Independence of Irrelevant Decisions in Stochastic Choice*

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

May 25, 2025

Abstract

We investigate stochasticity in choice behavior across diverse decisions. Each decision is modeled as a menu of actions with associated outcomes, and a stochastic choice rule assigns probabilities to actions based on the outcome profile. We characterize rules whose predictions are not affected by whether or not additional, irrelevant decisions are included in the model. Our main result is that such rules form the parametric family of mixed-logit rules.

1 Introduction

Consider an analyst who observes individuals choosing among different actions, each yielding a given payoff known to the analyst. Although people prefer higher payoffs, in

*A previous version of this paper was circulated under the title "Decomposable 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, Annie Liang, Daniel Litt, Yusufcan Masatlioglu, Jeffrey Mensch, Stephen Morris, Pietro Ortoleva, Ariel Pakes, Tom Palfrey, Marcin Pęski, 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

reality they do not always choose the action that yields the highest payoff, for various reasons, including cognitive limitations, errors in the decision-making process, and random shocks to the perceived payoff. The analyst would like to develop a prediction model to estimate out-of-sample choice probabilities, for example following a policy change or market restructuring. In this paper we use an axiomatic approach to restrict the analyst's set of models, abstracting away from the physical reasons for randomness in choice. We show that mixed logit, a commonly used model (also known as random coefficients logit), is unique in satisfying three simple axioms.

We model a single decision instance as a *menu* consisting of a finite set of actions, each associated with an outcome. For simplicity, we primarily focus on monetary outcomes, represented by real numbers. The actions are treated as mere labels, carrying no intrinsic meaning for the decision maker. This modeling approach abstracts from the structure of the alternatives and focuses instead on the distribution of payoffs. It parallels the random utility model literature, where utilities are primitives. Alternatively, one can view menus as one-player normal form games.

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).

Monotonicity requires that an action that yields a higher payoff is taken with higher probability. This axiom 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 the model predictions.

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 payoff from one choice does not affect the payoff from another. Suppose that an individual has to choose one of the two actions $A_1 = \{a, b\}$, and also one of the two actions $A_2 = \{s, t\}$. We can think of these two choices as a combined choice in a menu with action set $A = \{(a, s), (a, t), (b, s), (b, t)\}$, and an outcome function that assigns a monetary payoff to each action. A simple case is the one in which payoffs are additively separable, i.e., the outcome function is a sum of two functions, one that depends on the first coordinate, and another that depends on the second. In this case we say that the menu is a product menu.

For example, suppose that $\{a, b\}$ are two brands of soap, and $\{s, t\}$ are two brands of milk. Then it may be natural to model the joint problem as a separable one. However, if $\{a, b\}$ are two choices for a suit's jacket and $\{s, t\}$ are two choices for the suit's pants, then the appropriate model will not be a product menu, since there are strong complementarities.

Consider an analyst employed by a soap company who wants to predict the consumer share of each soap brand at a new supermarket, which also sells milk. The analyst could write a model that predicts the share of consumers that purchase each (soap, milk) pair. Because the menu is separable, the analyst could, alternatively, think of the soap decision separately, and predict choices in a menu that only includes the actions $\{a, b\}$, ignoring the consumer purchases of milk. Our IID axiom is the assumption that the predicted choice probabilities for soap will be independent of the two modeling options. Hence, the analyst may ignore the milk choice. Without IID, a model may provide different predictions for soap choices, depending on whether or not the choice of milk was included in the model, and furthermore on whether many other irrelevant choices were included, such as whether the shopper paid with cash or card, or what color shirt they chose to wear in the morning.

Note that IID does not exclude the possibility that these choices are correlated: it is possible that consumers who buy better soap are more likely to buy better milk. However, since there is no complementarity or substitutability between the two products, the menu of milks that consumers face is irrelevant to the menu of soaps. Accordingly, IID requires that a model yield the same predictions for soap, regardless of whether or not the choice of soap is modeled in conjunction with the choice of milk. In a sense, when an analyst decides to exclude an unrelated choice from their model, they are implicitly assuming that IID holds. We thus argue that IID is a reasonable positive assumption on the analyst and the set of possible models they may use to describe decision makers. Of course, for the choice of jacket and pants, where the menu is not a product menu, IID imposes no restrictions.

Our main result is that mixed-logit rules are the only ones that satisfy monotonicity, continuity and IID (Theorem 1). Moreover, the mixing measure over the logit parameter 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 mixed logit rule.

This result provides a simple, novel foundation for this widely used choice rule. It also shows that either IID or monotonicity is violated by all other stochastic choice rules, such as 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 mixed logit carries hidden global assumptions.[1] 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) or fail to choose better actions with higher probability (violating monotonicity).

A natural interpretation of our theorem is that it characterizes the behavior of a stochastic payoff maximizer in the presence of unobservable heterogeneity. Each realization of this heterogeneity gives rise to a distinct multinomial logit rule, corresponding to stochastic utility maximization with a particular level of utility shocks. The observed choice behavior, which takes the form of a mixture over such logits, thus reflects an average over this unobserved variation—whether across different individuals in a population or across multiple "selves" of a single decision maker, randomly realized at the time of choice. Importantly, this behavioral structure is not assumed and emerges as an inevitable consequence of our axioms.

Geometrically, the theorem shows that the set of stochastic choice rules satisfying the axioms is the convex hull of multinomial logit rules, which constitute its extreme points. These extreme rules can be characterized by strengthening IID to a more demanding axiom, which we call decomposability. Decomposability further requires that irrelevant choices are made independently, e.g., that decision makers choose soap and milk independently. We show that rules satisfying monotonicity, continuity and decomposability are exactly multinomial logit rules (Corollary 1).

According to Theorem 1, the mixing over multinomial logit rules remains constant across menus. An important consequence, explored in Proposition 1, is that any rule satisfying monotonicity, continuity, and IID must exhibit non-negative correlation in choice probabilities across product menus. Intuitively, individuals who experience less noise in one dimension of the menu will likewise experience less noise in the other

[1]Proposition 2 below shows that already on menus with three alternatives there are choice probabilities that are inconsistent with any mixed logit model.

dimension. The impossibility of negative correlation is not immediately apparent from the axioms. In particular, it is not a direct implication of IID, which allows for any correlation structure of choice probabilities in product menus. Instead, it is a joint implication of the three axioms put together.

To highlight another implication of our axioms, we show a novel property of mixed logit, which already applies to very simple menus with three actions. As an example, consider a menu with actions a, b and c, yielding payoffs 0, 2 and 7, respectively. Suppose that we observe that action a is chosen with probability 5%. What do the axioms imply for the choice probability of c? As it turns out (see §4), IID by itself does not imply any constraints. Monotonicity implies that the probability of choosing c is at least half of the complementary probability, 47.5%. Continuity imposes no restrictions. Interestingly, the three axioms together imply that the probability of c is at least 83.8%. This lower bound corresponds to the choice probability of c in the unique multinomial logit rule under which a is chosen with probability 5%. More generally, we show that under mixed logit, given the choice probability for the lowest payoff action, the choice probability for the highest payoff action is at least the choice probability under multinomial logit (Proposition 2).

