

A Separability Foundation for Random Coefficients Logit*

Fedor Sandomirskiy[†] Po Hyun Sung[‡] Omer Tamuz[§]
Ben Wincelberg[¶]

May 15, 2026

Abstract

We study stochastic choice across decision problems, each represented as a menu of action labels paired with observable outcome vectors. We propose a consistency condition for behavior in decision problems composed of two separable components: choice probabilities must agree with those obtained when each component is considered in isolation. Together with monotonicity and continuity, this separability requirement characterizes the family of random coefficients logit rules.

1 Introduction

Consider an analyst who studies individuals choosing among different actions, each yielding an observable outcome vector. From the analyst’s perspective, choice

*Previous versions of this paper were circulated under the titles “Decomposable Stochastic Choice” and “Independence of Irrelevant Decisions in Stochastic Choice.” We benefited from numerous suggestions and comments from our colleagues. We are grateful to Marina Agranov, Victor Aguiar, Marco Bernardi, Alexander Bloedel, Peter Caradonna, Gabriel Carroll, Daniel Chen, Geoffroy de Clippel, Federico Echenique, Francesco Fabbri, Drew Fudenberg, Michael Gibilisco, Alexander Guterman, Wade Hann-Caruthers, Jakub Kastl, Michael Keane, Annie Liang, Daniel Litt, Jay Lu, Yusufcan Masatlioglu, Jeffrey Mensch, Stephen Morris, Pietro Ortleva, Ariel Pakes, Tom Palfrey, Marcin Peški, Luciano Pomatto, Gil Refael, Philipp Sadowski, Todd Sarver, Ilya Segal, Barry Simon, Stanislav Smirnov, Tomasz Strzalecki, William Thomson, Aleh Tsyvinski, Christopher Turansick, Johan Ugander, Shoshana Vasserman, Ryan Webb, and Leeat Yariv.

[†]Princeton University. Email: fsandomi@princeton.edu

[‡]Caltech. Email: psung@caltech.edu

[§]Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a National Science Foundation CAREER award (DMS-1944153).

[¶]Caltech. Email: bwincelb@caltech.edu

behavior is stochastic: the population of decision makers may be heterogeneous, and, furthermore, they may choose actions with dominated outcomes for various reasons, including cognitive limitations, errors in the decision-making process, and random shocks unobserved by the analyst. The analyst would like to develop a model to predict out-of-sample choice probabilities. In this paper we use an axiomatic approach to restrict the analyst’s set of models. We show that random coefficients logit, a commonly used model, is unique in satisfying three simple axioms.¹

We model a single decision instance as a *menu* (A, o) consisting of a finite set of actions A , each associated with an outcome $o(a) \in \mathbb{R}^n$. Each dimension of the outcome is a quantity that can be added (i.e., an extensive quantity) such as money, time, fuel, and so on. Following the discrete choice literature (McFadden, 1981; Strzalecki, 2025), the actions are treated as mere labels, and the outcomes are assumed to carry all the information important to the decision maker. For example, consider the setting of a rideshare driver who faces the decision of which of several trips to accept. Each menu corresponds to a choice at a particular time. The actions in a menu are (meaningless) labels of trips, and the outcome associated with each action is a tuple $(r, -t, -g, -m)$ corresponding to the revenue from the trip, the time the trip will take, the gas it will consume and the mileage it will cover. More broadly, the decision maker can be thought of as a producer, facing the choice of which good to produce. Alternatively, the decision maker is a consumer, and outcomes represent consumption bundles. In either case, outcomes can incorporate risk, for instance through the variance of revenue or other uncertain quantities.

A stochastic choice rule assigns to each menu a probability distribution over the set of actions, which we interpret as the predicted choice probabilities. The collection of stochastic choice rules is a rich, non-parametric family that gives rise to the problem of model selection. We restrict this family by considering three axioms: *monotonicity*, *continuity*, and *independence of irrelevant decisions* (IID).

Our monotonicity axiom requires that an action that yields a dominating outcome (i.e., higher in every dimension) is taken with higher probability. This axiom

¹Random coefficients logit is a workhorse of empirical IO (Nevo, 2000; Moon et al., 2018; Gandhi and Nevo, 2021), popularized by the seminal paper of Berry, Levinsohn, and Pakes (1995). It belongs to the broader class of mixed logit models, although some empirical work uses the two terms interchangeably (see, e.g., the first paragraph in Nevo, 2000). Following Strzalecki (2025), we reserve the term mixed logit for mixtures over unrestricted utility functions and use random coefficients logit for mixtures of multinomial logit rules with linear utilities. See the literature review and §3.1 for further discussion.

places an ordinal restriction on the choice probabilities within a menu but places no restrictions across menus. Continuity is the technical assumption that a small change in outcomes leads to a small change in choice probabilities.

Our main axiom is independence of irrelevant decisions (IID). It is a condition imposed on additively separable menus, or, as we shall call them, *product menus*. Such menus represent situations where multiple choices are made together, and the outcome from one choice does not affect the outcome from another. Suppose that an individual has to choose one of the two actions from $A_1 = \{a, b\}$, and also one of the two actions from $A_2 = \{s, t\}$. We can think of these two choices as a combined choice in a menu with action set $A = \{(a, s), (a, t), (b, s), (b, t)\}$. When the outcome $o(a_1, a_2)$ of each combined action is a sum $o_1(a_1) + o_2(a_2)$ of outcomes of its two components, we call (A, o) a product menu. In the rideshare example, a product menu can correspond to the combination of two decisions at different points in time, where additive separability is natural since revenue, time, fuel, and mileage all aggregate across trips. Of course, not all menus of combined choices are product menus, as there may be interactions between the choices that make them non-separable.

IID requires that choice probabilities over A_1 are the same, regardless of whether (A_1, o_1) is faced in isolation or as part of the product menu (A, o) . The axiom does not rule out correlation across separable choices: a driver who favors shorter, less lucrative rides in the first trip may exhibit the same tendency in the second. Instead, IID captures a weak sense in which separability translates to behavior: decision makers' behavior in separable choices is consistent with their behavior when these choices are made in isolation. Importantly, for non-separable choices, where the menu is not a product menu, IID imposes no restrictions.

From the point of view of an analyst (say, employed by the rideshare company) who wants to predict the driver's choices, the IID assumption means that a stochastic choice rule will give the same prediction about a particular choice, regardless of whether it is considered together with other unrelated choices. Without IID, a model may provide different predictions for the driver's first choice of the day, depending on whether or not the analyst included later choices in the model. In a sense, when an analyst routinely excludes unrelated choices from their model, they are implicitly assuming that IID holds.

Our main result is that random coefficients logit rules are the only ones that satisfy monotonicity, continuity and IID (Theorem 1). Moreover, the coefficient distribution is identical across all menus. Thus, even though the IID axiom only

restricts predictions for product menus, its conjunction with monotonicity and continuity implies that all choices—including in non-product menus—are made according to the same rule.

This result provides a simple foundation for this widely used choice rule. It also shows that one of the axioms is violated by any other stochastic choice rule, including one-shot probit and separable probit. Indeed, one-shot probit violates IID, while separable probit violates monotonicity (see §5). The theorem thus highlights that modeling even one decision instance with a rule that is inconsistent with random coefficients logit carries hidden global assumptions. Regardless of how behavior is modeled on other menus, such a decision maker must either be influenced by the presence of irrelevant decisions (violating IID), fail to choose better actions with higher probability (violating monotonicity), or be extremely sensitive to outcomes (violating continuity).

In our main result we model separability by the addition of the outcomes in \mathbb{R}^n . In §6 we carry out a similar exercise, in a setting with an abstract outcome space and an abstract operation that captures separability. This highlights that it is separability—rather than the particular additive structure of \mathbb{R}^n —that is the main driver of our results.

1.1 Related Literature

In this paper we provide an axiomatization of random coefficients logit. This term is often conflated with mixed logit, and we follow [Strzalecki \(2025\)](#) to distinguish the two. Mixed logit allows unrestricted utility heterogeneity and can approximate any random utility model ([McFadden and Train, 2000](#)), so axiomatizing mixed logit is close in spirit to axiomatizing stochastic rationality more generally ([Manski, 1977](#); [Falmagne, 1978](#); [McFadden and Richter, 1990](#); [Clark, 1996](#)). A characterization of mixed logit along these lines was obtained by [Saito \(2018\)](#).

Random coefficients logit restricts heterogeneity to linear utility functions over outcomes in \mathbb{R}^n . This model has become the workhorse of empirical research owing to its parsimony and flexibility: it allows for structured taste heterogeneity and combines this with tractability and identification (see, e.g., [Berry et al., 1995](#); [Nevo, 2000](#); [Train, 2003](#); [Fox et al., 2012](#); [Moon et al., 2018](#); [Gandhi and Nevo, 2021](#)). To our knowledge, this paper is the first axiomatic characterization of this economically central subclass of mixed logit rules.

The restriction to linear utilities is substantive: linear-utility mixtures do not span all random utility behaviors unless the number of alternatives is small relative

to the outcome dimension (Saito, 2018; Lu and Saito, 2025). Such mixtures give rise to linear pure characteristic models (LPCM), which we discuss further in §3.1 alongside a characterization of a related class by Gul and Pesendorfer (2006).

In contrast to random coefficients logit, there is a large literature characterizing multinomial logit. Its early popularity stemmed from analytical tractability and its microfoundation as a random-utility model with Gumbel-distributed shocks (Luce and Suppes, 1965; McFadden, 1974). According to Luce (1959), a choice rule exhibits independence of irrelevant alternatives (IIA) if the relative probabilities for any subset of alternatives do not depend on the presence of other alternatives in the choice set. Luce (1959) demonstrated that any behavior satisfying IIA can be generated by multinomial logit for some choice of (possibly nonlinear) utilities. Variants of IIA have since been used to characterize various generalizations of multinomial logit (see, e.g., Gul, Natenzon, and Pesendorfer, 2014; Echenique, Saito, and Tserenjigmid, 2018; Ahn, Echenique, and Saito, 2018; Kovach and Tserenjigmid, 2022). Versions of multinomial logit with linear utilities can be characterized by augmenting IIA-type requirements with additional axioms pinning down the exponential dependence of likelihood ratios on outcomes (Yellott, 1977; Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini, 2021; Breitmoser, 2021; Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini, 2022). Unlike IIA, which constrains behavior across nearly identical menus such as those generated by duplicating an action as in (Debreu, 1960), IID places no restrictions in such settings and applies only to decisions made in separate, non-overlapping contexts.

A separate literature microfound multinomial logit and related rules through information or cognitive frictions. Matějka and McKay (2015) show that multinomial logit captures the behavior of a utility-maximizing individual with entropy-based attention costs (Sims, 2003). Woodford (2014) and Mattsson and Weibull (2002) derive related results for binary choices and costly effort, respectively. Steiner, Stewart, and Matějka (2017) obtain an entropy-cost characterization of dynamic logit; see also Fudenberg and Strzalecki (2015). In a multiperiod context, Fudenberg, Lanzani, and Strack (2025) show that mixed probit describes the limiting choice of an agent with bounded memory and Gaussian information. In contrast to this literature, our paper is agnostic about the origin of logit behavior, and instead shows that it is the only form of behavior compatible with consistent choices across separable decision problems.

Our analysis of separability in general outcome spaces (§6) connects to a classical literature showing that choice independence across dimensions yields additively

separable utility (Debreu, 1959; Fishburn, 1965; Segal, 1992; Wakker, 2013; Cho, 2022). Similarly, we demonstrate that choice independence across product menus implies that choices are driven by a utility function that is additively separable across the product components.

A closely related line of work on separability focuses on consistency of behavior across agents or periods. Chambers, Masatlioglu, and Turansick (2021) and Kashaev, Plávala, and Aguiar (2024) consider the choice behavior of two agents (or of a single agent over two periods) and study its separability, asking whether a joint distribution over choices can be rationalized by a single distribution over utility pairs. Similar cross-period separability restrictions appear in dynamic random-utility frameworks (Frick, Iijima, and Strzalecki, 2019; Li, 2021; Kashaev, Gauthier, and Aguiar, 2023). These papers share two features with ours: they impose consistency across joint decisions, and they conclude that separability pins down a stable underlying stochastic preference governing such choices. The key conceptual difference is that our separability requirement sharply disciplines model selection, leaving only random coefficients logit. A version of our separability-based axiomatic approach is explored by Sandomirskiy, Sung, Tamuz, and Wincelberg (2025) in a multi-agent strategic context, where it is used to axiomatize Nash and quantal response equilibria along with new solution concepts.

2 Model

Let \mathcal{A} be a universal set of actions. We assume that this set is non-empty and closed under the operation of forming ordered pairs. In other words, if $a_1, a_2 \in \mathcal{A}$ then the pair (a_1, a_2) is also an element of \mathcal{A} . We interpret (a_1, a_2) as a new action label representing the compound action of taking a_1 and a_2 together. Note that this condition implies that \mathcal{A} is infinite.² We further assume that \mathcal{A} is countable.

The set of possible outcomes of a decision is denoted by \mathcal{O} . A single decision instance is represented by a *menu* (A, o) , where $A \subset \mathcal{A}$ is a non-empty finite set of possible actions and $o: A \rightarrow \mathcal{O}$ assigns an outcome to each action. Note that the outcome of an action encapsulates all the information about this action relevant to the decision-maker. In contrast, the name of the action is just a label. In particular, the same action may appear in different menus and result in different outcomes.

The key setting we study is $\mathcal{O} = \mathbb{R}^n$; in §6 we generalize beyond \mathbb{R}^n . To build

²Indeed, if $a \in \mathcal{A}$, then $(a, a) \in \mathcal{A}$, and hence $((a, a), a) \in \mathcal{A}$, and so on.

intuition it is useful to keep in mind the example of $\mathcal{O} = \mathbb{R}$. This benchmark outcome space can be used to model decision-makers who compare actions by a single number, such as their monetary reward—and we accordingly refer to outcomes as payoffs. We use $\mathcal{O} = \mathbb{R}$ in the examples used to illustrate the definitions of this section.

We display a menu by showing each action’s outcome below it. For example,

$$(A, o) = \left\{ \begin{array}{cc} a & b \\ 3.14 & -17 \end{array} \right\}$$

is a menu with two actions, choosing a or b , with the former having an outcome of 3.14 and the latter having an outcome of -17 .

The collection of all menus is denoted by \mathcal{M} . Thus, when $\mathcal{O} = \mathbb{R}^n$,

$$\mathcal{M} = \{(A, o) : A \subset \mathcal{A} \text{ is finite and non-empty, } o : A \rightarrow \mathbb{R}^n\}.$$

Since actions and outcomes are specified separately, two menus may have the same action set but assign different outcomes to its elements.

