list. In each pair, the two preferences differ from one another only in their k -th subscript j_k .

Now, starting with any full support distribution of preferences π , suppose we change the probability of each $P_{j_1, \dots, j_{n-1}}$ by $\epsilon \cdot (-1)^{\sum_k j_k}$, which is $\pm\epsilon$ depending on whether the subscripts j_k sum to an even number or an odd number. As discussed above, any list outside the orbital partitions the 2^{n-1} preferences into pairs that lead to the same choice data in this list. In each of these pairs, the probability of one preference is increased by ϵ while that of the other is decreased by ϵ . Hence, such a change to π has no impact on the choice data in any list outside the orbital. This proves that a set of lists is sufficient for identification only if it contains a list in any given orbital.

Essentially the same proof applies if $N = 2n + 1$ is odd. In that case we consider 2^n preferences that all put x_{2n+1} (the last alternative in any list in a given orbital) as the most preferred alternative, followed by an ordering between x_{2n-1} and x_{2n} and so on. The remaining steps of the proof are unchanged.

7.3 Proof of Proposition 1

Sufficiency: Suppose an orbital selection $\mathcal{L}$ satisfies the premise of the proposition. Since it is an orbital selection, for any $a, b \in X$, there must be some list in $\mathcal{L}$ with $\{a, b\}$ in the first two positions. Suppose it is $[abcd]$. Then by switch we get $[bacd]$. If we could also obtain $[abdc]$ in the span, and then $[badc]$ via switch again, then $\mathcal{L}$ spans this orbital. Since this orbital is arbitrary, we will be able to conclude that $\mathcal{L}$ spans all lists.

Now, if $[adbc]$ is spanned by $\mathcal{L}$, then we obtain $[abdc]$ via $\text{merge}([abcd], [adbc])$. So suppose for the sake of contradiction that $[abdc]$ and $[adbc]$ are not in the span, which means that $[badc]$ and $[dabc]$ are also not, because of switch.

On the other hand, there exists some list in $\mathcal{L}$ that ends with the alternative c . As we have previously ruled out $[abdc]$, $[adbc]$, $[badc]$ and $[dabc]$, it must be either $[bdac]$ or $[dbac]$. So $[bdac]$ is spanned by $\mathcal{L}$. Recall that $[bacd]$ is also spanned, and so $\text{merge}([bacd], [bdac]) = [badc]$ is spanned, leading to a contradiction.

Necessity: The necessity of an orbital selection follows from Theorem 3. Now suppose that $\mathcal{L}$ is a collection of lists with none that end with the alternative a . Then it is not possible to generate such a list from another in $\mathcal{L}$ by switch since the operation leaves the final entry the same. To generate a list ending with a by merge, it must interchange the final two positions on some list, say $Lcab \in \mathcal{L}$. But this is also impossible because the other list used in the merge operation would be $Lbca$, ending with a . Thus, if no list in $\mathcal{L}$ ends with a , the same is true for its span. This completes the proof.

7.4 Proof of Proposition 2

Only If: Suppose that $\rho^\mathcal{L}$ is rationalizable. Then there exists $\pi \in \Delta(\mathcal{P})$ such that for all $s \in \mathbf{S}$ and $L \in \mathcal{L}$,

$$\rho^\mathcal{L}(s, L) = \rho_\pi(s, L).$$

If $B \in \mathcal{B}^L \cap \mathcal{B}^{L'}$ for some $L, L' \in \mathcal{L}$, then

$$\rho^\mathcal{L}(s^L(B), L) = \rho_\pi(s^L(B), L) = \sum_{> \in B} \pi(>) = \rho_\pi(s^{L'}(B), L') = \rho^\mathcal{L}(s^{L'}(B), L').$$

We can similarly show that if $B \in \mathcal{B}^L$ and $B^c \in \mathcal{B}^{L'}$, then

$$\rho^\mathcal{L}(s^L(B), L) = \sum_{> \in B} \pi(>) = 1 - \sum_{> \in B^c} \pi(>) = 1 - \rho^\mathcal{L}(s^{L'}(B^c), L').$$

So $\rho^\mathcal{L}$ is consistent.

To prove that the corresponding consistent data game has a non-empty core, we will show that π belongs to the core. This requires us to check the following:

$$\sum_{> \in B} \pi(>) \geq v(B), \quad \forall B \subseteq \mathcal{P}$$

with equality for $B = \mathcal{P}$. Equality holds because π is a probability distribution over $\mathcal{P}$, and $v(\mathcal{P})$ is defined to be 1.

As for the inequality, we only need to check those B for which $v(B)$ is strictly positive. By the definition of v , such a set of preferences B must have the property that either B or B^c belongs to $\mathcal{B}^L$ for some $L \in \mathcal{L}$. If $B \in \mathcal{B}^L$, then

$$\sum_{> \in B} \pi(>) = \rho_\pi(s^L(B), L) = \rho^\mathcal{L}(s^L(B), L) = v(B).$$

If instead $B^c \in \mathcal{B}^L$, then

$$\sum_{> \in B} \pi(>) = 1 - \sum_{> \in B^c} \pi(>) = 1 - \rho_\pi(s^L(B^c), L) = 1 - \rho^\mathcal{L}(s^L(B^c), L) = v(B).$$

Thus π belongs to the core, proving the only if direction of the result.

If: Now suppose that the partial data $\rho^\mathcal{L}$ is consistent, and the corresponding consistent data game has a non-empty core. Let π be an element of the core. In particular, π must be a probability distribution because $v(\mathcal{P}) = 1$. We will show that $\rho^\mathcal{L}(s, L) = \rho_\pi(s, L)$ for all s and all $L \in \mathcal{L}$.

From the definition of the consistent data game, if $B \in \mathcal{B}^L$ for some $L \in \mathcal{L}$, then $v(B) = \rho^\mathcal{L}(s^L(B), L)$ and $v(B^c) = 1 - \rho^\mathcal{L}(s^L(B), L)$. Since π is in the core, we have the two inequalities $\sum_{> \in B} \pi(>) \geq v(B)$ and $\sum_{> \in B^c} \pi(>) \geq v(B^c)$. Adding up yields $1 \geq 1$, so both inequalities must hold equal. That is, whenever $B \in \mathcal{B}^L$ for some $L \in \mathcal{L}$, we must have

$$v(B) = \sum_{> \in B} \pi(>)$$

Now, for any $L \in \mathcal{L}$ and $s \in \mathbf{S}$, there exists a partition cell $B \in \mathcal{B}^L$ such that $s = s^L(B)$. It follows that

$$\rho^\mathcal{L}(s, L) = \rho^\mathcal{L}(s^L(B), L) = v(B) = \sum_{> \in B} \pi(>) = \rho_\pi(s^L(B), L) = \rho_\pi(s, L).$$

Thus $\rho^\mathcal{L}$ is rationalizable, proving the if direction of the result.