In §6 we extend our main result to menus in which outcomes are not just one-dimensional payoffs, but rather take values in $\mathbb{R}^n$. We offer a number of interpretations to this settings, including choice among production plans, choice under ambiguity, and choice among Gaussian lotteries. In all of these cases, IID has a straightforward interpretation, as does the resulting mixed-logit rule.

In the first application, a menu represents a choice by a firm between production plans, where each plan is characterized by the quantities of different inputs it requires and the outputs it produces. The IID axiom means that a firm that has two factories that do not affect each other will make separate choices of plans. Our results show that firms act as if they are maximizing profit under some prices for the inputs and outputs. These prices can be inferred from the choice probabilities.

In the choice under ambiguity application, the outcome associated with an action is a Savage act, or a vector of state-contingent payoffs, rather than a fixed payoff, as in our one-dimensional model. Our results state that a model satisfying our axioms describes a population of subjective expected-payoff maximizing agents, each with a prior about the state and a level of noise. Since the only rationality assumption introduced directly by the axioms is monotonicity, beliefs and expectation maximization emerge from IID.

1.1 Related Literature

Mixed logit, a weighted average of multinomial logit rules, is a widely used model of randomness across various fields, including economics, psychology, statistics, machine learning, and statistical mechanics. Its simple structure and flexibility allow it to overcome the limitations of multinomial logit and accurately approximate empirical choice behavior (McFadden and Train, 2000; Train, 2009; Anderson, De Palma, and Thisse, 1992).

Despite its importance, characterizing mixed logit axiomatically—without relying on its functional form—has proven challenging. Saito (2017) provides two characterizations, though in a different setting. Saito considers a decision-maker with beliefs over possible menus, whose choice function is defined on sub-menus of a grand menu. One characterization requires that the agent's random choice be superior to the worst naive choice. The other, based on the positivity of the Block-Marschak polynomials, aligns with characterizations of general random utility models (Falmagne, 1978; Clark, 1996; McFadden and Richter, 1990). Axiomatic approaches have also been used for other generalizations of multinomial logit that share similarities with mixed logit, such as attribute rules or nested logit rules (Gul, Natenzon, and Pesendorfer, 2014; Kovach and Tserenjigmid, 2022) and multinomial logit with alternative priorities (Echenique, Saito, and Tserenjigmid, 2018).

In contrast to mixed logit, multinomial logit enjoys numerous characterizations. Its early popularity stemmed from its micro-foundation as a random-utility model with Gumbel-distributed shocks (Luce and Suppes, 1965), and its analytical tractability, providing explicit formulas for choice probabilities and welfare, unlike other random utility models that require Monte Carlo methods. Multinomial logit is also central to quantal response equilibrium, a generalization of Nash equilibrium for agents prone to errors (McKelvey and Palfrey, 1995).

According to Luce (1959), a choice rule exhibits independence of irrelevant alternatives (IIA) if the relative probabilities for a subset of alternatives do not depend on the presence of other alternatives in the choice set. 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. Instead, IID applies only to decisions made in separate, non-overlapping contexts.

Luce (1959) demonstrated that any behavior satisfying IIA can be generated by multinomial logit for some choice of utilities. In our setting—as in the analysis of random utilities—the scale of utilities is given. For a given scale, IIA implies that the probability of an alternative must be proportional to some fixed function of its utility. Multinomial logit corresponds to the exponential function, but IIA is also compatible with any other.

Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2021b) and Breitmoser (2021) characterize multinomial logit for a given utility scale. Both papers augment IIA with several other axioms to pin down the exponential dependence. Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2021b) characterize the whole one-parametric family of logit rules via axioms relating the rule's behavior across different noise levels and implying the multiplicative property of the exponent. Breitmoser (2021) pins down the exponential dependence by requiring translation invariance. Another related paper is Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2022), who characterize logit in a dynamic context. IIA underpins all these results; its rationality foundations are discussed by Cerreia-Vioglio, Lindberg, Maccheroni, Marinacci, and Rustichini (2021a). Yellott (1977) shows that a version of IIA pins down multinomial logit withing the class of independent random utility models with shock distribution fixed across menus; see also Luce and Suppes (1965) and Ragain and Ugander (2016). In contrast, our framework starts from a broader class of choice rules—including those with correlated or menu-dependent shocks, or not corresponding to stochastic utility maximization at all—and narrows it down purely through axioms. Absent IID, behavior across menus can be entirely unrelated.

Matějka and McKay (2015) develop a model combining choice with a given utility-scale and the rational-inattention framework of Sims (2003). They demonstrate that multinomial logit captures the behavior of a utility-maximizing individual with entropy-based attention cost. 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). The result of Matějka and McKay (2015) supports the conclusion of Camara (2022) that cognitive costs force decision-makers to split problems into unrelated sub-problems whenever possible.

Our axioms of independence of irrelevant decisions and decomposability have some similarities to separability notions in dynamic or multi-agent choice. 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, i.e., whether a joint distribution over choices is compatible with the existence of a single distribution over utility pairs; see also Frick, Iijima, and Strzalecki (2019); Li (2021); Kashaev, Gauthier, and Aguiar (2023) for multi-period dynamic random utility models. In a multi period context, Fudenberg, Lanzani, and Strack (2025) show that mixed probit describes the limiting choice of an agent with bounded memory and Gaussian information. Sandomirskiy, Sung, Tamuz, and Wincelberg (2025) explore a version of our decomposability axiom in multi-agent strategic environments, axiomatizing Nash and quantal response equilibria as well as new solution concepts.

2 Model

We study the choice behavior of a decision maker across a variety of decisions. Let $\mathcal{A}$ be a universal set of actions that the decision maker could possibly take. 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 $\mathcal{A}$. For example, if a_1 is the action of buying a certain soap and a_2 is the action of buying a certain milk, then (a_1, a_2) is the action of purchasing both. 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 finite set of possible actions and $o\colon A \to \mathcal{O}$ assigns an outcome to each action. 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, and we think of it as carrying no significance for the decision maker.

We formalize the model and discuss the results for $\mathcal{O} = \mathbb{R}$. 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. More general outcome spaces are considered in §6.

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

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

$$(A, o) = \begin{Bmatrix} \text{a} & \text{b} \\ 3.14 & -17 \end{Bmatrix}$$

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

The collection of all menus is denoted by $\mathcal{M}$. It consists of all pairs (A, o) where A is a finite subset of $\mathcal{A}$ and o is a function from A to $\mathbb{R}$. The richness of $\mathcal{M}$ distinguishes our approach from the standard stochastic choice setting in which menus are subsets of some fixed finite set of alternatives.