This framework sits between two standard stochastic-choice domains. In the classical random-utility tradition, a decision problem is a set of primitive alternatives, and all choice-relevant information is carried by the identities of those alternatives (see, e.g., [Block and Marschak, 1959](#); [Luce and Suppes, 1965](#); [Falmagne, 1978](#)). In pure-characteristics models, alternatives are represented directly as points in \mathbb{R}^n , so alternatives with the same vector are identified at the level of the model (see, e.g., [Saito, 2018](#); [Lu and Saito, 2025](#)). Our framework keeps these two roles distinct. As in characteristics-based models, the relevant information about action a in menu (A, o) is contained in $o(a)$. However, we do not reduce the menu to the set of outcomes $o(A) \subset \mathbb{R}^n$. Retaining action labels lets us form compound actions, such as (a_1, a_2) , and relate actions across different menus. The roles played by actions and outcomes in our framework fit the game-theoretic tradition in which the actions available to a player are mere labels and only the payoffs matter; indeed, for $\mathcal{O} = \mathbb{R}$, a menu (A, o) can be viewed as a one-player normal-form game with action set A and payoff function o .

A *stochastic choice rule* is a map Φ that assigns to each menu $(A, o) \in \mathcal{M}$ a probability distribution over A . We denote by $\Phi(a \mid A, o)$ the probability that $\Phi(A, o)$ assigns to $a \in A$. We think of Φ as describing or predicting the choices of a decision maker across different situations.

One family of widely used stochastic choice rules consists of the *independent*

additive random utility models (IARU), which are given by

$$\text{IARU}^u(a | A, o) = \mathbb{P}\left[u(o(a)) + \varepsilon_a = \max_{b \in A} u(o(b)) + \varepsilon_b\right],$$

where $u: \mathbb{R}^n \rightarrow \mathbb{R}$ is a utility function and $(\varepsilon_b)_{b \in A}$ are independent shocks with a common continuous CDF F .

When these shocks follow the Gumbel distribution, i.e., $F(t) = \exp(-\exp(-t))$, this is the *multinomial logit rule* (MNL), which is given by

$$\text{MNL}^u(a | A, o) = \frac{\exp(u(o(a)))}{\sum_{b \in A} \exp(u(o(b)))}.$$

A particularly important subclass is *linear multinomial logit*, in which $u(x) = \beta \cdot x$ for some $\beta \in \mathbb{R}^n$:

$$\text{MNL}^\beta(a | A, o) = \frac{\exp(\beta \cdot o(a))}{\sum_{b \in A} \exp(\beta \cdot o(b))}.$$

Multinomial logit is commonly used in the empirical literature for its computational tractability, but it does not allow for random taste variation and various substitution patterns (Train, 2003). These limitations are overcome by the *random coefficients logit rule* (RCL), which is a weighted average of linear multinomial logit rules, given by

$$\text{RCL}^\mu(a | A, o) = \int \text{MNL}^\beta(a | A, o) d\mu(\beta).$$

It is parameterized by a probability measure μ over the logit parameter $\beta \in \mathbb{R}^n$. More generally than random coefficients logit, one can consider mixtures of (non-linear) multinomial logit rules, which is the class of mixed logit rules. In the empirical literature these terms are often conflated (Nevo, 2000; Gandhi and Nevo, 2021), but in our setting the difference is important, as we discuss in §3.1.

We consider several properties of stochastic choice rules. The first one is monotonicity. We consider the set of outcomes \mathbb{R}^n together with the usual component-wise partial order \geq , so that higher outcomes are universally preferred.³ Monotonicity reflects this preference. In particular, it limits the possible influence of action labels on choice behavior.⁴

³Our results hold more generally for any partial order, including the trivial one in which each outcome is only comparable to itself. In §6 we formally consider general partial orders.

⁴While our model is motivated by the view that only outcomes matter to the decision maker, we do not directly impose this as an axiom or hardwire it into the definition of a choice rule. Nevertheless, dependence on labels will be ruled out by the combination of the other axioms we impose.

Axiom 1 (Monotonicity). A rule Φ is *monotone* if for any menu (A, o) in \mathcal{M} and any $a, a' \in A$ such that $o(a) \geq o(a')$ it holds that $\Phi(a | A, o) \geq \Phi(a' | A, o)$.

Note that this axiom does not impose any constraints across menus, but only within a given menu.

Monotonicity is satisfied by all IARU models with monotone utilities. Moreover, the class of stochastic choice rules that satisfy monotonicity is convex, i.e., a mixture of monotone rules is monotone. In particular, random coefficients logit rules whose coefficient distributions are supported on $\mathbb{R}_{\geq 0}^n$ satisfy monotonicity.

For a fixed set of actions A , we say that a sequence of menus (A, o_n) converges to (A, o) if $\lim_n o_n(a) = o(a)$ for all $a \in A$.

Axiom 2 (Continuity). A rule Φ is *continuous* if for any sequence of menus (A, o_n) from \mathcal{M} converging to (A, o) , we have $\lim_n \Phi(a | A, o_n) = \Phi(a | A, o)$ for all $a \in A$.

Alternatively put, continuity stipulates that very small changes in the outcomes result in very small changes in choice probabilities. This axiom excludes stochastic choice rules that describe individuals who pay excessive attention to even negligible differences in outcomes. For instance, it is violated by rules for $\mathcal{O} = \mathbb{R}$ that always select one of the highest-outcome actions, regardless of how small the advantage is. Nevertheless, continuity is a common modeling choice made for good reason: people do not always choose the dominant action, especially when the difference between outcomes is minuscule.

Our main axioms concern choice rule predictions on menus that represent multiple decisions with additively separable outcomes. We say that (A, o) is a *product menu* if

$$A = A_1 \times A_2 \quad \text{and} \quad o(a_1, a_2) = o_1(a_1) + o_2(a_2). \quad (1)$$

That is, A consists of action pairs $a = (a_1, a_2)$ with $a_1 \in A_1$ and $a_2 \in A_2$, and the outcome assigned to each pair is additively separable. We write $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ and refer to (A, o) as the product of (A_1, o_1) and (A_2, o_2) .

For example, suppose an experimenter runs two tests consecutively on the same subject. In the first test, the subject chooses between two actions a and b , and receives payoff 0 or 1, respectively. In the second, the subject chooses between s and t , and again receives 0 or 1. Indeed, many experiments contain comprehension questions and pay subjects for each correct answer.

The first test would be well-modeled by the menu (A_1, o_1) , with $A_1 = \{a, b\}$ and $o_1(a) = 0$, $o_1(b) = 1$. Likewise, the second test would be well-modeled by

(A_2, o_2) , with $A_2 = \{s, t\}$ and $o_2(s) = 0$, $o_2(t) = 1$. The joint decision the subject faces is the product menu

$$(A, o) = \left\{ \begin{array}{cc} a & b \\ 0 & 1 \end{array} \right\} \otimes \left\{ \begin{array}{cc} s & t \\ 0 & 1 \end{array} \right\} = \left\{ \begin{array}{cccc} (a,s) & (a,t) & (b,s) & (b,t) \\ 0 & 1 & 1 & 2 \end{array} \right\} = \left\{ \begin{array}{c|cc} & s & t \\ a & 0 & 1 \\ b & 1 & 2 \end{array} \right\}.$$

Alternatively, the experimenter could only award the subject if both questions are answered correctly. This experiment would be well-modeled by the menu

$$(A, o') = \left\{ \begin{array}{cccc} (a,s) & (a,t) & (b,s) & (b,t) \\ 0 & 0 & 0 & 1 \end{array} \right\} = \left\{ \begin{array}{c|cc} & s & t \\ a & 0 & 0 \\ b & 0 & 1 \end{array} \right\}.$$

This is not a product menu, even though the action set is a product set (indeed, the same product set), because the outcomes are not additively separable. This example highlights that we do not assume that arbitrary combinations of decisions give rise to product menus. While any such combination involves taking the Cartesian product of the underlying action sets, the associated outcomes need not take the separable form (1).

For an example with $\mathcal{O} = \mathbb{R}^n$, consider the setting of a rideshare driver who faces decisions of which of several trips to accept. Each menu corresponds to a choice at a particular time. The actions in a menu are (meaningless) labels of trips, and the outcome associated to each action is a tuple $(r, -t, -g, -m)$ corresponding to the revenue from the trip, the time the trip will take, the gas it will consume and the mileage it will cover. A product menu corresponds to the combined decision in two subsequent rides, where separability is natural, as all the above quantities are added across trips. Beyond the specific example of a driver, the decision maker can be thought of as a business owner, facing the choice of which jobs to accept, with a vector of various quantities describing the possible outcomes. Quantities could also include risk (e.g., the variance of the revenue or of other uncertain outcomes). A similar interpretation is that the decision maker is a consumer, and outcomes represent consumption bundles.

A multidimensional outcome space can also capture choice under ambiguity. Let Θ be a finite set of states. An outcome x is a function $x: \Theta \rightarrow \mathbb{R}$, specifying state-contingent payoffs. That is, x is a Savage act and the space of all outcomes can be identified with $\mathcal{O} = \mathbb{R}^\Theta$. A decision-maker is ambiguous about the state $\theta \in \Theta$ and so may take into account all the possible values x_θ . Product menus then correspond to combinations of two decisions leading to separable state-contingent payoffs.

Our main axiom, IID, restricts the choice rule only on product menus. It requires that choice predictions for a product menu be consistent with predictions for each component considered separately. We begin by discussing a stronger assumption, which additionally requires statistical independence of choices across the two dimensions of a product menu.

Axiom 3 (Decomposability). *A rule Φ satisfies **decomposability** if for all product menus $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ in \mathcal{M} , it holds that*

$$\Phi((a_1, a_2) \mid A, o) = \Phi(a_1 \mid A_1, o_1) \cdot \Phi(a_2 \mid A_2, o_2)$$

for all $(a_1, a_2) \in A$.

For example, suppose we observe the choice probabilities of (A_1, o_1) and (A_2, o_2) to be

$$\Phi \left(\begin{array}{cc} a & b \\ 0 & 1 \end{array} \right) = \Phi \left(\begin{array}{cc} s & t \\ 0 & 1 \end{array} \right) = (1/3, 2/3).$$

Decomposability requires that for the product menu,

$$\Phi \left(\begin{array}{c|cc} & s & t \\ \hline a & 0 & 1 \\ b & 1 & 2 \end{array} \right) = \begin{array}{c|cc} & s & t \\ \hline a & 1/9 & 2/9 \\ b & 2/9 & 4/9 \end{array}.$$

Decomposability means that, for product menus, the prediction is the same as when that decision is made in isolation. Moreover, the predicted distribution is statistically independent across the two dimensions. In the experimental lab example, this would imply that subjects choose the wrong answer independently in the two questions they are asked, when they are rewarded separately for each correct answer. Decomposability imposes no restriction on predicted behavior in situations where subjects are rewarded only for answering both questions correctly, as such situations do not correspond to product menus.

While decomposability is a simple separability assumption, in some settings its independence component may be unrealistic, especially in the presence of unobservable heterogeneity. In the lab example, one could expect that subjects who answer the first question correctly are more likely to also answer the second correctly. In the driver example, it is natural to expect that the choices a driver makes through the day are not necessarily independent of each other.

The IID axiom is a weakening of decomposability, allowing for such correlations while maintaining consistency with predictions for the components. Given a product menu $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$, we denote the marginal choice probability

of $a_1 \in A_1$ by

$$\Phi(a_1 | A, o) = \sum_{a_2 \in A_2} \Phi((a_1, a_2) | A, o).$$

Axiom 4 (IID). *A rule Φ satisfies **independence of irrelevant decisions** if for all product menus $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ in \mathcal{M} , it holds that*

$$\Phi(a_1 | A, o) = \Phi(a_1 | A_1, o_1) \tag{2}$$

for all $a_1 \in A_1$.

For example, for the menu $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ above, IID implies that if

$$\Phi \left(\begin{array}{cc} a & b \\ 0 & 1 \end{array} \right) = \Phi \left(\begin{array}{cc} s & t \\ 0 & 1 \end{array} \right) = (1/3, 2/3),$$

then

$$\Phi(a | A, o) = \Phi((a, s) | A, o) + \Phi((a, t) | A, o) = \frac{1}{3}.$$

Thus, IID allows for predictions such as

$$\Phi \left(\begin{array}{c|cc} & s & t \\ \hline a & 0 & 1 \\ b & 1 & 2 \end{array} \right) = \begin{array}{c|cc} & s & t \\ \hline a & 1/6 & 1/6 \\ b & 1/6 & 3/6 \end{array},$$

which is not a product measure, but it does not allow

$$\Phi \left(\begin{array}{c|cc} & s & t \\ \hline a & 0 & 1 \\ b & 1 & 2 \end{array} \right) = \begin{array}{c|cc} & s & t \\ \hline a & 1/8 & 2/8 \\ b & 2/8 & 3/8 \end{array}.$$

Like decomposability, the IID axiom imposes no restrictions on choice rule predictions for non-product menus such as (A, o') .

IID means that predictions are independent of the inclusion of an unrelated menu into the analysis. In other words, for product menus, the predicted choice frequency of an action in the first dimension is consistent with predictions when only that dimension is considered. Without the assumption of IID, a modeler would have to include all irrelevant decisions that the population faces in order to make an accurate prediction.

IID is satisfied by random coefficients logit, and in particular is satisfied by linear multinomial logit, which moreover satisfies decomposability. The class of

rules that satisfy IID is very large. For instance, we can construct such a rule Φ by specifying any choice distribution on each non-product menu, and then defining Φ on product menus to be the product measure given by Φ on each component menu. This construction will in fact satisfy decomposability, and is very flexible.

A natural example of this sort is a class of models that we call *separable IARU* (SIARU), which is defined as follows. Fix a utility $u: \mathbb{R}^n \rightarrow \mathbb{R}$. For non-product menus,⁵ it coincides with the one-shot rule IARU^u . For product menus, it is given by

$$\text{SIARU}^u((a_1, a_2) \mid A, o) = \text{SIARU}^u(a_1 \mid A_1, o_1) \cdot \text{SIARU}^u(a_2 \mid A_2, o_2). \quad (3)$$

That is, choices are made independently in each dimension. For example, if a consumer makes two unrelated choices—e.g., purchasing a milk and a soap—they choose the best milk subject to some noise and the best soap subject to some additional noise, with the noises independent for the two choices. Such a model could also be applied in a dynamic setting: a consumer chooses milk both today and tomorrow, and faces independent shocks for the two periods.⁶ See §5 for a formal definition and discussion.

2.1 Implications of the Axioms

The main result of this paper is a characterization of all monotone, continuous rules that satisfy IID. Before presenting our main result, we illustrate that while each of these axioms on its own is rather weak, together they have surprisingly strong implications, and already for $\mathcal{O} = \mathbb{R}$.