Another distinguishing feature of our approach is that it abstracts from the internal structure of alternatives and focuses solely on the payoffs induced by actions. This is in line with the random utility model literature, where utilities are taken as primitives. Our approach also admits a game-theoretic interpretation: a menu can be viewed as a one-player normal form game. Since actions are treated as mere labels, the same action may appear in different menus and result in different outcomes.[3]

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

$$\mathrm{IARU}(a \mid A, o) = \mathbb{P}\Big[o(a) + \varepsilon_a = \max_{b \in A} o(b) + \varepsilon_b\Big],$$

where $o(a)$ is interpreted as the utility of action a and $(\varepsilon_b)_{b \in A}$ are independent shocks with a common continuous CDF F. When these shocks follow the Gumbel distribution with a positive scale parameter $\frac{1}{\beta}$, i.e., $F(x) = \exp(-\exp(-\beta \cdot x))$, this is the *multinomial logit rule* (MNL), which is given by

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

While the multinomial logit rule is commonly used in the empirical literature for its computational tractability, it does not allow for random taste variation, various

[3] For example, an action a may represent the decision to buy a brand of soap in one menu, and represent the decision to wear a shirt of a particular color in another menu.

substitution patterns, and correlation in unobserved factors over time (Train, 2003). Some of these limitations are overcome by the *mixed-logit rule* (ML), which is a weighted average of multinomial logit rules, given by

$$\mathrm{ML}^{\mu}(a \mid A, o) = \int \mathrm{MNL}^{\beta}(a \mid A, o)\, \mathrm{d}\mu(\beta).$$

It is parameterized by a probability measure μ over the logit parameter β.

We consider several properties of stochastic choice rules. The first one is monotonicity. Monotonicity captures a sense in which the decision maker's choices are driven by preferences for higher payoffs. In particular, it limits the possible influence of action labels on choice behavior.[4]

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 $o(a) \geqslant o(a')$ it holds that $\Phi(a \mid A, o) \geqslant \Phi(a' \mid A, o)$.*

That is, a rule Φ is monotone if an action with a higher payoff is chosen with a higher probability than one with a lower payoff. Note that this axiom does not impose any constraints across menus, but only within a given menu.

Monotonicity is satisfied by all IARU models. Moreover, the class of stochastic choice rules that satisfy monotonicity is convex, i.e., a mixture of monotone rules is monotone. In particular, the mixed-logit and mixed-probit models 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 $\lim_n \Phi(a \mid A, o_n) = \Phi(a \mid A, o)$ 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 that always select one

[4]While our model is motivated by the view that only outcomes matter to the decision maker, the definition of a stochastic choice rule by itself does not preclude dependence of choice probabilities on labels, thus potentially allowing for framing effects. As we will see, such effects will be ruled out as a byproduct of our main result.

of the highest-payoff 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 combinations of unrelated decisions. We say that (A, o) is a *product menu* if

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

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, accordingly. 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

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

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

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

This is not a product menu, even though the action set is a product set (indeed, the same product set), because the payoffs are not additively separable.

Outside of the experimental lab, the menu (A, o) could represent a choice between two brands of soap, made together with a choice between two brands of milk. As is standard in the empirical literature, one can think of actual payoffs as comprising the payoffs in these menus, plus an idiosyncratic shock. The fact that (A, o) is a product menu would capture the lack of substitutability or complementarity between these products. In contrast, the menu (A, o') could be used to model a choice of jacket and pants, assuming that both have to be chosen correctly for the suit to work.

Product menus may remain appropriate when unrelated choices are influenced by prices. For instance, suppose the payoffs from consuming soaps and milks in the absence of prices are given by (A, o), and prices are $p(a), p(b)$ for soap and $q(s), q(t)$ for milk. Then the combined decision can be represented as

$$\left\{ \begin{array}{c|cc} & s & t \\ \hline a & 0 - p(a) - q(s) & 1 - p(a) - q(t) \\ b & 1 - p(b) - q(s) & 2 - p(b) - q(t) \end{array} \right\}.$$

Note that this model would only be appropriate in the absence of wealth effects and, in particular, no binding budget constraint. For example, a situation where the bundle (b, t) exceeds the budget, may be modeled by assigning a prohibitively negative payoff to that cell, thus breaking the product structure. More broadly, the presence of budget constraint or other joint constraints can generate interdependencies between otherwise unrelated decisions, in which case their combination would no longer correspond to a product menu.

Importantly, 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), as the examples above illustrate. Nonetheless, the richness of the menu collection $\mathcal{M}$ implies that a choice rule must provide predictions to what a decision maker will do when faced with a product menu.

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*

$$\Phi((a_1, a_2) \mid A, o) = \Phi(a_1 \mid A_1, o_1) \cdot \Phi(a_2 \mid A_2, o_2) \tag{2}$$

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

$$\Phi \begin{pmatrix} \text{a} & \text{b} \\ 0 & 1 \end{pmatrix} = \Phi \begin{pmatrix} \text{s} & \text{t} \\ 0 & 1 \end{pmatrix} = (1/3, 2/3).$$

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 soap-milk example, we might expect that consumers who are more careful when choosing soap are also more careful in their choice of milk, causing correlation when the two choices are made together.

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

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

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*

$$\Phi(a_1 \mid A, o) = \Phi(a_1 \mid A_1, o_1) \tag{3}$$

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

$$\Phi \begin{pmatrix} \text{a} & \text{b} \\ 0 & 1 \end{pmatrix} = \Phi \begin{pmatrix} \text{s} & \text{t} \\ 0 & 1 \end{pmatrix} = (1/3, 2/3),$$

then

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

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. Thus an analyst who wants to predict the consumer share of soap brands would have to investigate other, unrelated decisions consumers face, such as which milk they buy or which color shirt they choose to wear. On the other hand, with IID, these various decisions can all be ignored without affecting the relevant predictions.

IID is satisfied by multinomial logit, and, more generally, by a class of models that we call *separable IARU*; in fact, these models satisfy decomposability. In these models, shocks are added independently for each dimension of a product menu. This is in contrast with the standard IARU models in which shocks are added directly to choice pairs in a product menu as if the choices were made in *one shot*.

In separable IARU models, if a consumer is 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.[5] In one-shot IARU models, consumers choose the best milk-soap pair subject to some noise that is added to each pair. For a non-product menu, separable IARU and one-shot IARU models coincide.

We formally define separable IARU models recursively, as follows. Φ is a separable IARU model if there exists a shock distribution ε such that the following holds.

- For a menu $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$, which is a non-trivial product, i.e., with $|A_1| \geqslant 2$ and $|A_2| \geqslant 2$, it holds that

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

  That is, shocks are applied independently to each dimension.

- Otherwise,

$$\Phi(a \mid A, o) = \mathbb{P}[o(a) + \varepsilon_a = \max_{b \in A} o(b) + \varepsilon_b], \tag{4}$$

  where $(\varepsilon_b)_{b \in A}$ are independent and distributed as ε.

Interestingly, while the one-shot logit rule can be written as a separable IARU model, the one-shot probit rule cannot. Indeed, it turns out that the one-shot probit rule violates IID. See §5.

The IID axiom by itself allows for a wide range of choice rules. These include all separable IARU models, but also their mixtures, since the class of stochastic choice models that satisfy IID is convex. In particular, mixed logit satisfies IID.

[5] We thank an anonymous referee for suggesting this example.

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.

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

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

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 $\Phi(a_3 \mid A, o) \geqslant 47.5\%$, since $\Phi(a_2 \mid A, o) + \Phi(a_3 \mid A, o) = 95\%$ and $\Phi(a_2 \mid A, o) \leqslant \Phi(a_3 \mid A, o)$. 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.

Surprisingly, this intuition is wrong. It turns out that if we assume that Φ is monotone and satisfies IID, then $\Phi(a_1 \mid A, o) = 5\%$ implies that $\Phi(a_3 \mid A, o) \geqslant 83.8\%$. This bound is tight: if we make the stronger assumption that Φ is monotone and decomposable, then $\Phi(a_1 \mid A, o) = 5\%$ implies that $\Phi(a_3 \mid A, o) = 83.8\%$. As we explain in detail below, this is a consequence of our main results. Note that our main results also assume continuity, but this axiom is not needed in this example, since all payoffs are integer. Continuity allows us to extend to real payoffs. Indeed, there is a unique Φ that is monotone, decomposable, continuous, and satisfies $\Phi(a_1 \mid A, o) = 5\%$.

3 IID and Mixed Logit

Recall that for μ supported on $\mathbb{R}_+$, the mixed-logit rule given by

$$\mathrm{ML}^{\mu} = \int \mathrm{MNL}^{\beta} \, \mathrm{d}\mu(\beta) \tag{5}$$

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

Theorem 1. *Let Φ satisfy monotonicity, continuity, and IID for the outcome space $\mathcal{O} = \mathbb{R}$. Then Φ coincides with a mixed-logit rule ML^μ for some μ supported on $\mathbb{R}_+$.*

Theorem 1 is proved in §7. An interpretation of the theorem is that any decision maker satisfying the assumptions behaves as a stochastic payoff maximizer, whose magnitude of payoff shocks, captured by the parameter β, is itself random. Conditional on each realization of β, choices follow a multinomial logit rule. The realized magnitude is unobservable and thus the observable choice takes the form of a mixture over such logits, with mixing distribution μ. This distribution may reflect heterogeneity in noise magnitude across a population or variability of the internal state of a single individual. Notably, the fact that choice probabilities follow a random utility model emerges from the axioms and is not assumed a priori.

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 mixed-logit rule governs choice even in compound decisions that are not representable as product menus—for example, those involving interrelated choices or shared constraints.

Theorem 1 also has implications for possible choice correlation across unrelated decisions. Because the mixing distribution μ is fixed across all menus, choices in product menus must exhibit non-negative correlation: lower shock magnitudes (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. We are not aware of a direct proof of the fact that our axioms imply non-negative correlation.

Formally, we define correlation as follows. Given a menu (A, o) and a function $f\colon A \to \mathbb{R}$ 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 $\Phi(\cdot|A, o)$. 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

$$\Phi\Big(o_1(a_1) \cdot o_2(a_2) \,\Big|\, A, o\Big) \geqslant \Phi\Big(o_1(a_1) \,\Big|\, A_1, o_1\Big) \cdot \Phi\Big(o_2(a_2) \,\Big|\, A_2, o_2\Big). \tag{6}$$

We say that Φ exhibits *zero correlation* if (6) 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 first question correctly they are (weakly) more likely to answer the second one correctly.

Proposition 1. *For any probability measure μ, ML^μ exhibits non-negative correlation on every product menu. Moreover, ML^μ exhibits zero correlation on a product of non-trivial menus if and only if μ is a Dirac measure.*

The proof of Proposition 1 appears in §B, and is straightforward. What is more surprising is that every rule that satisfies our axioms has non-negative correlation. 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.

As a direct corollary of Theorem 1 and Proposition 1, we characterize multinomial logit rules by strengthening IID to decomposability.

Corollary 1. *Let Φ satisfy monotonicity, continuity and decomposability for the outcome space $\mathcal{O} = \mathbb{R}$. Then Φ coincides with a multinomial logit rule MNL^β for some $\beta \geqslant 0$.*

Instead of decomposability in Corollary 1, it is enough to assume that Φ satisfies IID and exhibits zero correlation on a product of a pair of non-trivial menus. Proposition 1 also implies 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.[6]

[6] An extreme point of a convex set is one that cannot be written as a non-trivial convex combination of others. In this light, representation (5) mirrors the Choquet theorem: any choice rule satisfying the axioms can be expressed as a mixture of extreme points. Fox et al. (2012) show that this representation is unique, that is $\mathrm{ML}^\mu = \mathrm{ML}^{\mu'}$ 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.

To illustrate the ideas behind Theorem 1, we sketch a direct proof for Corollary 1. The proof of Theorem 1 follows a similar path, but does it in a more complicated, random setting, making 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 payoffs) 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

$$(B, r) = \begin{Bmatrix} b_0 & b_1 \\ 0 & 1 \end{Bmatrix}.$$

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

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

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

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

$$(A, o) \otimes \underbrace{(B, r) \otimes (B, r) \otimes \cdots \otimes (B, r)}_{n \text{ times}}, \tag{7}$$

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

$$\Phi(a_3 \mid A, o) \cdot \Phi(b_0 \mid B, r)^{59} = \Phi(a_2 \mid A, o) \cdot \Phi(b_1 \mid B, r)^{59}. \tag{8}$$

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

4 Restrictions of Monotonicity and IID

The family of mixed-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 above, for the menu

$$(A, o) = \begin{Bmatrix} a_1 & a_2 & a_3 \\ 0 & 2 & 7 \end{Bmatrix}$$

it holds for every mixed-logit rule Φ that $\Phi(a_1 \mid A, o) = 5\%$ implies that $\Phi(a_3 \mid A, o) \geqslant 83.8\%$. Indeed, this lower bound of the choice probability of a_3 turns out to be $\mathrm{MNL}^\beta(a_3 \mid A, o)$ for the β that satisfies $\mathrm{MNL}^\beta(a_1 \mid A, o) = 5\%$. The following proposition shows that this is a general property of mixed-logit rules.

Proposition 2. *Let $A = \{a_1, \dots, a_n\}$ and $o(a_1) \leqslant o(a_2) \leqslant \cdots \leqslant o(a_n)$. Let $\beta \in \mathbb{R}$ be the (unique) logit parameter satisfying $\mathrm{MNL}^\beta(a_1 \mid A, o) = \Phi(a_1 \mid A, o)$. If Φ is a mixed-logit rule, then $\Phi(a_n \mid A, o) \geqslant \mathrm{MNL}^\beta(a_n \mid A, o)$.*

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 $\mathrm{MNL}^\gamma(a_n \mid A, o) = \Phi(a_n \mid A, o)$, then $\Phi(a_1 \mid A, o) \geqslant \mathrm{MNL}^\gamma(a_1 \mid A, o)$.

Proposition 2 highlights a property of mixed logit, independently of our main results. But since our axioms imply mixed 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.

Proof. If $o(a_1) = \cdots = o(a_n)$, then it is trivial, since Φ and MNL^β are uniform for all β, by monotonicity. Suppose then, that at least one of the inequalities is strict. Without loss of generality, suppose $o(a_1) = 0$. Define $d(\beta) = \sum_i \mathrm{e}^{\beta \cdot o(a_i)}$. Let $L(\beta) = \frac{1}{d(\beta)}$ and $H(\beta) = \frac{\mathrm{e}^{\beta \cdot o(a_n)}}{d(\beta)}$ denote the logit choice probabilities of the low and high alternatives, respectively, as a function of β.

Denote $p = \Phi(a_1 \mid A, o)$. Since $d(\beta)$ is strictly increasing, 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.