For example, consider an analyst who is interested in choice probabilities in the menu

$$(A, o) = \left\{ \begin{array}{ccc} a_1 & a_2 & a_3 \\ 0 & 2 & 7 \end{array} \right\}.$$

Suppose that the analyst observes the choice probability of the action a_1 to be 5%, i.e., $\Phi(a_1 \mid A, o) = 5\%$.

Clearly, monotonicity implies $\Phi(a_3 \mid A, o) \geq 47.5\%$, since $\Phi(a_2 \mid A, o) + \Phi(a_3 \mid A, o) = 95\%$ and $\Phi(a_2 \mid A, o) \leq \Phi(a_3 \mid A, o)$. Without monotonicity, IID yields no constraints for $\Phi(a_3 \mid A, o)$, since (A, o) is not a product menu (indeed, even decomposability has no further implications). Naively, the combination of monotonicity and IID does not seem to imply any further constraints.

⁵For present purposes, these include products in which one of the menus is trivial, in the sense that it includes only one action.

⁶We thank an anonymous referee for suggesting this example.

Surprisingly, this intuition is wrong. It turns out that if we assume that Φ is monotone and satisfies IID, then $\Phi(a_1 | A, o) = 5\%$ implies that $\Phi(a_3 | A, o) \geq 83.8\%$. This bound is tight: if we make the stronger assumption that Φ is monotone and decomposable, then $\Phi(a_1 | A, o) = 5\%$ implies that $\Phi(a_3 | A, o) = 83.8\%$. As we explain in detail below, this is a consequence of our main results. Furthermore, there is a unique Φ that is monotone, decomposable, continuous, and satisfies $\Phi(a_1 | A, o) = 5\%$; see Appendix §F for details.

3 IID and Random Coefficients Logit

Consider the outcome space $\mathcal{O} = \mathbb{R}^n$. For μ supported on $\mathbb{R}_{\geq 0}^n$, the random coefficients logit rule given by

$$\text{RCL}^\mu = \int \text{MNL}^\beta d\mu(\beta) \quad (4)$$

satisfies monotonicity, continuity, and IID. Our main theorem states that no other rules satisfy these axioms.

Theorem 1. *Let Φ satisfy monotonicity, continuity, and IID. Then Φ coincides with a random coefficients logit rule RCL^μ for some μ supported on $\mathbb{R}_{\geq 0}^n$.*

The theorem is proved in §7 for $\mathcal{O} = \mathbb{R}$ and the multidimensional case is covered in Appendix C. Some of the core ideas are explained in §4.

An interpretation of the theorem is that any decision maker satisfying the assumptions behaves as a stochastic utility maximizer with linear utility, random taste, and random noise level: the taste is encoded by the direction of β , while the noise level is inversely proportional to its magnitude. The variation in β may reflect heterogeneity across a population or variability of the internal state of a single individual. Conditional on each realization of β , choices follow a linear multinomial logit rule. The realized β is unobservable and thus the observable choice takes the form of a mixture over such logits, with mixing distribution μ . Notably, the fact that choice probabilities follow a random utility model emerges from the axioms and is not assumed a priori.

In the one-dimensional case $\mathcal{O} = \mathbb{R}$ where outcomes are interpreted as payoffs, $\beta \in \mathbb{R}_{\geq 0}$ determines the (random) noise level of a decision maker who is optimizing, but may stochastically choose a dominated outcome. In the example of the rideshare driver with outcomes $(r, -t, -g, -m)$ where the revenue r is measured in dollars, we can write $\beta = \gamma \cdot p$, where γ is in $\mathbb{R}_{\geq 0}$ and $p = (p_1, p_2, p_3, p_4)$ is a non-negative vector normalized to $p_1 = 1$. We can now interpret the components

p_2, p_3 and p_4 as the prices that the driver assigns to their time, gas and mileage. The driver is stochastically optimizing their expected profit, with noise γ . In the choice under ambiguity example, where outcomes are vectors of payoffs indexed by state θ , one can instead normalize p to sum to one, recasting it as a probability measure over the state space Θ . Under this normalization, the decision maker becomes a Bayesian stochastic expected payoff maximizer.

When some coordinates of the outcome vector represent costs rather than payoffs, it is natural to adjust the notion of monotonicity so that the decision maker favors actions with lower values in those coordinates.⁷ Theorem 1 extends straightforwardly to arbitrary partial orders on \mathbb{R}^n , since its proof invokes the monotonicity axiom only to ensure that two actions with identical outcomes are chosen with equal probability. Accordingly, if monotonicity is imposed with respect to a given partial order \succsim' on \mathbb{R}^n , the corresponding version of Theorem 1 states that the rule coincides with RCL $^\mu$ for a measure μ supported on the set of $\beta \in \mathbb{R}^n$ for which the linear function $x \rightarrow \beta \cdot x$ is \succsim' -monotone. At the extreme, taking \succsim' to be the trivial order ($x \succsim' y$ if and only if $x = y$) reduces monotonicity to the requirement that equal-outcome actions are selected with equal probability, and μ can be supported on all of \mathbb{R}^n . We discuss monotonicity with respect to general partial orders further in §6.

The theorem highlights several additional consequences of the axioms, beyond explicitly characterizing the form of choice rules consistent with them. First, while our definition of a stochastic choice rule and each axiom in isolation permit dependence on action labels, the axioms jointly rule this out: only the profile of outcomes in the menu can affect choice probabilities, eliminating framing effects. Second, although IID—the only axiom linking behavior across menus—places no restrictions on non-product menus by itself, its interaction with the other axioms pins down behavior on all menus. As a result, the same random coefficients logit rule governs choice even in compound decisions that are not representable as product menus—for example, those involving interrelated choices or shared constraints.

⁷For instance, if outcomes in the rideshare example are written as (r, t, g, m) rather than $(r, -t, -g, -m)$, it is natural to impose monotonicity with respect to the order \succsim' on \mathbb{R}^4 defined by $(r, t, g, m) \succsim' (r', t', g', m')$ whenever $r \geq r'$, $t \leq t'$, $g \leq g'$, and $m \leq m'$.

3.1 Mixed Logit, Random Coefficients Logit, and Universal Approximation

It is instructive to discuss the connection between our characterization of random coefficients logit and the universal approximation property of mixed logit rules established by [McFadden and Train \(2000\)](#). Effectively, they show that any aggregate choice behavior of a population of rational individuals can be approximated with general mixed logit rules, i.e., mixtures of MNL^u over different utilities u , as long as one can choose utilities from a rich enough class such as all polynomials of high degree.

This property of mixed logit highlights that it is an extremely flexible model, making it challenging to identify and estimate. In contrast, random coefficients logit—in which the utilities u are restricted to be linear—is widely used because it is parsimonious and hence identifiable, while retaining sufficient flexibility to capture many phenomena.⁸

Theorem 1 does not characterize the entire class of mixed logit rules, but rather just the subclass of mixtures corresponding to linear utilities of the form $u(x) = \beta \cdot x$, which gives rise to the random coefficients logit rules. In particular, not all mixed logit rules satisfy IID: among the continuous and monotone ones, IID is satisfied only by random coefficients logit.

An approximation result for random coefficients logit was explored by [Lu and Saito \(2025\)](#), who show that such mixtures can approximate any *linear pure characteristic model* (LPCM). These rules represent the aggregate choice behavior of a population of rational individuals with linear utilities and are defined as follows

$$\text{LPCM}^\mu(a \mid A, o) = \int \frac{\mathbb{1}_{a \in \arg \max_{b \in A} \beta \cdot o(b)}}{|\arg \max_{b \in A} \beta \cdot o(b)|} d\mu(\beta).$$

Since $\text{MNL}^{t\beta}$ with $t \rightarrow +\infty$ approximates the choice of a rational individual with $u(x) = \beta \cdot x$, one can show that $\text{LPCM}^\mu(a \mid A, o)$ can be approximated by mixing over such linear logits with β distributed according to μ . As limits of rules satisfying IID and monotonicity, LPCM^μ with μ supported on $\mathbb{R}_{\geq 0}^n$ also satisfies

⁸Of course, if one expands the vector of outcomes associated with an action by adding extra coordinates equal to non-linear functions of the original outcome vector (e.g., powers of other coordinates, as in [McFadden and Train, 2000](#)), then the linear functions effectively describe previously non-linear ones. However, this hurts identification, and hence cannot be over-used (see, e.g., [Fox, il Kim, Ryan, and Bajari, 2012](#), Assumption 2.1). Relatedly, the distinction between linear and non-linear utilities vanishes when the number of alternatives is smaller than the number of outcomes, as in [Saito \(2018\)](#).

these axioms (this can also be verified directly). However, LPCM^u violates the continuity assumption and thus is ruled out in Theorem 1.⁹

4 Correlations, Decomposability, and Linear Multinomial Logit

The IID axiom does not preclude correlations between separable choices. In this section we tackle two questions. First, we want to better understand what types of correlations IID allows, once it is taken in conjunction with continuity and monotonicity. Second, we would like to understand the implications of requiring that there are no correlations, i.e., replacing IID with the stronger axiom of decomposability.

Our first observation is that in the one-dimensional $\mathcal{O} = \mathbb{R}$ case, rules satisfying IID, monotonicity and continuity (i.e., random coefficients logit rules, by Theorem 1) cannot display negative correlations. Since variability in choice in the one-dimensional case stems from mistakes rather than taste variations, we interpret this result as stating that mistakes must be non-negatively correlated.

Formally, we define correlation as follows. Given a menu (A, o) and a function $f: A \rightarrow \mathbb{R}$ we will write $\Phi(f(a) \mid A, o) = \sum_a f(a)\Phi(a \mid A, o)$ for the expectation of f under the probability measure $\Phi(\cdot \mid A, o)$. A stochastic choice rule Φ exhibits *non-negative correlation* on a product menu $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ if

$$\Phi\left(o_1(a_1) \cdot o_2(a_2) \mid A, o\right) \geq \Phi\left(o_1(a_1) \mid A, o\right) \cdot \Phi\left(o_2(a_2) \mid A, o\right). \quad (5)$$

We say that Φ exhibits *zero correlation* if (5) holds with equality. A menu (A, o) is said to be non-trivial if o is non-constant.

In the example of the lab experiment in which every correct answer yields a payoff of one, non-negative correlation means that when a subject answers the

⁹A characterization of a class of rules similar to LPCM, though on a different domain, is obtained by Gul and Pesendorfer (2006). They consider a decision maker facing finite choice sets of lotteries and characterize random expected utility. Interpreting lottery weights as outcome vectors, random expected utility models become mathematically equivalent to LPCM. The key axioms are linearity, playing the role of the random-choice analogue of the von Neumann–Morgenstern independence axiom (equivalently, invariance to affine transformations of the outcome vectors); monotonicity across nested choice sets; and extremeness requiring that only choices maximizing some linear utility function receive positive weight. Random coefficients logit violates linearity because rescaling outcomes alters the relative magnitude of logit noise, and it violates extremeness because logit assigns positive probability to every choice, including dominated ones.

first question correctly they are (weakly) more likely to answer the second one correctly.

Proposition 1. *For any distribution μ on \mathbb{R} , RCL^μ exhibits non-negative correlation on every product menu. Moreover, RCL^μ exhibits zero correlation on a product of non-trivial menus if and only if μ is a Dirac measure.*

Intuitively, because the mixing distribution μ is fixed across all menus, lower noise levels (and thus better decisions) tend to occur simultaneously across the components of a product menu. This pattern is a joint consequence of the three axioms rather than of IID, which allows for arbitrary correlation in product menus.

The proof of Proposition 1 appears in §B, and is straightforward. Together with Theorem 1, we conclude that every rule that satisfies our axioms has non-negative correlation, which is not immediately evident from the axioms. In §5, we provide an example demonstrating that probit models do not always exhibit non-negative correlation, highlighting the connection between this property and our axioms.

Next, we ask which of the rules satisfying our axioms furthermore satisfy decomposability, i.e., exhibit no correlations across product menus. The following result holds for outcomes $\mathcal{O} = \mathbb{R}^n$.

Proposition 2. *Let Φ satisfy monotonicity, continuity and decomposability. Then Φ coincides with a linear multinomial logit rule MNL^β for some $\beta \in \mathbb{R}_{\geq 0}^n$.*

These propositions imply the well-known fact (see Fox, il Kim, Ryan, and Bajari, 2012) that a non-degenerate mixture of multinomial logit rules cannot itself be a multinomial logit rule. This yields a geometric interpretation of Theorem 1: the set of choice rules satisfying monotonicity, continuity, and IID is convex, with multinomial logit rules as its extreme points.¹⁰

The proof of Proposition 2 appears in Appendix D, and uses Theorem 1 and Proposition 1. We now sketch a direct proof for Proposition 2 for the case $\mathcal{O} = \mathbb{R}$.

¹⁰An extreme point of a convex set is one that cannot be written as a non-trivial convex combination of others. In this light, representation (4) mirrors the Choquet theorem: any choice rule satisfying the axioms can be expressed as a mixture of extreme points. Fox, il Kim, Ryan, and Bajari (2012) show that this representation is unique, that is $\text{RCL}^\mu = \text{RCL}^{\mu'}$ implies $\mu = \mu'$, so the space of such rules forms an infinite-dimensional simplex. This fact also follows from Proposition 5 in §7, which is a part of the proof of our main theorem. In fact, this proposition shows that μ is already identified by the choice probabilities on menus with integer outcomes.

This also provides a simple explanation of the core ideas of the proof of our main result, Theorem 1. The formal proof of Theorem 1 follows a similar path and makes use of the De Finetti Theorem to show that Φ is a convex combination of decomposable rules.

Suppose Φ is decomposable and monotone. We show that Φ (restricted to menus with rational outcomes) is identified by its predictions for a single menu, and that this implies that Φ is a multinomial logit rule. Assume we know $\Phi(B, r)$ for

$$(B, r) = \left\{ \begin{array}{cc} b_0 & b_1 \\ 0 & 1 \end{array} \right\}.$$

For the sake of this proof sketch, suppose also that both b_0 and b_1 are chosen with positive probability. Our goal is to show how this knowledge pins down Φ on any given menu (A, o) . Since multinomial logit rules are monotone and decomposable, this will immediately imply that Φ is a multinomial logit rule, with the parameter β chosen to agree with Φ on (B, r) .

For clarity, consider the particular example

$$(A, o) = \left\{ \begin{array}{ccc} a_1 & a_2 & a_3 \\ -17 & -17 & 42 \end{array} \right\}.$$

The same idea will apply to any (A, o) .

By monotonicity $\Phi(a_1 | A, o) = \Phi(a_2 | A, o)$. We will demonstrate that $\Phi(a_2 | A, o)$ and $\Phi(a_3 | A, o)$ satisfy a certain identity. Consider the product of (A, o) with the n -fold product of (B, r) :

$$(A, o) \otimes \underbrace{(B, r) \otimes (B, r) \otimes \cdots \otimes (B, r)}_{n \text{ times}}, \quad (6)$$

where $n = o(a_3) - o(a_2) = 59$. In this menu, the two actions $(a_3, b_0, b_0, \dots, b_0)$ and $(a_2, b_1, b_1, \dots, b_1)$ have the same outcome, and thus have the same probability by monotonicity. Therefore, decomposability implies

$$\Phi(a_3 | A, o) \cdot \Phi(b_0 | B, r)^{59} = \Phi(a_2 | A, o) \cdot \Phi(b_1 | B, r)^{59}. \quad (7)$$

Combined with the identities $\Phi(a_1 | A, o) = \Phi(a_2 | A, o)$ and $\Phi(a_1 | A, o) + \Phi(a_2 | A, o) + \Phi(a_3 | A, o) = 1$, this equation pins down $\Phi(A, o)$, which is therefore determined by $\Phi(B, r)$. Since $|B| = 2$, we can always choose $\beta \geq 0$ such that $\Phi(B, r) = \text{MNL}^\beta(B, r)$. Since multinomial logit also satisfies the same identities, we conclude that $\Phi(A, o) = \text{MNL}^\beta(A, o)$.

5 Probit Rules

Probit rules are a natural and widely used family of stochastic choice rules. In this section we study two natural variants of these rules and explore their relation to our axioms. For simplicity, both are defined for outcomes $\mathcal{O} = \mathbb{R}$.

One-shot linear probit is an independent additive random utility (IARU) model with normal shocks, so that

$$\text{Probit}(a \mid A, o) = \mathbb{P}\left[o(a) + \varepsilon_a = \max_{b \in A} o(b) + \varepsilon_b\right],$$

where ε_b are independent Gaussians. *Separable linear probit* is a separable IARU model in which each component in a product menu receives its own independent normal shocks (see the formal definition below). Our main result implies that these rules must violate our axioms. We show that the popular one-shot probit rule violates IID. Perhaps more surprisingly, we demonstrate that separable probit—which does satisfy IID—violates monotonicity. We also show that probit rules may exhibit negative correlation, unlike random coefficients logit.

Consider the simple one-shot probit rule with standard Gaussian $N(0, 1)$ shocks. We examine its predictions for the menu

$$(B, r) = \left\{ \begin{array}{cc} b_0 & b_1 \\ 0 & 1 \end{array} \right\} \quad (8)$$

and its “square”

$$(C, s) = (B, r) \otimes (B, r) = \begin{array}{c|cc} & b_0 & b_1 \\ \hline b_0 & 0 & 1 \\ b_1 & 1 & 2 \end{array} .$$

The probit predictions are

$$\text{Probit}(B, r) \simeq \begin{array}{cc} b_0 & b_1 \\ \hline 0.24 & 0.76 \end{array} \quad \text{and} \quad \text{Probit}(C, s) \simeq \begin{array}{c|cc} & b_0 & b_1 \\ \hline b_0 & 0.033 & 0.175 \\ b_1 & 0.175 & 0.617 \end{array} .$$

Since the marginal probability of choosing b_1 from (C, s) is $0.175 + 0.617 = 0.792$, which is greater than the choice probability of b_1 from (B, r) , the probit rule violates IID. We also note that, unlike multinomial logit, which exhibits zero correlation in product menus, the probit predictions for (C, s) exhibit negative correlation, as

$$\text{Probit}((b_1, b_1) \mid C, s) \simeq 0.617 < 0.627 \simeq \text{Probit}(b_1 \mid C, s)^2 .$$

The reason underlying both effects is that the Gaussian distribution's tails are too light. When outcome differences are large, the light tails cause probit weights to drop too quickly, much more so than in the logit case. When the differences are small relative to the noise variance, probit is less sensitive to those differences than logit. This leads to insufficient mass on (b_0, b_0) and excessive probability on mixed pairs like (b_1, b_0) and (b_0, b_1) , thus generating negative correlation and IID violation.

Separable linear probit (SProbit) is a separable IARU model, where the utility is linear and shocks are Gaussian. As in (3), it coincides with one-shot linear probit for non-product menus, and for product menus it is given by

$$\text{SProbit}((a_1, a_2) \mid A, o) = \text{SProbit}(a_1 \mid A, o) \cdot \text{SProbit}(a_2 \mid A, o).$$

Unlike one-shot linear probit, in separable linear probit shocks are applied independently in each dimension. Since utilities are linear, we can think of the decision maker as choosing separately in each dimension, or choosing the best combined choice; both yield the same choice probabilities.

By construction, separable linear probit satisfies decomposability and thus IID. We now show that it violates monotonicity. Consider the menus

$$(A_1, o_1) = \left\{ \begin{array}{cc} a & b \\ 0 & 9 \end{array} \right\}, \quad (A_2, o_2) = \left\{ \begin{array}{cc} c & d \\ 0 & 6 \end{array} \right\}, \quad (A_3, o_3) = \left\{ \begin{array}{cc} e & f \\ 0 & 6 \end{array} \right\},$$

and let

$$(A, o) = (A_1, o_1) \otimes (A_2, o_2) \otimes (A_3, o_3).$$

Let G denote the CDF of the difference of two independent standard normals, i.e., a normal distribution with mean zero and variance 2. Then $\text{SProbit}(b \mid A_1, o_1) = G(9)$ and $\text{SProbit}(d \mid A_2, o_2) = \text{SProbit}(f \mid A_3, o_3) = G(6)$. Since SProbit satisfies decomposability,

$$\begin{aligned} \text{SProbit}((a, d, f) \mid A, o) &= (1 - G(9)) \cdot G(6)^2 \\ &< G(9) \cdot (1 - G(6))^2 \\ &= \text{SProbit}((b, c, e) \mid A, o). \end{aligned}$$

This violates monotonicity, since (a, d, f) has a higher outcome $o(a, d, f) = o_1(a) + o_2(d) + o_3(f) = 12$, while $o(b, c, e) = 9$. This again traces back to the light tails of the Gaussian: the probit weights decrease too sharply, leading to a reversal of monotonicity when the rule is extended to product menus in a decomposable way.

The results above demonstrate that widely used rules, such as probit, violate our axioms even in very simple menus.

6 Utility Representations for Abstract Outcome Spaces \mathcal{O}

So far, we have focused on the outcome spaces $\mathcal{O} = \mathbb{R}$ and more generally $\mathcal{O} = \mathbb{R}^n$, and have studied IID and decomposability with respect to addition on these spaces. In the case of $\mathcal{O} = \mathbb{R}$, this is a direct assumption on how payoffs affect behavior. In this section we take a more abstract approach that avoids using payoffs as a primitive, elucidating the content of our assumptions. For simplicity, we focus in this section on decomposable rules.

Let \mathcal{O} be a metrizable topological space of outcomes, endowed with a partial order \leq . One example is, of course, \mathbb{R}^n with the usual partial order. Another example is $\mathcal{O} = \{\text{bounded continuous } f: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}\}$, which can represent a decision-maker who cares about infinite payoff streams and compares them pointwise. The notions of a menu (A, o) , a collection of menus \mathcal{M} , and a stochastic choice rule Φ extend straightforwardly, as do continuity and monotonicity (with respect to \leq).

To define product menus, we endow the outcome space \mathcal{O} with a binary operation $*$ corresponding to combining outcomes in a separable way. To build intuition, consider $\mathcal{O} = \mathbb{R} \times \mathbb{R}_{\geq 0}$, where the first component is interpreted as the mean and the second component as the standard deviation of a stochastic Gaussian monetary reward. In this setting, product menus correspond to combined choices in which the stochastic payoffs are independent of each other, so that the operation $*$ is given by

$$\begin{pmatrix} m_1 \\ \sigma_1 \end{pmatrix} * \begin{pmatrix} m_2 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} m_1 + m_2 \\ \sqrt{\sigma_1^2 + \sigma_2^2} \end{pmatrix}. \quad (9)$$

This structure of outcome space \mathcal{O} implies that choice probabilities are determined by the expectation and standard deviation. The IID axiom then implies that choice probabilities are not affected by additional decisions that yield independent outcomes. This captures a common assumption reflecting invariance to wealth effects or background risk.

Formally, given two menus (A_1, o_1) and (A_2, o_2) , we define their product $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ by

$$A = A_1 \times A_2, \quad \text{and} \quad o(a_1, a_2) = o_1(a_1) * o_2(a_2).$$

A rule Φ is decomposable if

$$\Phi((a_1, a_2) \mid A, o) = \Phi(a_1 \mid A_1, o_1) \cdot \Phi(a_2 \mid A_2, o_2)$$

for all $(A_1, o_1), (A_2, o_2) \in \mathcal{M}$. Decomposability is a separability assumption of independent behavior in product menus.

The only requirement that we impose on $*$ is the existence of an irrelevant outcome: there exists $e \in \mathcal{O}$ such that $e * x = x * e = x$ for any $x \in \mathcal{O}$. The decision-maker does not care about an action a having outcome e in the sense that combining a with any other action b with outcome x does not change the decision-maker's perception of b . The operation $*$ defined in (9) admits the irrelevant outcome $(0, 0)$. It is also commutative and associative, but in general we assume neither commutativity nor associativity.

For simplicity, we will make an additional assumption of positivity in this section:

Axiom 5. *A rule Φ is **positive** if for any menu (A, o) in \mathcal{M} and any $a \in A$ it holds that $\Phi(a \mid A, o) > 0$.*

The question that we ask in this general setting is: what are the monotone, continuous, positive, decomposable rules?

Before stating our main result of this section we will need an additional definition. A function $u: \mathcal{O} \rightarrow \mathbb{R}$ is called a *utility representation* of $(\mathcal{O}, *)$ if

$$u(x * y) = u(x) + u(y) \quad \text{for all } x, y \in \mathcal{O}. \quad (10)$$

In other words, a utility representation assigns a numerical value to each outcome so that combining outcomes in a separable way corresponds to summing their utilities. We define monotonicity and continuity of a utility representation with respect to the topology and partial order on the outcome space.

For example, when $(\mathcal{O}, *) = (\mathbb{R}^n, +)$, any linear function $u(x) = \beta \cdot x$ is a utility representation. Moreover, all continuous utility representations take this form, which follows from a fact about the Cauchy functional equation $u(x) + u(y) = u(x + y)$.¹¹ Monotonicity then pins down the signs of the coefficients β_i : under the standard partial order, it requires $\beta_i \geq 0$. Similarly, when $(\mathcal{O}, *)$ is the Gaussian outcome space (9), the function $u(m, \sigma) = \beta m + \gamma \sigma^2$ is a utility representation. This functional form again exhausts all continuous representations, and the natural monotonicity requirement, which reflects that lower-variance outcomes are chosen more frequently, yields $\gamma \leq 0$.

Our usage of the term utility representation for functions $u: \mathcal{O} \rightarrow \mathbb{R}$ satisfying (10) can be motivated by the well-established connection between cardinal utility values and separability: since we think of $x * y$ as a separable combination of

¹¹For $(\mathcal{O}, *) = (\mathbb{R}, +)$, this dates back to Cauchy; for a general Banach space, see, e.g., [Kuczma \(2009\)](#).

outcomes x and y , the identity (10) is the respective separability condition on utility u . This interpretation relates our analysis to the characterization of separable utility by [Debreu \(1959\)](#), who shows that choice independence across dimensions pins down a separable utility (uniquely up to affine transformations).

In a similar vein, our second main result relates decomposability and utility representations.

Theorem 2. *Let Φ satisfy monotonicity, continuity, positivity and decomposability for $(\mathcal{O}, *, \leq)$. Then there exists a continuous, monotone utility representation u of $(\mathcal{O}, *)$ such that*

$$\Phi(a \mid A, o) = \frac{\exp\left(u(o(a))\right)}{\sum_{b \in A} \exp\left(u(o(b))\right)} \quad (11)$$

for any menu $(A, o) \in \mathcal{M}$.

Informally, [Theorem 2](#) says that there is a canonical way to assign utilities to elements of the outcome space so that choices are governed by a multinomial logit rule with respect to these utilities; moreover, summing these utilities corresponds to combining outcomes in a separable way. The logit scale parameter is normalized to one, as any positive scale factor can always be absorbed by u .

[Theorem 2](#) is a generalization of [Proposition 2](#), which shows a similar result for the case of $(\mathcal{O}, *) = (\mathbb{R}, +)$. Furthermore, for $(\mathcal{O}, *) = (\mathbb{R}^n, +)$, all continuous utility representations are of the form $\beta \cdot x$, and so [Theorem 2](#) is aligned with [Theorem 1](#) and [Proposition 2](#). The conceptual conclusion we wish to draw from [Theorem 2](#) is that our main results, characterizing linear logit-type models, hinge on the separability properties of the decomposability and IID axioms, and do not require strong assumptions such as the additive structure of outcomes in \mathbb{R}^n .

7 Proof of [Theorem 1](#) for $\mathcal{O} = \mathbb{R}$

The remainder of this paper is a proof of our main theorem, which we hope is of technical interest in its own right. We prove the case of $\mathcal{O} = \mathbb{R}$ here and extend to \mathbb{R}^n in [Appendix C](#). We first describe a classical result of De Finetti regarding partially exchangeable processes. Then, using IID, we extend our stochastic choice rules to infinite products of menus, to which we apply De Finetti's Theorem. Monotonicity ensures the partial exchangeability of a process defined by the extended choice rule. The mixture of i.i.d. conclusion of De Finetti's theorem then implies that the choice rule is a mixture of multinomial logit rules.

7.1 De Finetti's Theorem for Partially Exchangeable Processes

An important tool in our proof is the De Finetti Theorem for partially exchangeable processes. Let (X_1, X_2, \dots) be a sequence of random variables, each taking values in some finite set. We denote by \mathcal{T} the tail sigma-algebra of this sequence, i.e., the collection of all events that depend only on the values of X_i for large enough i and are unaffected by modifications to any finite prefix. Let $(X_{i_1}, X_{i_2}, \dots)$ be a subsequence. We say that this subsequence is exchangeable if its joint distribution is invariant to any finite permutation of the coordinates.

Suppose that $\mathbb{N} = \{1, 2, \dots\}$ can be written as a disjoint union of infinite sets $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \dots$. We say that (X_1, X_2, \dots) is partially exchangeable with respect to this partition if its law is invariant under every finite permutation keeping each N_k intact.

The following is a classical result due to [De Finetti \(1980\)](#). It is a generalization of the well-known De Finetti Theorem for exchangeable processes.