We show in Lemma 4 in the appendix that

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

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 mixed logit, there is a probability measure μ such that $\Phi(a_n \mid A, o) = \int H \,\mathrm{d}\mu$ and $p = \int L \,\mathrm{d}\mu$. For such μ, we have

$$\Phi(a_n \mid A, o) = \int H \,\mathrm{d}\mu = \int H \circ L^{-1} \circ L \,\mathrm{d}\mu \geqslant H \circ L^{-1} \left(\int L \,\mathrm{d}\mu \right)$$
$$= H \circ L^{-1}(p) = H(\beta) = \mathrm{MNL}^{\beta}(a_n \mid A, o).$$

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

5 Probit Rules

Probit rules are a natural and widely used family of stochastic choice rules. *One-shot probit* is an independent additive random utility (IARU) model with normal shocks, so that

$$\Phi(a \mid A, o) = \mathbb{P}\Big[o(a) + \varepsilon_a = \max_{b \in A} o(b) + \varepsilon_b\Big],$$

were ε_b are independent Gaussians. *Separable probit* is a separable IARU model in which each component in a product menu receives its own independent normal shocks (see (4)). 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 mixed logit.

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

$$(B, r) = \begin{Bmatrix} b_0 & b_1 \\ 0 & 1 \end{Bmatrix} \tag{9}$$

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

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

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.

We now show that separable probit violates monotonicity. Let Φ be a separable IARU with standard normal shocks. That is, Φ coincides with the one-shot probit on each menu that cannot be represented as a non-trivial product, while the prediction for product menus is defined as the product of predictions. This ensures decomposability and thus IID. Consider the menus

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

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 $\Phi(b \mid A_1, o_1) = G(9)$ and $\Phi(d \mid A_2, o_2) = \Phi(f \mid A_3, o_3) = G(6)$. Since Φ satisfies decomposability,

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

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 Higher Dimensional Outcomes

Theorem 1 generalizes to a higher dimensional setting, in which the outcome space is $\mathcal{O} = \mathbb{R}^n$ rather than $\mathbb{R}$. For menus with outcomes in $\mathbb{R}^n$, the multinomial logit rule with parameter $\beta \in \mathbb{R}^n$ is given by

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

where $\beta \cdot o(a)$ is the dot product. Likewise, the mixed-logit rule with mixing measure $\mu \in \Delta(\mathbb{R}^n)$ is given by

$$\mathrm{ML}^{\mu}(a \mid A, o) = \int \mathrm{MNL}^{\beta}(a \mid A, o)\, \mathrm{d}\mu(\beta).$$

The axioms extend verbatim to vector-valued outcomes, with $o(a) \geqslant o(a')$ in the monotonicity axiom interpreted as component-wise dominance.

Theorem 2. *Let Φ satisfy monotonicity, continuity, and IID for the outcome space $\mathcal{O} = \mathbb{R}^n$. Then Φ coincides with a mixed-logit rule ML^{μ} for some μ supported on $\mathbb{R}^n_+$.*

Theorem 2 is proved in Appendix D. It follows the proof strategy of Theorem 1, and moreover uses that theorem as a building block.

In the following subsections, we provide a few applications of this theorem. These applications differ in how elements in the outcome space are interpreted.

6.1 Choice under ambiguity

Let Θ be a finite set of states. An outcome x is a function $x\colon \Theta \to \mathbb{R}$, specifying state-contingent payoffs. That is, x is a Savage act and the space of all outcomes can be identified with $\mathcal{O} = \mathbb{R}^{\Theta}$.

Decision-makers are ambiguous about the state $\theta \in \Theta$ and so may take into account all the possible values x_θ. Monotonicity, continuity and IID translate naturally to this setting.

By Theorem 2, if Φ satisfies these axioms then there is a distribution μ over $\mathbb{R}^{\Theta}_+$ such that

$$\Phi(a \mid A, o) = \int \frac{\exp(\beta \cdot o(a))}{\sum_{b \in A} \exp(\beta \cdot o(b))} \mathrm{d}\mu(\beta).$$

We can write every nonzero $\beta \in \mathbb{R}^n_+$ as $\beta = \gamma \cdot p$, where $\gamma \in \mathbb{R}_+$ is given by $\gamma = \sum_\theta \beta_\theta$, and $p \in \Delta(\Theta)$ is given by $p_\theta = \beta_\theta/\gamma$. We can thus reparametrize by letting μ be a distribution over $\mathbb{R}_+ \times \Delta(\Theta)$. Let (γ, p) be drawn from μ, so that γ is a random positive number, and p is a random probability measure over Θ.

The choice probabilities thus correspond to a population of decision makers parametrized by (γ, p), each behaving as if they have a belief p over Θ according to which they stochastically maximize their subjective expected payoff $u(a) = p \cdot o(a)$, choosing an action a with probability

$$\frac{\exp(\gamma \cdot u(a))}{\sum_{b \in A} \exp(\gamma \cdot u(b))}.$$

When μ is a point mass (i.e., in the decomposable, multinomial logit case), there is one belief p and one parameter γ, so that behavior is consistent with a stochastic expected-payoff maximizer.

6.2 Mean-Variance Preferences

Our one-dimensional setting considers a set of actions, each yielding a deterministic payoff. In this section we apply our higher dimensional setting to study actions yielding random payoffs. We will assume that these payoffs are normally distributed, with known expectation and variance, and that decision makers prefer higher expectation and lower variance.

Accordingly, let the outcome space be $\mathcal{O} = \mathbb{R} \times \mathbb{R}_+$, where the first component is interpreted as the mean and the second component as the variance of a stochastic Gaussian monetary reward. The monotonicity axiom implies that individuals choose a with higher probability than b if a yields a stochastic payoff that has a higher expected value and lower variance. As an example, consider the the following menu, in which the decision-maker compares two investment decisions that differ substantially by expected rewards and variances

$$(A_1, o_1) = \left\{ \begin{matrix} \text{bond} & \text{crypto} \\ \binom{5}{2} & \binom{10}{10} \end{matrix} \right\}.$$

Another example is choice between health insurance plans:

$$(A_2, o_2) = \left\{ \begin{matrix} \text{a} & \text{b} \\ \binom{-2}{1} & \binom{-1}{4} \end{matrix} \right\},$$

where plan a has lower expectation and lower variance, corresponding to higher cost and better coverage.

In this setting, product menus correspond to combined choices in which the stochastic payoffs are independent of each other, so that the expectations and variances are simply added. For example, the product menu

$$(A_1, o_1) \otimes (A_2, o_2) = \left\{ \begin{matrix} \text{(bond,a)} & \text{(bond,b)} & \text{(crypto,a)} & \text{(crypto,b)} \\ \binom{3}{3} & \binom{4}{6} & \binom{8}{11} & \binom{9}{14} \end{matrix} \right\}$$

corresponds to a choice of investment and a choice of health insurance, which have independent outcomes, so that the variances are indeed summed.

Theorem 2 implies that decision probabilities correspond to a population of agents with various mean-variance preferences. I.e., there is a distribution μ over $(\beta_1, \beta_2) \in \mathbb{R}^2_+$ such that agents evaluate an action a yielding a lottery with mean m and variance v according to $u(a) = \beta_1 m - \beta_2 v$, and choose a with probability

$$\frac{\exp(u(a))}{\sum_{b \in A} \exp(u(b))}.$$

6.3 Production

A firm uses m inputs to produce n outputs. The outcome space is $\mathcal{O} = \mathbb{R}^m_- \times \mathbb{R}^n_+$, so that each outcome is a *production plan* specifying a quantity for each input and output. A menu is a choice of production plans that are feasible to the firm. We think of inputs as negative and outputs as positive, so that the monotonicity axiom implies that decision makers are more likely to choose a plan with less inputs and more outputs. A choice in a product menu corresponds to choosing to implement two production plans separately, which yields the sum of the two production plans. For example, this could model a choice of what to implement in two separate factories that do not affect each other. The IID axiom implies that the same choices will be made in the first factory, regardless of the existence of the second. This mirrors a basic feature of standard general equilibrium models in which combining firms does not change their optimal production plans.

Theorem 2 implies that the choice probabilities correspond to aggregate choices of a collection of firms that face various prices and stochastically maximize their profits. Namely, there is a distribution μ over prices in $\mathbb{R}^{m+n}_+$ such that each firm faces a random price $p \sim \mu$ and chooses a from A with probability

$$\frac{\exp(p \cdot o(a))}{\sum_{b \in A} \exp(p \cdot o(b))}.$$

The particular case of a multinomial logit rule corresponds to the existence of one fixed (rather than random) price faced by all firms.

7 Proof of Theorem 1

The remainder of this paper is a proof of our main theorem, which we hope is of technical interest in its own right. 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 $(X_1, X_2, \ldots)$ 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 $(X_{i_1}, X_{i_2}, \ldots)$ 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 $\mathbb{N} = \{1, 2, \ldots\}$ can be written as a disjoint union of infinite sets $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \cdots$ such that for all k it holds that $(X_i)_{i \in N_k}$ is exchangeable. In this case we say that $(X_1, X_2, \ldots)$ is a partially exchangeable process.

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 $(X_1, X_2, \ldots)$ be a partially exchangeable process, witnessed by the partition $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \cdots$. 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* $(X_1, X_2, \ldots)$ *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 $\{i_1, i_2, \ldots\}$ 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 lemma shows that there is a unique way to extend Φ to countable products of menus taking the form $(A_1, u_1) \otimes (A_2, u_2) \otimes \cdots$. We denote the set of all countable products of menus with $\mathcal{M}^\infty$.[7]

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.

Let $(A_1, o_1), (A_2, o_2), \ldots$ be a sequence of menus, and let $M = (A_1, o_1) \otimes (A_2, o_2) \otimes \cdots \in \mathcal{M}^\infty$ be their product. Denote by $\Omega = \prod_{i=1}^{\infty} A_i$ the set of sequences $(a_1, a_2, \ldots)$ corresponding to a choice in each of the menus. For a finite product $(A, o) = (A_1, o_1) \otimes \cdots \otimes (A_n, o_n)$, 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 $(X_1, X_2, \ldots)$ the random variables corresponding to the choice in each sub-menu, that is, $X_i(a_1, a_2, \ldots) = a_i$.

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

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

for all n *and all* $(b_1, b_2, \ldots, b_n)$. *Moreover, if* $(A_{i_1}, o_{i_1}) = (A_{i_2}, o_{i_2}) = \cdots$, *then the sequence of random variables* $(X_{i_1}, X_{i_2}, \ldots)$ *is exchangeable.*

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

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 $B_{b_1,\ldots,b_n} = \{a \in \Omega \mid a_1 = b_1, a_2 = b_2, \ldots, a_n = b_n\}$ are clopen. In particular, given $n \geqslant 1$ and a probability measure μ_n on $\prod_{i=1}^{n} A_i$, denote by $\mathcal{P}(\mu_0)$ the set of probability measures on Ω that agree with μ_n on sets of the form $B_{b_1,\ldots,b_n}$:

$$\mathcal{P}(\mu_n) = \Big\{\mu \,:\, \mu(B_{b_1,\ldots,b_n}) = \mu_n(\{(b_1,\ldots,b_n)\})\Big\}.$$

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 $\Phi(\otimes_{i=1}^{n}(A_i, o_i))$. By IID, μ_{n+1} agrees with μ_n on $B_{b_1,\ldots,b_n}$, 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 $B_{b_1,\ldots,b_n}$. Since the latter sets form a subbase of the topology, the measure μ is unique: $\{\mu\} = \cap_i \mathcal{P}(\mu_n)$.

Finally, suppose that $(A_{i_1}, o_{i_1}) = (A_{i_2}, o_{i_2}) = \cdots$; we would like to show that sequence of random variables $(X_{i_1}, X_{i_2}, \ldots)$ is exchangeable. 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 $(X_1, X_2, \ldots, X_n)$ of the entire sequence remains unchanged. This follows from the monotonicity of Φ, since the joint distribution of $(X_1, X_2, \ldots, X_n)$ 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 Mixed Logit from De Finetti

The next proposition is the heart of the proof of Theorem 1.

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

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

$$(B_0, u_0) = \begin{Bmatrix} \ell & h \\ 0 & 1 \end{Bmatrix}.$$

Define the sequence of menus $(A_1, o_1), (A_2, o_2), \ldots$ as follows. Write the natural numbers $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \cdots$ 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 $\mathcal{M}^\infty$ 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 $\Psi(\bigotimes_{i=1}^\infty (A_i, o_i))$, and let X_n be the coordinate projections, i.e., $X_n(a_1, a_2, \ldots) = a_n$, with the tail σ-algebra of $(X_1, X_2, \ldots)$ denoted by $\mathcal{T}$.

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

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

Enumerate $N_0 = \{j_1, j_2, \ldots\}$. Let

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

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

By Proposition 3 we know that $(Y_1, Y_2, \ldots)$ are exchangeable, as are $(Z_1, Z_2, \ldots)$. 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 $(X_1, X_2, \ldots)$ are independent, with Y_i chosen from F, and Z_i chosen from G, so that

$$F(b_k) = \mathbb{P}[Y_i = b_k \,|\, \mathcal{T}] \quad \text{and} \quad G(b_0) = \mathbb{P}[Z_i = b_0 \,|\, \mathcal{T}], \quad \text{for any } i. \tag{10}$$

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 $\mathrm{ML}^\mu(B_k, u_k)$. 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 $u_k(a) \geqslant u_k(a')$ and let $d = u_k(a) - u_k(a')$. By (10),

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

Since Y_i and Z_i are independent conditioned on the tail,

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

and so by the law of total expectation

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

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 $d = o_1(a) - o_1(a')$,

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

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

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

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

$$\mathbb{E}\big[F(a) \cdot G(\ell)^d \cdot P\big] = \mathbb{E}\big[F(a') \cdot G(h)^d \cdot P\big]. \tag{11}$$

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

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

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 (11) 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

$$F(a)G(\ell)^d = F(a')G(h)^d \text{ almost surely.} \tag{12}$$

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 (12) that $F(a) = 0$ conditioned 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 $\beta = +\infty$, $F = \mathrm{MNL}^{+\infty}(B_k, u_k)$, and likewise conditioned on $\beta = -\infty$, $F = \mathrm{MNL}^{-\infty}(B_k, u_k)$.