Theorem 3 (De Finetti). *Let (X_1, X_2, \dots) be a partially exchangeable process, witnessed by the partition $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \dots$. Then for each k there exists a tail-measurable random variable F_k such that, conditioned on the tail sigma-algebra, it holds that (i) the random variables (X_1, X_2, \dots) are independent, and (ii) for $i \in N_k$ the random variable X_i has distribution F_k .*

The random distributions F_k are the empirical measures. That is, if $\{i_1, i_2, \dots\}$ is an enumeration of N_k , then

$$F_k(x) = \lim_n \frac{1}{n} \sum_{m=1}^n 1_{X_{i_m}=x}.$$

7.2 Extending Φ to Infinite Products

Let \mathcal{M}_X denote the subset of menus with outcomes in $X \subseteq \mathbb{R}$ and let $\mathcal{M} = \mathcal{M}_{\mathbb{R}}$. Since \mathcal{M} is closed under \otimes , a stochastic choice rule Φ is defined for any finite product of menus. The following proposition shows that there is a unique way to extend Φ to countable products of menus taking the form $(A_1, u_1) \otimes (A_2, u_2) \otimes \dots$. We denote the set of all countable products of menus by \mathcal{M}^∞ .¹²

The next proposition shows that when a rule satisfies monotonicity and IID, we can extend it to countable products of menus. Moreover, this extension satisfies partial exchangeability.

¹²Formally, we may identify \mathcal{M}^∞ with the set of sequences in \mathcal{M} . We only refer to infinite products of menus for notational convenience.

Let $(A_1, o_1), (A_2, o_2), \dots$ be a sequence of menus, and let $M = (A_1, o_1) \otimes (A_2, o_2) \otimes \dots \in \mathcal{M}^\infty$ be their product. Denote by $\Omega = \prod_{i=1}^\infty A_i$ the set of sequences (a_1, a_2, \dots) corresponding to a choice in each of the menus. For a finite product $(A, o) = (A_1, o_1) \otimes \dots \otimes (A_n, o_n)$, a probability measure $\Phi(A, o)$ over $\prod_{i=1}^n A_i$ describes the probability of each choice. We will extend Φ to assign to the infinite product menu M a probability distribution over Ω . Given such a measure, denote by (X_1, X_2, \dots) the random variables corresponding to the choice in each sub-menu, that is, $X_i(a_1, a_2, \dots) = a_i$.

Proposition 3. *If Φ satisfies monotonicity and IID, then there is a unique Ψ defined on \mathcal{M}^∞ such that for every $M = (A_1, o_1) \otimes (A_2, o_2) \otimes \dots \in \mathcal{M}^\infty$, $\Psi(M)$ is a probability measure on $\Omega = \prod_{i=1}^\infty A_i$ satisfying*

$$\Psi(M) \left(\left\{ a \in \Omega \mid a_1 = b_1, a_2 = b_2, \dots, a_n = b_n \right\} \right) = \Phi \left((b_1, b_2, \dots, b_n) \left| \bigotimes_{i=1}^n (A_i, o_i) \right. \right)$$

for all n and all (b_1, b_2, \dots, b_n) . Moreover, if \mathbb{N} is partitioned into classes on which the menus (A_i, o_i) are identical, then (X_1, X_2, \dots) is partially exchangeable with respect to this partition.

Proof. Equip $\Omega = \prod_{i=1}^\infty A_i$ with the product topology, under which it is compact. Note that the sets of the form $B_{b_1, \dots, b_n} = \{a \in \Omega \mid a_1 = b_1, a_2 = b_2, \dots, a_n = b_n\}$ are clopen. In particular, given $n \geq 1$ and a probability measure μ_n on $\prod_{i=1}^n A_i$, denote by $\mathcal{P}(\mu_n)$ the set of probability measures on Ω that agree with μ_n on sets of the form B_{b_1, \dots, b_n} :

$$\mathcal{P}(\mu_n) = \left\{ \mu : \mu(B_{b_1, \dots, b_n}) = \mu_n(\{(b_1, \dots, b_n)\}) \right\}.$$

Then $\mathcal{P}(\mu_n)$ is a compact subset of the probability measures on Ω . It is also easily seen to be nonempty.

Suppose that Φ satisfies IID. Denote by μ_n the measure on $\prod_{i=1}^n A_i$ given by $\Phi(\bigotimes_{i=1}^n (A_i, o_i))$. By IID, μ_{n+1} agrees with μ_n on B_{b_1, \dots, b_n} , so that $\mathcal{P}(\mu_{n+1}) \subseteq \mathcal{P}(\mu_n)$. Since these sets are compact and non-empty, their intersection is non-empty, and so there exists a probability measure μ on Ω that agrees with μ_n on B_{b_1, \dots, b_n} . Since the latter sets generate the sigma-algebra, the measure μ is unique: $\{\mu\} = \bigcap_i \mathcal{P}(\mu_n)$.

Finally, suppose that \mathbb{N} is partitioned as $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \dots$, and that for each block N_k , $(A_{i_1}, o_{i_1}) = (A_{i_2}, o_{i_2}) = \dots$ for indices i_1, i_2, \dots from N_k . We want to show that the law of (X_1, X_2, \dots) is invariant under every finite permutation that maps each block N_k into itself. To this end, it suffices to show, without loss of generality, that if we permute X_{i_1} and X_{i_2} then the joint distribution of any long

enough prefix (X_1, X_2, \dots, X_n) of the entire sequence remains unchanged. This follows from the monotonicity of Φ , since the joint distribution of (X_1, X_2, \dots, X_n) is given by $\Phi(\otimes_{i=1}^n (A_i, o_i))$, and by monotonicity, this distribution assigns equal probabilities to sequences yielding the same total payoff, and payoffs are preserved by permuting copies of the same menu, such as X_{i_1} and X_{i_2} , as long as n is larger than both i_1 and i_2 . \square

7.3 Random Coefficients Logit from De Finetti

The next proposition is the heart of the proof of Theorem 1. Recall that the linear multinomial logit rule is given by

$$\text{MNL}^\beta(a \mid A, o) = \frac{\exp(\beta \cdot o(a))}{\sum_{b \in A} \exp(\beta \cdot o(b))}$$

for $\beta \in \mathbb{R}$. We extend this definition to $\beta \in \{-\infty, +\infty\}$ by letting $\text{MNL}^{+\infty}$ be the rule in which the decision maker chooses uniformly at random one of the actions with the highest outcomes. Likewise, $\text{MNL}^{-\infty}$ is the rule in which the decision maker chooses uniformly at random one of the actions with the lowest outcomes.

We accordingly extend random coefficients logit to allow the random parameter β to be chosen from a distribution μ over the extended reals.

Proposition 4. *If Φ satisfies monotonicity and IID, then there is a random coefficients logit rule RCL^μ with μ supported on $[0, \infty]$ such that $\Phi|_{\mathcal{M}_{\mathbb{Z}}} = \text{RCL}^\mu|_{\mathcal{M}_{\mathbb{Z}}}$.*

Proof. Let $(B_1, u_1), (B_2, u_2), \dots$ be an enumeration of $\mathcal{M}_{\mathbb{Z}}$, the set of all the menus with integer outcomes. Note that $\mathcal{M}_{\mathbb{Z}}$ is countable, because the set of actions \mathcal{A} is countable. We choose

$$(B_0, u_0) = \left\{ \begin{array}{cc} \ell & h \\ 0 & 1 \end{array} \right\}.$$

Define the sequence of menus $(A_1, o_1), (A_2, o_2), \dots$ as follows. Write the natural numbers $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \dots$ as a disjoint union of infinite sets, and given $i \in N_k$, set $(A_i, o_i) = (B_k, u_k)$. Hence, there are infinitely many copies of each (B_k, u_k) in the sequence $(A_i, o_i)_i$.

By Proposition 3, there is a unique Ψ defined on \mathcal{M}^∞ that marginalizes to Φ on each finite product of menus. Recall that $\Omega = \prod_{i=1}^\infty A_i$. Let \mathbb{P} denote the probability measure on Ω given by $\Psi(\otimes_{i=1}^\infty (A_i, o_i))$, and let X_n be the coordinate projections, i.e., $X_n(a_1, a_2, \dots) = a_n$, with the tail σ -algebra of (X_1, X_2, \dots) denoted by \mathcal{T} .

Fix any $k \geq 1$. Enumerate $N_k = \{i_1, i_2, \dots\}$ and let

$$(Y_1, Y_2, Y_3, \dots) = (X_{i_1}, X_{i_2}, X_{i_3}, \dots).$$

Enumerate $N_0 = \{j_1, j_2, \dots\}$ and let

$$(Z_1, Z_2, Z_3, \dots) = (X_{j_1}, X_{j_2}, X_{j_3}, \dots).$$

Hence Y_i corresponds to the choice in the i th copy of B_k and Z_i to the i th copy of B_0 .

By Proposition 3, the entire process (X_1, X_2, \dots) is partially exchangeable with respect to the partition $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \dots$. It therefore follows by De Finetti (Theorem 3) that there are tail-measurable random distributions F and G , where F is a distribution over B_k and G is a distribution over $B_0 = \{\ell, h\}$, and such that conditioned on the tail we have that (X_1, X_2, \dots) are independent, with Y_i chosen from F , and Z_i chosen from G , so that

$$F(b_k) = \mathbb{P}[Y_i = b_k \mid \mathcal{T}] \quad \text{and} \quad G(b_0) = \mathbb{P}[Z_i = b_0 \mid \mathcal{T}], \quad \text{for any } i. \quad (12)$$

The distributions F and G are therefore the (random) empirical distributions of the actions, i.e.,

$$F(b_k) = \lim_n \frac{1}{n} \sum_{i=1}^n 1_{Y_i = b_k} \quad \text{and} \quad G(b_0) = \lim_n \frac{1}{n} \sum_{i=1}^n 1_{Z_i = b_0}.$$

Denote

$$\beta = \log \frac{G(h)}{G(\ell)},$$

and note that β is a random variable taking values in $[-\infty, +\infty]$. Let μ be the distribution of β . We claim that $\Phi(B_k, u_k)$ is equal to $\text{RCL}^\mu(B_k, u_k)$. Since k is arbitrary, and since $(B_k, u_k)_k$ enumerate $\mathcal{M}_{\mathbb{Z}}$, showing this will complete the proof.

Choose any $a, a' \in B_k$ such that $u_k(a) \geq u_k(a')$ and let $d = u_k(a) - u_k(a')$. By (12),

$$\mathbb{E}[F(a) \cdot G(\ell)^d] = \mathbb{E}[\mathbb{P}[Y_1 = a \mid \mathcal{T}] \cdot \mathbb{P}[Z_1 = \ell \mid \mathcal{T}] \cdots \mathbb{P}[Z_d = \ell \mid \mathcal{T}]].$$

Since Y_i and Z_i are independent conditioned on the tail,

$$\mathbb{E}[F(a) \cdot G(\ell)^d] = \mathbb{E}[\mathbb{P}[Y_1 = a, Z_1 = \ell, \dots, Z_d = \ell \mid \mathcal{T}]],$$

and so by the law of total expectation

$$\mathbb{E}[F(a) \cdot G(\ell)^d] = \mathbb{P}[Y_1 = a, Z_1 = \ell, \dots, Z_d = \ell].$$

Note that the probability on the right hand side is the probability, under Ψ , of choosing a in the first copy of (B_k, u_k) , and choosing ℓ in the first d copies of (B_0, u_0) . Since Ψ agrees with Φ on finite products, and since Φ is monotone, it follows that this probability is invariant to changing the choices to another set that yields the same total payoff. Hence, since $d = u_k(a) - u_k(a')$,

$$\mathbb{E}[F(a) \cdot G(\ell)^d] = \mathbb{P}[Y_1 = a', Z_1 = h, \dots, Z_d = h].$$

By the same argument used above, we have that the right hand side is equal to $\mathbb{E}[F(a') \cdot G(h)^d]$, and so we have shown that

$$\mathbb{E}[F(a) \cdot G(\ell)^d] = \mathbb{E}[F(a') \cdot G(h)^d].$$

We will need to show a stronger version of this equality. In particular, let P be monomial in $F(a), F(a'), G(\ell), G(h)$. Then we claim that

$$\mathbb{E}[F(a) \cdot G(\ell)^d \cdot P] = \mathbb{E}[F(a') \cdot G(h)^d \cdot P]. \quad (13)$$

For example, when $P = F(a')G(\ell)$, then, following the argument above,

$$\begin{aligned} \mathbb{E}[F(a) \cdot G(\ell)^d \cdot P] &= \mathbb{P}[Y_1 = a, Z_1 = \ell, \dots, Z_d = \ell, Y_2 = a', Z_{d+1} = \ell] \\ &= \mathbb{P}[Y_1 = a', Z_1 = h, \dots, Z_d = h, Y_2 = a', Z_{d+1} = \ell] \\ &= \mathbb{E}[F(a') \cdot G(h)^d \cdot P]. \end{aligned}$$

The general case follows the same idea, introducing an event of the form $Y_i = a, Y_i = a', Z_i = \ell$ or $Z_i = h$ for each term in the monomial, using distinct indices i each time.

By the linearity of expectation, we have that (13) holds for any polynomial P . Thus, taking $P = F(a)G(\ell)^d - F(a')G(h)^d$, we have that $\mathbb{E}[P^2] = 0$, so that

$$F(a)G(\ell)^d = F(a')G(h)^d \text{ almost surely.} \quad (14)$$

Note that if $d = 0$ then this proof yields that $F(a) = F(a')$ almost surely. Otherwise, $d = u_k(a) - u_k(a') > 0$. Let E_h be the event that $G(h) = 0$, which is the event that $\beta = -\infty$. It follows from (14) that $F(a) = 0$ on E_h . Since a and a' are an arbitrary pair such that $u_k(a) > u_k(a')$, we have $F(b) = 0$ for any b that does not yield lowest payoff. Likewise, we have $F(c) = 0$ for any c that does not

yield the highest payoff conditioned on the event E_ℓ where $G(\ell) = 0$ and $\beta = +\infty$. We thus have that conditioned on $\beta = +\infty$, $F = \text{MNL}^{+\infty}(B_k, u_k)$, and likewise conditioned on $\beta = -\infty$, $F = \text{MNL}^{-\infty}(B_k, u_k)$.

Outside the union of E_h and E_ℓ , β is finite, and it follows from (14) that

$$\frac{F(a)}{F(a')} = e^{\beta(u_k(a) - u_k(a'))} = \frac{e^{\beta u_k(a)}}{e^{\beta u_k(a')}}.$$

Hence, also on the event $(E_\ell \cup E_h)^c$ we have that $F = \text{MNL}^\beta(B_k, u_k)$. We have thus shown that $F = \text{MNL}^\beta(B_k, u_k)$, and so

$$\Phi(a \mid B_k, u_k) = \mathbb{P}[Y_i = a] = \mathbb{E}[F(a)] = \mathbb{E}[\text{MNL}^\beta(a \mid B_k, u_k)] = \text{RCL}^\mu(a \mid B_k, u_k).$$

By Lemma 1 from Appendix A, monotonicity implies that μ is supported on $[0, \infty]$. \square

7.4 Final Steps

In the next proposition, we show that the mixing measure of a random coefficients logit rule is uniquely identified by the restriction to $\mathcal{M}_{\mathbb{Z}}$.