Outside the union of E_h and E_ℓ, β is finite, and it follows from (12) that

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

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

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

□

7.4 Final Steps

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

Proposition 5. *Suppose $ML^\mu|_{\mathcal{M}_\mathbb{Z}} = ML^\nu|_{\mathcal{M}_\mathbb{Z}}$. Then $\mu = \nu$.*

Proof. Given $\gamma > 0$, define the sequence of menus $(A_1, o_1^\gamma), (A_2, o_2^\gamma), \ldots$ by $A_n = \{a_1, \ldots, a_n, b\}$, $o_n^\gamma(a_k) = 0$, and $o_n^\gamma(b) = \lfloor \frac{\log(n)}{\gamma} \rfloor$. Then

$$\lim_n \mathrm{MNL}^\beta(b \mid A_n, o_n^\gamma) = \begin{cases} 0 & \text{if } \beta < \gamma \\ 1/2 & \text{if } \beta = \gamma \\ 1 & \text{if } \beta > \gamma. \end{cases}$$

It follows that

$$\lim_n \mathrm{ML}^\mu(b \mid A_n, o_n^\gamma) = \frac{1}{2}\mu(\{\gamma\}) + \mu((\gamma, \infty]),$$

and so μ is identified by ML^μ restricted to $\mathcal{M}_\mathbb{Z}$. □

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

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

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

$$\Phi^k(A, o) = \Phi\left(A, \frac{1}{k} \cdot o\right). \tag{13}$$

Note that such rules satisfy monotonicity and IID, since Φ does. Hence, by Proposition 4 again, each $\Phi^k|_{\mathcal{M}_\mathbb{Z}} = \mathrm{ML}^{\mu_k}|_{\mathcal{M}_\mathbb{Z}}$ for some μ_k. By (13) and Proposition 5, it follows that $\mu = k_*\mu_k$.

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,

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

Thus, $\Phi|_{\mathcal{M}_\mathbb{Q}} = \mathrm{ML}^\mu_{\mathcal{M}_\mathbb{Q}}$. 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 $\lim_n \Phi(a \mid A, o_n) = \frac{1}{2}$. If, however, $\mu(\{\infty\}) = \varepsilon > 0$, then $\lim_n \Phi(a \mid A, o_n) \leqslant \frac{1}{2}(1 - \varepsilon)$, violating continuity.

Fix any menu (A, o) and $a \in A$. For $n = 1, 2, \ldots$, define $\bar{o}_n \colon A \to \mathbb{Q}$ by $\bar{o}_n(a) = \frac{1}{n}\lceil n \cdot o(a) \rceil$. Since $\bar{o}_n(a) \to o(a)$, by continuity, we have $\Phi(A, o) = \lim_{n\to\infty} \Phi(A, \bar{o}_n)$. Hence, Φ is uniquely determined by $\Phi|_{\mathcal{M}_\mathbb{Q}}$. Since $\Phi|_{\mathcal{M}_\mathbb{Q}} = \mathrm{ML}^\mu|_{\mathcal{M}_\mathbb{Q}}$, and since ML^μ is continuous, it follows that $\Phi = \mathrm{ML}^\mu$. □

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 unrelated decisions are combined. 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 (3), the defining property of IID.

While the axioms we impose are quite mild, the richness of the domain allows a strong conclusion: the individual's behavior must follow a fixed mixed logit rule across all contexts. This result allows one to predict choice behavior when the payoffs are observable. But it also has implications for an analyst who only observes choice probabilities and aims to estimate the payoffs driving this choice behavior. If the analyst endorses the IID assumption with respect to these unobserved payoffs, they must deploy a mixed-logit rule to estimate payoffs from observed choice probabilities, and for subsequent counterfactual analysis.

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 mixed logit.

A more subtle extension would involve restricting the richness of the menu domain, for instance, by limiting the number of primitive menus used to form product menus. Our current proof relies on de Finetti's theorem for infinite exchangeable sequences, which requires taking products over arbitrarily many copies of the same menu. It may be possible to adapt the approach to finite exchangeability, following Diaconis and Freedman (1980). In such a setting, one might hope to recover approximate representations, with the quality of approximation depending on the sequence length. We leave these questions for future work.

The richness of our domain of menus may raise the question of whether behavior over such a large domain can be observed by an analyst. Even if we restrict attention to a finite domain, it would likely need to be large to guarantee a good approximation—perhaps unrealistically large for being observed in practice. A more practical view of our results is that the analyst assumes the decision maker could potentially face many menus and would behave consistently with IID. This assumption pins down the mixed logit rule, which can then be estimated from behavior over a smaller subset of observed menus.

References

S. P. Anderson, A. De Palma, and J.-F. Thisse. *Discrete choice theory of product differentiation*. MIT press, 1992.

Y. Breitmoser. An axiomatic foundation of conditional logit. *Economic Theory*, 72 (1):245–261, 2021.

M. K. Camara. Computationally tractable choice. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 28–28, 2022.

S. Cerreia-Vioglio, P. O. Lindberg, F. Maccheroni, M. Marinacci, and A. Rustichini. A canon of probabilistic rationality. *Journal of Economic Theory*, 196:105289, 2021a.

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, 2021b.

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.

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. Review of individual choice behavior by RD Luce. *American Economic Review*, 50(1):186–188, 1960.

P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, pages 745–764, 1980.

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.

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.

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.

R. Li. Dynamic random choice. *arXiv preprint arXiv:2102.00143*, 2021.

R. D. Luce. *Individual Choice Behavior: A theoretical analysis*. John Wiley and Sons, Inc, 1959.

R. D. Luce and P. Suppes. Preferences, utility and subjective probability. *Handbook of mathematical psychology*, pages 249–410, 1965.

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 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.

R. D. McKelvey and T. R. Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2): 122–136, 1964. ISSN 00129682, 14680262.

S. Ragain and J. Ugander. Pairwise choice markov chains. *Advances in neural information processing systems*, 29, 2016.

K. Saito. Axiomatic characterizations of mixed logit. Technical Report 1433, Caltech Social Science, 2017.

F. Sandomirskiy, P. H. Sung, O. Tamuz, and B. Wincelberg. Narrow bracketing and risk in games. *arXiv preprint arXiv:2502.11243*, 2025.

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.

K. E. Train. *Mixed Logit*, page 138–154. Cambridge University Press, 2003.

K. E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

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 Mixed Logit

Recall that the multinomial logit rule is given by

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

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

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

The following lemma shows that in extended mixed 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 mixed-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 $k = 1, 2, \ldots,$ $(A, o_k) = (\{a, b, c\}, (0, 1, k))$, and suppose, for the sake of contradiction, that $\mu([-\infty, 0)) > 0$. Then

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

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

□

B Proof of Proposition 1

Lemma 2 (See, e.g., Hardy, Littlewood, and Pólya (1952)). *If $f, g\colon \mathbb{R} \to \mathbb{R}$ are strictly increasing functions and μ is a probability measure then $\int fg \, \mathrm{d}\mu \geqslant \int f \, \mathrm{d}\mu \cdot \int g \, \mathrm{d}\mu$, with equality if and only if μ is a Dirac measure.*

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

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

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

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

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

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

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

where the equality holds when $u_i \cdot \sqrt{\mathrm{e}^{\beta u_i}} = b \cdot \sqrt{\mathrm{e}^{\beta u_i}}$ for all i, i.e., when u_i is constant. □

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

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

and

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

It follows that

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

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

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

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

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

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

□

C Proof of Proposition 2

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

Proof. Let o_i denote $o(a_i)$, and let $d(\beta) = \sum_i \mathrm{e}^{\beta o_i}$. Then

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

The derivatives of L are given by:

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

Without loss of generality, assume $o(a_1) = 0$ so that $H = \mathrm{e}^{\beta o_n} L$. Then

$$H' = o_n \mathrm{e}^{\beta o_n} L + e^{\beta o_n} L' = \mathrm{e}^{\beta o_n}(o_n L + L')$$
$$H'' = o_n^2 \mathrm{e}^{\beta o_n} L + 2o_n \mathrm{e}^{\beta o_n} L' + \mathrm{e}^{\beta o_n} L''.$$

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

$$\frac{d'' - o_n d'}{d^3} < 0,$$

so the overall expression is positive, as desired.

□

D Proof of Theorem 2

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

Since Λ_t satisfies monotonicity, continuity and IID, by Theorem 1, $\Lambda_t = \mathrm{ML}^{\nu_t}$ for some ν_t supported on $\mathbb{R}_+$.

Let $(B_{n+1}, u_{n+1}), (B_{n+2}, u_{n+2}), \ldots$ 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 $t = 1, \ldots, n$ choose

$$(B_t, u_t) = \begin{Bmatrix} \ell_t & h_t \\ 0 & e_t \end{Bmatrix}. \tag{14}$$

Since $(B_{n+1}, u_{n+1}), (B_{n+2}, u_{n+2}), \dots$ is an enumeration, the menus (14) 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 $\mathbb{N} = N_1 \cup N_2 \cup \cdots$ 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.[8] Let $\mathbb{P}$ denote the probability measure on Ω given by $\Psi(\bigotimes_{i=1}^{\infty}(A_i, o_i))$, 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

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

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 ith copy of B_k and Z_i^t to the ith copy of B_t, as defined in (14).

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

$$F(b_k) = \mathbb{P}[Y_i = b_k \mid \mathcal{T}] \;\text{ and }\; G_t(b_t) = \mathbb{P}[Z_i^t = b_t \mid \mathcal{T}], \text{ for any } i. \tag{15}$$

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

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

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

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 $\mathrm{ML}^\mu(B_k, u_k)$. 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 (15), Y_i and $Z_i^1, \ldots, Z_i^t$ are independent conditioned on the tail. By the same argument used in the proof of Proposition 4 we conclude that

$$\begin{aligned} F(a) &\cdot G_1(\ell_1)^{d_1^+} \cdots G_n(\ell_n)^{d_n^+} \cdot G_1(h_1)^{d_1^-} \cdots G_n(h_n)^{d_n^-} \\ &= F(a') \cdot G_1(\ell_1)^{d_1^-} \cdots G_n(\ell_n)^{d_n^-} \cdot G_1(h_1)^{d_1^+} \cdots G_n(h_n)^{d_n^+} \quad \text{almost surely.} \end{aligned} \tag{16}$$

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

$$\frac{F(a)}{F(a')} = \mathrm{e}^{\beta_t(u_t(a) - u_t(a'))} = \frac{\mathrm{e}^{\beta_t u_t(a)}}{\mathrm{e}^{\beta_t u_t(a')}},$$

and we may conclude that $\Lambda_t|_{\mathcal{M}_\mathbb{Z}} = \mathrm{ML}^{\mu_t}|_{\mathcal{M}_\mathbb{Z}}$. Thus, by Proposition 5, $\mu_t = \nu_t$, so the support of μ_t is contained in $\mathbb{R}_+$. Thus, β is finite (and F is non-degenerate) almost surely. For any (B_k, u_k) it follows from (16) that

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

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

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

We now show that $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}} = \mathrm{ML}^\mu|_{\mathcal{M}_{\mathbb{Q}^n}}$. As in the proof of Theorem 1, we define, for $k = 1, 2, \ldots,$

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

Since each of these rules satisfy monotonicity and IID they are mixed-logit rules. We now show that their mixing measures must be $k_*^{-1}\mu$ by proving an analogue of Proposition 5 for multidimensional mixed-logit rules.

Given $\gamma \in \mathbb{R}^n_{++}$ and $m \in \mathbb{N}$, define the menus $(C_m, o_1^m), \ldots, (C_m, o_n^m)$ in $\mathcal{M}_{\mathbb{Z}^n}$ by $C_m = (a_1, \ldots, a_m, b)$, $o_t^m(a_k) = 0$, and $o_t^m(b) = \lfloor \frac{\ln(m)}{\gamma_t} \rfloor \cdot e_t$, for $t = 1, \ldots, n$. Let $(A_m, o_m) = (C_m, o_1^m) \otimes \cdots \otimes (C_m, o_n^m)$. Then

$$\lim_{m} \mathrm{MNL}^{\beta}((b,\ldots,b) \mid A_m, o_m) = \begin{cases} 0 & \text{if } \beta_t < \gamma_t \text{ for some } t \\ \frac{1}{2^k} & \text{if } \beta \geqslant \gamma \text{ and } |\{t : \beta_t = \gamma_t\}| = k. \end{cases}$$

Thus μ is identified by ML^{μ} restricted to $\mathcal{M}_{\mathbb{Z}^n}$.

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

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

Thus, $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}} = \mathrm{ML}^{\mu}|_{\mathcal{M}_{\mathbb{Q}^n}}$. Fix any menu (A, o) and $a \in A$. For $n = 1, 2, \ldots$, define $\bar{o}_n \colon A \to \mathbb{Q}^n$ by $\bar{o}_n(a) = \frac{1}{n}\lceil n \cdot o(a) \rceil$. Since $\bar{o}_n(a) \to o(a)$, by continuity, we have $\Phi(A, o) = \lim_{n\to\infty} \Phi(A, \bar{o}_n)$. Hence, Φ is uniquely determined by $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}}$. Since $\Phi|_{\mathcal{M}_{\mathbb{Q}^n}} = \mathrm{ML}^{\mu}|_{\mathcal{M}_{\mathbb{Q}^n}}$, and since ML^{μ} is continuous, it follows that $\Phi = \mathrm{ML}^{\mu}$.

□

In some natural applications of stochastic choice, outcomes are limited to a subset of Euclidean space, such as the positive orthant (see, e.g., §6.2 and §6.3). 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 2. *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 $\Phi = \mathrm{ML}^{\mu}$ for some μ supported on $\mathbb{R}^n_+$.*

Proof. Let C and Φ 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 $a \in A$. Note that, by IID, Ψ is well-defined, since any such c leads to the same distribution $\Psi(A, o)$. It is obvious that Ψ is monotone. Ψ is moreover continuous as c may be chosen continuously. 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

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

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