Proposition 5. *Suppose $\text{RCL}^\nu|_{\mathcal{M}_{\mathbb{Z}}} = \text{RCL}^{\nu'}|_{\mathcal{M}_{\mathbb{Z}}}$ with ν and ν' supported on $[0, \infty]$. Then $\nu = \nu'$.*

Proof. Given $\gamma > 0$, define $N_m = \lfloor e^{\gamma m} \rfloor$ and let (A_m, o_m^γ) be the menu with $A_m = \{a_1, \dots, a_{N_m}, b\}$, $o_m^\gamma(a_j) = 0$ for $j = 1, \dots, N_m$, and $o_m^\gamma(b) = m$. Then

$$\lim_m \text{MNL}^\beta(b \mid A_m, o_m^\gamma) = \mathbf{1}_{\{\beta > \gamma\}} + \frac{1}{2} \mathbf{1}_{\{\beta = \gamma\}}.$$

Hence, by dominated convergence,

$$\lim_m \text{RCL}^\nu(b \mid A_m, o_m^\gamma) = \nu((\gamma, \infty]) + \frac{1}{2} \nu(\{\gamma\}).$$

The same identity holds for ν' . Since the two rules agree on $\mathcal{M}_{\mathbb{Z}}$, we get $\nu((\gamma, \infty]) = \nu'((\gamma, \infty])$ at every γ that is not an atom of either ν or ν' . Since such γ 's are dense, the intervals $(\gamma, \infty]$ generate all the Borel sets of $[0, \infty]$. Thus $\nu = \nu'$. \square

Given Propositions 4 and 5, we are ready to prove our main theorem.

Proof of Theorem 1 for $\mathcal{O} = \mathbb{R}$. We first show that Φ coincides with some RCL^μ on $\mathcal{M}_{\mathbb{Q}}$. By Proposition 4 there is some μ such that Φ restricted to $\mathcal{M}_{\mathbb{Z}}$ coincides with RCL^μ .

Define for $k = 1, 2, \dots$, the rules Φ^k by

$$\Phi^k(A, o) = \Phi\left(A, \frac{1}{k} \cdot o\right). \quad (15)$$

Note that such rules satisfy monotonicity and IID, since Φ does. Hence, by Proposition 4 again, each $\Phi^k|_{\mathcal{M}_{\mathbb{Z}}} = \text{RCL}^{\mu_k}|_{\mathcal{M}_{\mathbb{Z}}}$ for some μ_k . By (15) and Proposition 5, it follows that μ is equal to the push-forward of μ_k under the map $\beta \rightarrow k\beta$.

For $(A, o) \in \mathcal{M}_{\mathbb{Q}}$, there is a positive integer k such that $(A, k \cdot o) \in \mathcal{M}_{\mathbb{Z}}$. Thus,

$$\Phi(A, o) = \Phi^k(A, k \cdot o) = \text{RCL}^{\mu_k}(A, k \cdot o) = \text{RCL}^{\mu}(A, o).$$

Thus, $\Phi|_{\mathcal{M}_{\mathbb{Q}}} = \text{RCL}^{\mu}|_{\mathcal{M}_{\mathbb{Q}}}$. By Lemma 1, μ is supported on the non-negative extended reals. By continuity, $\mu(\{+\infty\}) = 0$. Indeed, consider menus (A, o_n) where $A = \{a, b\}$, $o_n(a) = 0$, and $o_n(b) = \frac{1}{n}$. By continuity and monotonicity $\lim_n \Phi(a | A, o_n) = \frac{1}{2}$. If, however, $\mu(\{+\infty\}) = \varepsilon > 0$, then $\lim_n \Phi(a | A, o_n) \leq \frac{1}{2}(1 - \varepsilon)$, violating continuity.

Fix any menu (A, o) and $a \in A$. For $k = 1, 2, \dots$, define $\bar{o}_k : A \rightarrow \mathbb{Q}$ by $\bar{o}_k(a) = \frac{1}{k}[k \cdot o(a)]$. Since $\bar{o}_k(a) \rightarrow o(a)$, by continuity, we have $\Phi(A, o) = \lim_k \Phi(A, \bar{o}_k)$. Hence, Φ is uniquely determined by $\Phi|_{\mathcal{M}_{\mathbb{Q}}}$. Since $\Phi|_{\mathcal{M}_{\mathbb{Q}}} = \text{RCL}^{\mu}|_{\mathcal{M}_{\mathbb{Q}}}$, and since RCL^{μ} is continuous, it follows that $\Phi = \text{RCL}^{\mu}$. \square

8 Conclusion

This paper explores a novel approach to stochastic choice. We model the behavior of an individual across a rich variety of situations, under the key assumption that choices remain consistent when decisions are combined in a separable way. Though often implicit, this assumption underpins the validity of experimental analyses that focus on isolated decision problems. In this paper, the corresponding IID axiom is taken as an assumption. But this assumption can be easily tested in a lab, by providing subjects with different menus yielding carefully selected monetary payoffs, and seeing whether their choice probabilities satisfy (2), the defining property of IID.

While each axiom we impose is quite mild, their combination leads to a strong conclusion: choice behavior must follow a random coefficients logit rule with a mixing measure fixed across all contexts. This conclusion is robust to changes in the outcome space and, plausibly, to other features of the model. While we focus on exact adherence to the axioms, real decision-makers only satisfy them approximately. We believe our framework can be extended to accommodate such

deviations—approximate equality in the definition of IID or approximate inequality in the definition of monotonicity—to yield approximate random coefficients logit.

In our approach, a choice rule predicts behavior in any decision problem, modeled as a finite collection of vectors interpreted as observable covariates and associated with action labels. As is standard in full-domain characterizations, the axioms are imposed on the entire domain. The practical counterpart of this conceptual exercise is that the analyst assumes the decision maker could in principle face a rich class of menus, and that behavior across these menus would respect separability. This assumption may be motivated by observations from a limited set of decisions, but complete verification would require observing behavior on the whole domain. Once this assumption is adopted, the analyst is naturally led to the random coefficients logit rule, which can then be estimated using a smaller subset of menus. Whether a variant of our axiomatic approach can deliver a (possibly approximate) characterization on a smaller domain remains a natural question for future work.

References

- D. S. Ahn, F. Echenique, and K. Saito. On path independent stochastic choice. *Theoretical Economics*, 13(1):61–85, 2018.
- S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.
- H. D. Block and J. Marschak. Random orderings and stochastic theories of response. 1959.
- Y. Breitmoser. An axiomatic foundation of conditional logit. *Economic Theory*, 72(1):245–261, 2021.
- S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and A. Rustichini. Axiomatic tests for the Boltzmann distribution. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(1):013406, 2021.
- S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and A. Rustichini. Multinomial logit processes and preference discovery: Inside and outside the black box. *The Review of Economic Studies*, 08 2022. ISSN 0034-6527.

- C. P. Chambers, Y. Masatlioglu, and C. Turansick. Correlated choice. *arXiv preprint arXiv:2103.05084*, 2021.
- W. J. Cho. How to add apples and oranges: Aggregating performances of different nature. *Games and Economic Behavior*, 131:222–244, 2022.
- S. A. Clark. The random utility model with an infinite choice space. *Economic Theory*, 7:179–189, 1996.
- B. De Finetti. On the condition of partial exchangeability. *Studies in Inductive Logic and Probability*, 2:193–205, 1980.
- G. Debreu. Topological methods in cardinal utility theory. *Cowles Foundation Discussion Papers*, 299, 1959.
- G. Debreu. Review of individual choice behavior by RD Luce. *American Economic Review*, 50(1):186–188, 1960.
- F. Echenique, K. Saito, and G. Tserenjigmid. The perception-adjusted Luce model. *Mathematical Social Sciences*, 93:67–76, 2018.
- J.-C. Falmagne. A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18(1):52–72, 1978.
- P. C. Fishburn. Independence in utility theory with whole product sets. *Operations research*, 13(1):28–45, 1965.
- J. T. Fox, K. il Kim, S. P. Ryan, and P. Bajari. The random coefficients logit model is identified. *Journal of Econometrics*, 166(2):204–212, 2012.
- M. Frick, R. Iijima, and T. Strzalecki. Dynamic random utility. *Econometrica*, 87(6):1941–2002, 2019.
- D. Fudenberg and T. Strzalecki. Dynamic logit with choice aversion. *Econometrica*, 83(2):651–691, 2015.
- D. Fudenberg, G. Lanzani, and P. Strack. Limited memory, learning, and stochastic choice. Technical report, Working paper, 2025.
- A. Gandhi and A. Nevo. Empirical models of demand and supply in differentiated products industries. In *Handbook of industrial organization*, volume 4, pages 63–139. Elsevier, 2021.

- F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.
- F. Gul, P. Natenzon, and W. Pesendorfer. Random choice as behavioral optimization. *Econometrica*, 82(5):1873–1912, 2014.
- G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 2nd edition, 1952.
- N. Kashaev, C. Gauthier, and V. H. Aguiar. Dynamic and stochastic rational behavior. *arXiv preprint arXiv:2302.04417*, 2023.
- N. Kashaev, M. Plávala, and V. H. Aguiar. Entangled vs. separable choice. *arXiv preprint arXiv:2403.09045*, 2024.
- M. Kovach and G. Tserenjigmid. Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, 130(9):2411–2461, 2022.
- M. Kuczma. An introduction to the theory of functional equations and inequalities. Birkhäuser Basel, 2009.
- R. Li. Dynamic random choice. *arXiv preprint arXiv:2102.00143*, 2021.
- J. Lu and K. Saito. Mixed logit versus pure characteristics models. 2025.
- R. D. Luce. *Individual choice behavior*, volume 4. Wiley New York, 1959.
- R. D. Luce and P. Suppes. Preferences, utility and subjective probability. *Handbook of Mathematical Psychology*, pages 249–410, 1965.
- C. F. Manski. The structure of random utility models. *Theory and decision*, 8(3):229, 1977.
- F. Matějka and A. McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298, 2015.
- L.-G. Mattsson and J. W. Weibull. Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior*, 41(1):61–78, 2002.
- D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142, 1974.

- D. McFadden. Econometric models of probabilistic choice. In C. F. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. The MIT Press, Cambridge, MA, 1981.
- D. McFadden and M. K. Richter. Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, Westview Press: Boulder, CO, pages 161–186, 1990.
- D. McFadden and K. Train. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470, 2000.
- H. R. Moon, M. Shum, and M. Weidner. Estimation of random coefficients logit demand models with interactive fixed effects. *Journal of Econometrics*, 206(2): 613–644, 2018.
- A. Nevo. A practitioner’s guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548, 2000.
- J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2): 122–136, 1964.
- K. Saito. Axiomatizations of the mixed logit model. 2018.
- F. Sandomirskiy, P. H. Sung, O. Tamuz, and B. Wincelberg. Monotonicity and bracketing in games. *arXiv preprint arXiv:2502.11243*, 2025.
- U. Segal. Additively separable representations on non-convex sets. *Journal of Economic Theory*, 56(1):89–99, 1992.
- C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.
- J. Steiner, C. Stewart, and F. Matějka. Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 85(2):521–553, 2017.
- T. Strzalecki. Stochastic choice theory. *Cambridge Books*, 2025.
- K. E. Train. *Mixed Logit*, page 138–154. Cambridge University Press, 2003.
- P. P. Wakker. *Additive representations of preferences: A new foundation of decision analysis*. Springer Science & Business Media, 2013.
- M. Woodford. Stochastic choice: An optimizing neuroeconomic model. *American Economic Review*, 104(5):495–500, 2014.

J. I. Yellott. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2):109–144, 1977.

A Monotonicity of Random Coefficients Logit

Recall that an extended random coefficients logit is given by RCL^μ for some measure μ on the extended reals. The following lemma shows that in extended random coefficients logit the distribution of the random response parameter must be supported on the non-negative extended reals in order to achieve monotonicity.

Lemma 1. *The extended random coefficients logit rule with random parameter $\beta \sim \mu$ satisfies monotonicity if and only if it satisfies monotonicity for menus in $\mathcal{M}_{\mathbb{Z}}$ if and only if $\mu([-\infty, 0)) = 0$.*

Proof of Lemma 1. Define, for $k = 1, 2, \dots$, $(A, o_k) = (\{a, b, c\}, (0, 1, k))$, and suppose, for the sake of contradiction, that $\mu([-\infty, 0)) > 0$. Then

$$\begin{aligned} \lim_k \Phi(A, o_k)(a) &= \frac{1}{3}\mu(0) + \lim_k \int_{[-\infty, 0)} \frac{1}{1 + e^\beta + e^{\beta \cdot k}} d\mu(\beta) \\ &= \frac{1}{3}\mu(0) + \int_{[-\infty, 0)} \frac{1}{1 + e^\beta} d\mu(\beta) \\ &> \frac{1}{3}\mu(0) + \int_{[-\infty, 0)} \frac{e^\beta}{1 + e^\beta} d\mu(\beta) = \lim_k \Phi(A, o_k)(b), \end{aligned}$$

so monotonicity is violated for large enough k . On the other hand, when $\mu([-\infty, 0)) = 0$, it is clear that RCL^μ satisfies monotonicity. □

B Proof of Proposition 1

Lemma 2 (See, e.g., [Hardy, Littlewood, and Pólya \(1952\)](#)). *If $f, g: \mathbb{R} \rightarrow \mathbb{R}$ are strictly increasing functions, μ is a probability measure, and f, g , and fg are μ -integrable, then $\int fg d\mu \geq \int f d\mu \cdot \int g d\mu$, with equality if and only if μ is a Dirac measure.*

Lemma 3. *Let (A, o) be a menu, and let*

$$f(\beta) = \sum_{a \in A} o(a) \cdot \text{MNL}^\beta(a \mid A, o).$$

Then $f(\beta)$ is increasing in β , and strictly increasing if A is non-trivial.

Proof. Let $A = \{a_1, \dots, a_n\}$ and denote $u_i = o(a_i)$. Let $d(\beta) = \sum_i e^{\beta u_i}$. Then

$$d'(\beta) = \sum_i u_i \cdot e^{\beta u_i}, \quad d''(\beta) = \sum_i u_i^2 \cdot e^{\beta u_i}.$$

We can rewrite $f(\beta)$ as

$$f(\beta) = \frac{d'(\beta)}{d(\beta)},$$

and

$$f'(\beta) = -\frac{d'(\beta)^2}{d(\beta)^2} + \frac{d''(\beta)}{d(\beta)} = -\frac{1}{d(\beta)^2}(d'(\beta)^2 - d(\beta) \cdot d''(\beta)).$$

By the Cauchy-Schwarz inequality, we have

$$d'(\beta)^2 - d(\beta) \cdot d''(\beta) = \left(\sum_i u_i \cdot e^{\beta u_i} \right)^2 - \left(\sum_i e^{\beta u_i} \right) \cdot \left(\sum_i u_i^2 \cdot e^{\beta u_i} \right) \leq 0,$$

where the equality holds when $u_i \cdot \sqrt{e^{\beta u_i}} = \text{const} \cdot \sqrt{e^{\beta u_i}}$ for all i , i.e., when u_i itself is a constant. \square

Proof of Proposition 1. Let $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ be a product of non-trivial menus and let

$$f_1(\beta) = \sum_{a \in A_1} o_1(a) \cdot \text{MNL}^\beta(a \mid A_1, o_1)$$

and

$$f_2(\beta) = \sum_{a \in A_2} o_2(a) \cdot \text{MNL}^\beta(a \mid A_2, o_2).$$

It follows that

$$\begin{aligned} & \sum_{a \in A_1} \sum_{b \in A_2} o_1(a) \cdot o_2(b) \cdot \text{RCL}^\mu((a, b) \mid A, o) \\ &= \sum_{a \in A_1} \sum_{b \in A_2} o_1(a) \cdot o_2(b) \cdot \int \text{MNL}^\beta((a, b) \mid A, o) \, d\mu(\beta) \\ &= \int \sum_{a \in A_1} \sum_{b \in A_2} o_1(a) \cdot o_2(b) \cdot \text{MNL}^\beta((a, b) \mid A, o) \, d\mu(\beta), \end{aligned}$$

by a change of the order of summation. By the decomposability of MNL^β ,

$$\begin{aligned} &= \int \sum_{a \in A_1} \sum_{b \in A_2} o_1(a) \cdot o_2(b) \cdot \text{MNL}^\beta(a \mid A_1, o_1) \cdot \text{MNL}^\beta(b \mid A_2, o_2) \, d\mu(\beta) \\ &= \int \sum_{a \in A_1} o_1(a) \cdot \text{MNL}^\beta(a \mid A_1, o_1) \sum_{b \in A_2} o_2(b) \cdot \text{MNL}^\beta(b \mid A_2, o_2) \, d\mu(\beta) \\ &= \int f_1(\beta) f_2(\beta) \, d\mu(\beta). \end{aligned}$$

Since f_1 and f_2 are strictly increasing in β (Lemma 3),

$$\begin{aligned} &\geq \left(\int f_1(\beta) d\mu(\beta) \right) \cdot \left(\int f_2(\beta) d\mu(\beta) \right) \\ &= \left(\sum_{a \in A_1} o_1(a) \cdot \text{RCL}^\mu(a \mid A_1, o_1) \right) \cdot \left(\sum_{b \in A_2} o_2(b) \cdot \text{RCL}^\mu(b \mid A_2, o_2) \right). \end{aligned}$$

Moreover, by Lemma 2, the inequality holds with equality if and only if μ is a Dirac measure. □

C Proof of Theorem 1 for $\mathcal{O} = \mathbb{R}^n$ with $n \geq 2$

Proof. Let Φ be a monotone, continuous rule on $\mathcal{M}_{\mathbb{R}^n}$ satisfying IID. Let $\mathcal{B} = \{e_1, \dots, e_n\}$ denote the basis of standard unit vectors of \mathbb{R}^n . For $t = 1, \dots, n$, define the rule Λ_t on $\mathcal{M}_{\mathbb{R}}$ by $\Lambda_t(A, o) = \Phi(A, o \cdot e_t)$, where $o \cdot e_t: A \rightarrow \mathbb{R}^n$ maps a to $o(a) \cdot e_t$.

Since Λ_t satisfies monotonicity, continuity and IID, the one-dimensional case proved in §7 gives $\Lambda_t = \text{RCL}^{\nu_t}$ for some ν_t supported on $\mathbb{R}_{\geq 0}$.

Let $(B_{n+1}, u_{n+1}), (B_{n+2}, u_{n+2}), \dots$ be an enumeration of $\mathcal{M}_{\mathbb{Z}^n}$, the set of all the menus with outcomes in \mathbb{Z}^n . Note that $\mathcal{M}_{\mathbb{Z}^n}$ is countable, because the set of actions \mathcal{A} is countable. For $t = 1, \dots, n$ choose

$$(B_t, u_t) = \left\{ \begin{array}{cc} \ell_t & h_t \\ 0 & e_t \end{array} \right\}. \quad (16)$$

Since $(B_{n+1}, u_{n+1}), (B_{n+2}, u_{n+2}), \dots$ is an enumeration, the menus (16) will appear twice in $(B_1, u_1), (B_2, u_2), \dots$

Define the sequence of menus $(A_1, o_1), (A_2, o_2), \dots$ in $\mathcal{M}_{\mathbb{Z}^n}$ as follows. Write the natural numbers $\mathbb{N} = N_1 \cup N_2 \cup \dots$ as a disjoint union of infinite sets, and given $i \in N_k$, set $(A_i, o_i) = (B_k, u_k)$. Hence, there are infinitely many copies of each (B_k, u_k) in the sequence $(A_i, o_i)_i$. Let $\Omega = \prod_{i=1}^{\infty} A_i$. By Proposition 3, there is a unique Ψ defined on $\mathcal{M}_{\mathbb{R}^n}^{\infty}$ that marginalizes to Φ on each finite product of menus.¹³ Let \mathbb{P} denote the probability measure on Ω given by $\Psi(\otimes_{i=1}^{\infty} (A_i, o_i))$, and let X_n be the coordinate projections, i.e., $X_n(a_1, a_2, \dots) = a_n$, with the tail σ -algebra of (X_1, X_2, \dots) denoted by \mathcal{T} .

Fix any $k > n$. Enumerate $N_k = \{i_1, i_2, \dots\}$ and let

$$(Y_1, Y_2, Y_3, \dots) = (X_{i_1}, X_{i_2}, X_{i_3}, \dots).$$

¹³Formally, Proposition 3 is stated for $\mathcal{M}_{\mathbb{R}}$, but the exact same proof applies to $\mathcal{M}_{\mathbb{R}^n}$.

For $t = 1, \dots, n$, let $N_t = \{j_1^t, j_2^t, \dots\}$. Let

$$(Z_1^t, Z_2^t, Z_3^t, \dots) = (X_{j_1^t}, X_{j_2^t}, X_{j_3^t}, \dots),$$

Hence Y_i corresponds to the choice in the i th copy of B_k and Z_i^t to the i th copy of B_t , as defined in (16).

By Proposition 3, we know that (Y_1, Y_2, \dots) are exchangeable, as are (Z_1^t, Z_2^t, \dots) . It therefore follows by De Finetti (Theorem 3) that there are tail-measurable random distributions F, G_1, \dots, G_n , where F is a distribution over B_k and G_t is a distribution over $B_t = \{\ell_t, h_t\}$, and such that conditioned on the tail we have that (X_1, X_2, \dots) are independent, with Y_i chosen from F , and Z_i^t chosen from G_t , so that

$$F(b_k) = \mathbb{P}[Y_i = b_k | \mathcal{T}] \quad \text{and} \quad G_t(b_t) = \mathbb{P}[Z_i^t = b_t | \mathcal{T}], \quad \text{for any } i. \quad (17)$$

The distributions F, G_1, \dots, G_n are therefore the (random) empirical distributions of the actions, i.e.,

$$F(b_k) = \lim_n \frac{1}{n} \sum_{i=1}^n 1_{Y_i=b_k} \quad \text{and} \quad G_t(b_t) = \lim_n \frac{1}{n} \sum_{i=1}^n 1_{Z_i^t=b_t} \quad \text{a.s.}$$

Denote by β the n -dimensional random variable with coordinates

$$\beta_t = \log \frac{G_t(h_t)}{G_t(\ell_t)},$$

and note that β takes values in $[-\infty, +\infty]^n$. Let μ be the distribution of β and μ_t be the marginal distribution of β_t .

We claim that $\Phi(B_k, u_k)$ is equal to $\text{RCL}^\mu(B_k, u_k)$. To this end, choose any $a, a' \in B_k$, let d_t denote the t -th coordinate of $u_k(a) - u_k(a')$, and let $d_t^+ = \max\{d_t, 0\}$ and $d_t^- = \max\{-d_t, 0\}$. By (17), Y_i and Z_i^1, \dots, Z_i^n are independent conditioned on the tail. By the same argument used in the proof of Proposition 4 we conclude that

$$\begin{aligned} & F(a) \cdot G_1(\ell_1)^{d_1^+} \cdots G_n(\ell_n)^{d_n^+} \cdot G_1(h_1)^{d_1^-} \cdots G_n(h_n)^{d_n^-} \\ &= F(a') \cdot G_1(\ell_1)^{d_1^-} \cdots G_n(\ell_n)^{d_n^-} \cdot G_1(h_1)^{d_1^+} \cdots G_n(h_n)^{d_n^+} \quad \text{almost surely.} \end{aligned} \quad (18)$$

Let \mathbb{R}_t denote the one-dimensional vector space spanned by e_t . Note that for $(B_k, u_k) \in \mathcal{M}_{\mathbb{R}_t}$, $d_i^+ = d_i^- = 0$ for all $i \neq t$. Hence, in this case we have

$$\frac{F(a)}{F(a')} = e^{\beta_t d_t} = \frac{e^{\beta_t u_k(a)_t}}{e^{\beta_t u_k(a')_t}},$$

and we may conclude that $\Lambda_t|_{\mathcal{M}_{\mathbb{Z}}} = \text{RCL}^{\mu_t}|_{\mathcal{M}_{\mathbb{Z}}}$. Thus, by Proposition 5, $\mu_t = \nu_t$, so the support of μ_t is contained in $\mathbb{R}_{\geq 0}$. Thus, β is finite (and F is non-degenerate) almost surely. For any (B_k, u_k) it follows from (18) that

$$\frac{F(a)}{F(a')} = e^{\beta \cdot (u_k(a) - u_k(a'))} = \frac{e^{\beta \cdot u_k(a)}}{e^{\beta \cdot u_k(a')}}.$$

We have thus shown that $F = \text{MNL}^\beta(B_k, u_k)$, and so

$$\Phi(a | B_k, u_k) = \mathbb{P}[Y_i = a] = \mathbb{E}[F(a)] = \mathbb{E}[\text{MNL}^\beta(a | B_k, u_k)] = \text{RCL}^\mu(a | B_k, u_k).$$

We now show that $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}} = \text{RCL}^\mu|_{\mathcal{M}_{\mathbb{Q}^n}}$. As in the proof of the one-dimensional case, we define, for $k = 1, 2, \dots$,

$$\Phi^k(A, o) = \Phi\left(A, \frac{1}{k} \cdot o\right).$$

The integer-outcome argument above applies to each Φ^k , so $\Phi^k|_{\mathcal{M}_{\mathbb{Z}^n}} = \text{RCL}^{\mu_k}|_{\mathcal{M}_{\mathbb{Z}^n}}$ for some measure μ_k on $\mathbb{R}_{\geq 0}^n$. Let $\tilde{\mu}_k$ be the image of μ_k under the map $\beta \rightarrow k\beta$. We claim that $\mu = \tilde{\mu}_k$.

We use the following identification fact generalizing Proposition 5. Let ν, ν' be mixing measures on $\mathbb{R}_{\geq 0}^n$ such that $\text{RCL}^\nu|_{\mathcal{M}_{\mathbb{Z}^n}} = \text{RCL}^{\nu'}|_{\mathcal{M}_{\mathbb{Z}^n}}$. We claim that $\nu = \nu'$. Fix nonempty $S \subseteq \{1, \dots, n\}$ and $\gamma \in \mathbb{R}_{> 0}^S$. For $t \in S$, set $N_{t,m} = \lfloor e^{\gamma_t m} \rfloor$ and consider the menu with $N_{t,m}$ actions of outcome 0 and one action b_t of outcome $m e_t$. In the product of these menus over $t \in S$, the probability of choosing all distinguished actions under MNL^β converges to

$$\prod_{t \in S} \left(\mathbb{1}_{\{\beta_t > \gamma_t\}} + \frac{1}{2} \mathbb{1}_{\{\beta_t = \gamma_t\}} \right).$$

Thus, whenever the coordinates of γ are not atoms of the corresponding one-dimensional marginals,

$$\nu(\{\beta_t > \gamma_t \ \forall t \in S\}) = \nu'(\{\beta_t > \gamma_t \ \forall t \in S\}).$$

Such γ 's are dense. Hence, the corresponding upper cylinders, together with $\mathbb{R}_{\geq 0}^n$, generate the Borel sets of $\mathbb{R}_{\geq 0}^n$.¹⁴

Applying this identification fact with $\nu = \mu$ and $\nu' = \tilde{\mu}_k$, we obtain $\mu = \tilde{\mu}_k$. Now consider $(A, o) \in \mathcal{M}_{\mathbb{Q}^n}$ and choose a positive integer k such that $(A, k \cdot o) \in \mathcal{M}_{\mathbb{Z}^n}$. Thus,

$$\Phi(A, o) = \Phi^k(A, k \cdot o) = \text{RCL}^{\mu_k}(A, k \cdot o) = \text{RCL}^\mu(A, o).$$

¹⁴The sets with $S = \{1, \dots, n\}$ are enough to generate the Borel sigma-algebra of $\mathbb{R}_{> 0}^n$. Smaller S are needed for the faces of $\mathbb{R}_{\geq 0}^n$, where some coordinates are zero.

Thus, $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}} = \text{RCL}^\mu|_{\mathcal{M}_{\mathbb{Q}^n}}$. Fix any menu (A, o) and $a \in A$. For $k = 1, 2, \dots$, define $\bar{o}_k: A \rightarrow \mathbb{Q}^n$ by $\bar{o}_k(a) = \frac{1}{k}[k \cdot o(a)]$, where the ceiling is applied componentwise. Since $\bar{o}_k(a) \rightarrow o(a)$, by continuity, we have $\Phi(A, o) = \lim_k \Phi(A, \bar{o}_k)$. Hence, Φ is uniquely determined by $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}}$. Since $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}} = \text{RCL}^\mu|_{\mathcal{M}_{\mathbb{Q}^n}}$, and since RCL^μ is continuous, it follows that $\Phi = \text{RCL}^\mu$. \square

In some natural applications of stochastic choice, outcomes are limited to a subset of Euclidean space, such as the positive orthant. Our characterization still applies, provided the outcome space is rich enough. Indeed, as the following corollary shows, a stochastic choice rule defined for menus with outcomes in a convex cone that satisfies the axioms can be extended to a rule on \mathbb{R}^n that satisfies the axioms.

Corollary 1. *Let $C \subseteq \mathbb{R}^n$ be a full-dimensional convex cone, and let Φ be a stochastic choice rule on \mathcal{M}_C that satisfies monotonicity, continuity, and IID. Then $\Phi = \text{RCL}^\mu$ for some μ supported on $\mathbb{R}_{\geq 0}^n$.*

Proof. Let C and Φ be as in the corollary. Define Ψ on $\mathcal{M}_{\mathbb{R}^n}$ by $\Psi(A, o) = \Phi(A, o + c)$, where $c \in C$ is such that $o(a) + c \in C$ for all $a \in A$. By IID, this definition does not depend on the choice of c . To see this, suppose c and c' are both admissible shifts. By the cone property, $c + c'$ is also admissible. Applying IID to the product of $(A, o + c)$ with a singleton menu having outcome c' gives $\Phi(A, o + c) = \Phi(A, o + c + c')$, and, similarly, applying IID to the product of $(A, o + c')$ with a singleton menu having outcome c results in $\Phi(A, o + c') = \Phi(A, o + c + c')$. Hence $\Phi(A, o + c) = \Phi(A, o + c')$, so Ψ is well-defined. It is obvious that Ψ is monotone. Ψ is moreover continuous as the same c from the interior of C works for all menus sufficiently close to (A, o) . Finally, for $\Psi(A_1, o_1) = \Phi(A_1, o_1 + c_1)$ and $\Psi(A_2, o_2) = \Phi(A_2, o_2 + c_2)$ and $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$, we have

$$\Psi(a_1 | A, o) = \Phi(a_1 | A, o + c_1 + c_2) = \Phi(a_1 | A_1, o_1 + c_1) = \Psi(a_1 | A_1, o_1),$$

for all $a_1 \in A_1$. Thus Ψ satisfies IID and is a random coefficients logit rule. Since Ψ agrees with Φ on C , Φ is a random coefficients logit rule. \square

D Proof of Proposition 2

Proof. Since decomposability implies IID, by Theorem 1, $\Phi = \text{RCL}^\mu$ for some μ supported on $\mathbb{R}_{\geq 0}^n$. For $k = 1, \dots, n$, define the stochastic choice rule Φ^k for the

outcome space $\mathcal{O} = \mathbb{R}$ by $\Phi^k(a \mid A, o) = \Phi(a \mid A, o_k)$, where $o_k(a) = o(a) \cdot e_k$ where e_k is the standard basis vector whose k th component is 1 and whose other components are 0. It follows that

$$\begin{aligned}\Phi^k(a \mid A, o) &= \int \text{MNL}^\beta(a \mid A, o_k) \, d\mu(\beta) \\ &= \int \text{MNL}^{\beta_k}(a \mid A, o) \, d\mu(\beta) \\ &= \int \text{MNL}^{\beta_k}(a \mid A, o) \, d\mu_k(\beta_k),\end{aligned}$$

where μ_k is the k th marginal of μ . Note that Φ^k inherits decomposability from Φ . By Proposition 1, μ_k is a Dirac measure, since decomposability implies that Φ^k exhibits zero correlation on product menus. Since each marginal of μ is a Dirac measure, μ itself is a Dirac measure, i.e., $\Phi = \text{MNL}^\beta$ for some $\beta \in \mathbb{R}_{\geq 0}^n$. \square

E Proof of Theorem 2

Proof. Recall that e denotes the identity element of $(\mathcal{O}, *)$. For each element $x \in \mathcal{O}$ of the outcome space, fix a menu

$$(A_x, o_x) = \left\{ \begin{array}{cc} a_e & a_x \\ e & x \end{array} \right\}$$

that has two actions with outcomes e and x . Let $p_x = \Phi(a_x \mid A_x, o_x)$ be the probability that the action with outcome x is chosen in this menu.

Define the function $u: \mathcal{O} \rightarrow \mathbb{R}$ by

$$u(x) = \ln \frac{p_x}{1 - p_x}. \quad (19)$$

Note that this logarithm is finite by the positivity of Φ .

We now demonstrate that u is a utility representation of \mathcal{O} , i.e., that it satisfies the generalized Cauchy equation $u(x * y) = u(x) + u(y)$ for all $x, y \in \mathcal{O}$. Consider a product menu

$$(B, o) = \left((A_x, o_x) \otimes (A_y, o_y) \right) \otimes (A_{x*y}, o_{x*y}).$$

Recall that the associativity of $*$ is not assumed, so we must be careful about the order of operations. The constructed product menu contains actions

$$b = ((a_e, a_e), a_{x*y}) \quad \text{and} \quad b' = ((a_x, a_y), a_e).$$

Computing their outcomes, we get

$$o(b) = (e * e) * (x * y) = x * y \quad \text{and} \quad o(b') = (x * y) * e = x * y,$$

where we used the fact that e is both a left and a right identity. Since the outcomes of b and b' are the same, monotonicity of Φ implies

$$\Phi(b \mid B, o) = \Phi(b' \mid B, o).$$

By decomposability of Φ , this identity can be rewritten as follows

$$(1 - p_x) \cdot (1 - p_y) \cdot p_{x*y} = p_x \cdot p_y \cdot (1 - p_{x*y}).$$

Taking the logarithm and using the definition of u , we obtain

$$u(x * y) = u(x) + u(y)$$

and conclude that u is a utility representation of \mathcal{O} . We stress that the proof of this fact uses neither associativity nor commutativity of $*$.

We now consider an arbitrary menu (A, o) and demonstrate that $\Phi(A, o)$ is given by multinomial logit with the constructed utility function (11). Let $a, b \in A$ be two distinct actions. We express the ratio $\Phi(a \mid A, o)/\Phi(b \mid A, o)$ in terms of the constructed utility representation u . Denote the outcomes by $x = o(a)$ and $y = o(b)$ and consider a new menu

$$\left((A_x, o_x) \otimes (A, o) \right) \otimes (A_y, o_y).$$

The two actions

$$\left((a_e, a), a_y \right) \quad \text{and} \quad \left((a_x, b), a_e \right)$$

have equal outcomes. Indeed, $(e * x) * y = x * y$ and $(x * y) * e = x * y$ by the assumption that e is both a left and a right identity element. By monotonicity, Φ assigns equal probabilities to these actions. Applying decomposability, we obtain

$$(1 - p_x) \cdot \Phi(a \mid A, o) \cdot p_y = p_x \cdot \Phi(b \mid A, o) \cdot (1 - p_y).$$

This equality can be rewritten as

$$\frac{\Phi(a \mid A, o)}{\Phi(b \mid A, o)} = \frac{p_x/(1 - p_x)}{p_y/(1 - p_y)}$$

and thus

$$\frac{\Phi(a \mid A, o)}{\Phi(b \mid A, o)} = \frac{\exp\left(u(o(a))\right)}{\exp\left(u(o(b))\right)}.$$

Since a and b were arbitrary, and $\sum_{b \in A} \Phi(b \mid A, o) = 1$, we conclude that

$$\Phi(a \mid A, o) = \frac{\exp\left(u(o(a))\right)}{\sum_{b \in A} \exp\left(u(o(b))\right)}, \quad (20)$$

i.e., Φ is multinomial logit.

To verify that u is monotone, take $x \leq y$ and consider the binary menu with outcomes x and y . By monotonicity of Φ , the action with outcome y is chosen with probability at least as large as the action with outcome x . Using the representation (20), this implies $\exp(u(y)) \geq \exp(u(x))$, and hence $u(y) \geq u(x)$.

It remains to check the equivalence between the continuity of Φ and that of u . If $u: \mathcal{O} \rightarrow \mathbb{R}$ is a continuous utility representation, then Φ given by (20) is continuous since the right-hand side of (20) is a continuous function of the profile of outcomes. In the opposite direction, we suppose that Φ is continuous and show that $u(x_n) \rightarrow u(x)$ for any sequence $x_n \rightarrow x$ in \mathcal{O} . Fix a binary set of actions $A = \{a, b\}$ and consider outcome functions o_n and o given by $o_n(a) = o(a) = x$, $o_n(b) = x_n$, and $o(b) = x$. Thus (A, o_n) converges to the menu with identical outcomes (A, o) . By continuity, $\Phi(A, o_n)$ converges to the uniform distribution $\Phi(A, o)$. By (20), we get

$$\frac{\exp(u(x_n))}{\exp(u(x))} = \frac{\Phi(b | A, o_n)}{\Phi(a | A, o_n)} \rightarrow \frac{\Phi(b | A, o)}{\Phi(a | A, o)} = 1$$

and so $u(x_n)$ converges to $u(x)$. Thus u is continuous. □

F Restrictions of Monotonicity and IID

The family of random coefficients logit rules is parameterized by a probability measure, which is an infinite dimensional object. This gives this family considerable flexibility to model a wide range of behavior. Nevertheless, knowing that a rule belongs to this family imposes significant restrictions on the choice probabilities, even within simple menus. As we discussed in §2.1, for the menu

$$(A, o) = \left\{ \begin{array}{ccc} a_1 & a_2 & a_3 \\ 0 & 2 & 7 \end{array} \right\}$$

it holds for every random coefficients logit rule Φ that $\Phi(a_1 | A, o) = 5\%$ implies that $\Phi(a_3 | A, o) \geq 83.8\%$. Indeed, this lower bound of the choice probability of a_3 turns out to be $\text{MNL}^\beta(a_3 | A, o)$ for the β that satisfies $\text{MNL}^\beta(a_1 | A, o) = 5\%$. The following proposition shows that this is a general property of random coefficients logit rules for $\mathcal{O} = \mathbb{R}$.

Proposition 6. *Let $A = \{a_1, \dots, a_n\}$ and $o(a_1) \leq o(a_2) \leq \dots \leq o(a_n)$ with at least one strict inequality. Let $\beta \in \mathbb{R}$ be the unique parameter satisfying $\text{MNL}^\beta(a_1 |$*

$A, o) = \Phi(a_1 | A, o)$. If Φ is a random coefficients logit rule, then $\Phi(a_n | A, o) \geq \text{MNL}^\beta(a_n | A, o)$.

By a symmetric argument, we can get a lower bound on the probability of a_1 , given the probability of a_n : if γ is such that $\text{MNL}^\gamma(a_n | A, o) = \Phi(a_n | A, o)$, then $\Phi(a_1 | A, o) \geq \text{MNL}^\gamma(a_1 | A, o)$.

Proposition 6 highlights a property of random coefficients logit, independently of our main results. But since our axioms imply random coefficients logit, it follows from this proposition and our main results that every rule that satisfies our axioms also satisfies this property. We do not know a direct proof of this.

Before proving this proposition we state and prove a lemma.

Lemma 4. Let $L(\beta) = \text{MNL}^\beta(a_1 | A, o)$ and $H(\beta) = \text{MNL}^\beta(a_n | A, o)$, where $A = \{a_1, \dots, a_n\}$ and $o(a_1) \leq o(a_2) \leq \dots \leq o(a_n)$, with at least one strict inequality. Then $-\frac{H''}{H'} \leq -\frac{L''}{L'}$.

Proof. Because logit probabilities are invariant to adding a common constant to all outcomes, assume without loss of generality that $o(a_1) = 0$. Let o_i denote $o(a_i)$, and let $d(\beta) = \sum_i e^{\beta o_i}$. Then $0 \leq o_i \leq o_n$ for all i , with $o_n > 0$, and

$$\begin{aligned} d'(\beta) &= \sum_i o_i e^{\beta o_i} < o_n \cdot d(\beta), \\ d''(\beta) &= \sum_i o_i^2 e^{\beta o_i} \leq o_n \cdot d'(\beta). \end{aligned}$$

The derivatives of L are given by:

$$\begin{aligned} L' &= -\frac{d'}{d^2} > -\frac{o_n \cdot d}{d^2} = -o_n \cdot L \\ L'' &= \frac{2(d')^2 - d'' \cdot d}{d^3} \end{aligned}$$

Since $o_1 = 0$, we have $H = e^{\beta o_n} L$. Then

$$\begin{aligned} H' &= o_n e^{\beta o_n} L + e^{\beta o_n} L' = e^{\beta o_n} (o_n L + L') \\ H'' &= o_n^2 e^{\beta o_n} L + 2o_n e^{\beta o_n} L' + e^{\beta o_n} L'' \end{aligned}$$

Thus, we have:

$$\begin{aligned} \frac{H''}{H'} - \frac{L''}{L'} &= \frac{o_n^2 L + 2o_n L' + L''}{o_n L + L'} - \frac{L''}{L'} \\ &= \frac{o_n^2 L \cdot L' + 2o_n (L')^2 - o_n L \cdot L''}{L'(o_n L + L')}. \end{aligned}$$

Since $L' < 0$ and $o_n L + L' > 0$, the denominator of the above expression is negative. Moreover the numerator simplifies to

$$\frac{o_n(d'' - o_n d')}{d^3} \leq 0,$$

so the overall expression is non-negative, as desired. \square

Proof of Proposition 6. Consider $A = \{a_1, \dots, a_n\}$ and outcomes $o(a_1) \leq o(a_2) \leq \dots \leq o(a_n)$ with at least one strict inequality. Without loss of generality, suppose $o(a_1) = 0$. Define $d(\beta) = \sum_i e^{\beta \cdot o(a_i)}$. Let $L(\beta) = \frac{1}{d(\beta)}$ and $H(\beta) = \frac{e^{\beta \cdot o(a_n)}}{d(\beta)}$ denote the logit choice probabilities of the low and high alternatives, respectively, as a function of β .

Denote $p = \Phi(a_1 | A, o)$. Since $L(\beta) = 1/d(\beta)$ is strictly decreasing, the β solving $L(\beta) = p$ is uniquely pinned down by p . Let $F(\beta) = -L(\beta)$, so that H and F are increasing functions.

By Lemma 4,

$$-\frac{H''}{H'} \leq -\frac{L''}{L'} = -\frac{F''}{F'}.$$

Hence, by Pratt (1964), $F \circ H^{-1}$ is increasing and concave, i.e., $L \circ H^{-1}$ is decreasing and convex, and its inverse $H \circ L^{-1}$ is convex. Because Φ is random coefficients logit, there is a probability measure μ such that $\Phi(a_n | A, o) = \int H d\mu$ and $p = \int L d\mu$. For such μ , we have

$$\begin{aligned} \Phi(a_n | A, o) &= \int H d\mu = \int H \circ L^{-1} \circ L d\mu \geq H \circ L^{-1} \left(\int L d\mu \right) \\ &= H \circ L^{-1}(p) = H(\beta) = \text{MNL}^\beta(a_n | A, o). \end{aligned}$$

The inequality follows from Jensen's inequality since $H \circ L^{-1}$ is convex. \square