Decomposable Stochastic Choice*

Fedor Sandomirskiy† Omer Tamuz‡

December 7, 2023

Abstract

We investigate inherent stochasticity in individual choice behavior across diverse decisions. Each decision is modeled as a menu of actions with outcomes, and a stochastic choice rule assigns probabilities to actions based on the outcome profile. Outcomes can be monetary values, lotteries, or elements of an abstract outcome space. We characterize decomposable rules: those that predict independent choices across decisions not affecting each other. For monetary outcomes, such rules form the one-parametric family of multinomial logit rules. For general outcomes, there exists a universal utility function on the set of outcomes, such that choice follows multinomial logit with respect to this utility. The conclusions are robust to replacing strict decomposability with an approximate version or allowing minor dependencies on the actions' labels. Applications include choice over time, under risk, and with ambiguity.

*The paper 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, Geoffroy de Clippel, Federico Echenique, Francesco Fabbri, Alexander Guterman, Wade Hann-Caruthers, Annie Liang, Daniel Litt, Yusufcan Masatlioglu, Jeffrey Mensch, Stephen Morris, Pietro Ortoleva, Tom Palfrey, Marcin Pęski, Luciano Pomatto, Gil Refael, Philipp Sadowski, Todd Sarver, Ilya Segal, Barry Simon, Stanislav Smirnov, William Thomson, Aleh Tsyvinski, Christopher Turansick, Shoshana Vasserman, Ryan Webb, Leeat Yariv, and seminar participants at Caltech, Brown University, Duke University, the Hebrew University of Jerusalem, HSE University, the Ohio State University, the University of North Carolina at Chapel Hill, the University of Toronto, the Fields Institute, and North America Summer Meeting of Econometric Society 2023.

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

‡Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a grant from the Simons Foundation (#419427), a BSF award (#2018397), and a National Science Foundation CAREER award (DMS-1944153).

1 Introduction

Consider an individual choosing an action from a finite menu of actions, say, whether to wear a red, green, or blue shirt on a particular day. Such a choice may look stochastic to an analyst unaware of some of the factors affecting the choice, e.g., an individual's favorite color or whether it is St. Patrick's Day. This perceived stochasticity motivated the widespread empirical use of random utility models, which assume that individuals are rational utility maximizers, but the utilities contain random unobserved components originating from unobservable latent variables.

By contrast, a substantial body of research suggests that choice behavior may be inherently stochastic, i.e., randomness would not be completely eliminated even if the analyst had perfect access to the individual's type and all the external factors affecting the decision. There is neither consensus on the origin of inherent stochasticity nor on the way to model it. For example, inherent stochasticity may originate from a preference for randomization, or from ambiguity aversion/regret minimization (see, e.g., Machina, 1985; Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella, 2019; Agranov and Ortoleva, 2017, 2022, and references therein), cognitive costs (see Matějka and McKay, 2015), or the neuro-physiological origin of the decision-making process (e.g., Webb, 2019, and references therein).

Our paper develops an axiomatic approach to model the inherent stochasticity of choice behavior. Our approach has two key features. First, it is agnostic to the origin of stochasticity. Instead of deriving a choice rule from a particular mechanism behind stochasticity, we start from a broad family of choice rules, allowing for irrational behaviors and so do not assume that individuals are random utility maximizers.[1] We impose assumptions (axioms) on the choice rules and characterize choice rules compatible with these axioms. Second, we assume that the choice rule governs individual's behavior—and thus axioms apply—over a broad range of decisions. This enables strong conclusions based on seemingly weak axioms.

Each decision is modeled as a finite menu of actions in which each action is assigned an outcome. The outcomes represent all relevant information about the actions. We

[1]Necessary conditions for a stochastic choice rule to originate from a random utility model were derived by Block and Marschak (1959). Falmagne (1978) demonstrated their sufficiency; see also McFadden and Richter (1990). By allowing irrational behaviors, we also allow for violating these conditions.

first study the baseline case, where the relevant information about each action is summarized by a single number. Such a numerical outcome may represent monetary rewards or costs associated with an action. However, we do not assume that the decision-maker is engaged in maximization of any sort.

A stochastic choice rule is a model that describes how an individual chooses their actions across different decisions. Formally, they are maps that assign to each menu a choice probability for each action. This rich non-parametric family of rules leads to the problem of model selection. We take an axiomatic approach to this problem, based on a concept that we call decomposability.

Decomposability can be motivated by the following example. Consider choosing which of three shirts to wear and which of two brands of cereals to eat for breakfast. While this is a choice from a menu of six actions, the decision naturally decomposes into two separate ones that every human would treat separately. Decomposability is an Occam's Razor that extends this logic, imposing independent choices across sub-decisions that are unrelated in the sense that there is no complementarity or substitutability in their outcomes. By contrast, the choice of shirt may not be unrelated to the choice of pants, and so decomposability would not imply that these are chosen independently.

Formally, a stochastic choice rule is decomposable if, whenever a decision decomposes into unrelated ones, the choices are made as if they were made in isolation, and independently of each other. In particular, consider two unrelated decisions, such as selecting a shirt and cereal. We can apply a choice rule to each of the two corresponding menus separately and then sample a pair of actions independently from the corresponding distributions. Alternatively, we can compose the two menus into one by pairing actions and summing outcomes, and then apply the rule to this composed menu. Summing the outcomes of the two actions manifests that actions are unrelated in the sense that the reward or cost of one does not affect that of the other. We say that a rule is decomposable if both routes yield the same distribution over pairs of actions.

At first glance, decomposability is an extremely weak requirement, as it places no restriction on decisions that are not composed of independent sub-decisions. Contrary to this intuition, decomposability imposes enough constraints across decisions to pin down a one-parametric family of rules.

Our first main result is the characterization of all decomposable stochastic choice rules. Under mild additional assumptions, we show that multinomial logit is the only such rule. Moreover, the parameter in the logit (the coefficient in the exponent) is identical across all decisions. This result provides a simple, novel foundation for this widely used choice rule.

Curiously, decomposability implies rational utility-maximizing behavior. Indeed, multinomial logit corresponds to a random utility model with shocks following the Gumbel distribution (Luce and Suppes, 1965).[2] Hence, the individual behaves as if she was maximizing or minimizing the outcome of a chosen action plus noise with magnitude depending on the parameter. Thus the outcomes can be interpreted as utilities or disutilities, respectively, even though no rationality assumption was made a priori.

The theorem highlights that an analyst selecting a stochastic choice rule for one particular decision implicitly makes assumptions about the counterfactual behavior of the individual. Unless multinomial logit is selected for the decision of interest, it is possible to find a pair of unrelated decision problems that would be solved together differently than they would in isolation.

The theorem also demonstrates a logical relation between seemingly unrelated behavioral patterns. Indeed, multinomial logit has a number of properties distinguishing it from a generic stochastic choice rule, and our theorem demonstrates that all these properties are implied by decomposability. For example, logit has the property of independence of irrelevant alternatives (IIA) that is commonly criticized as too strong and lacking experimental evidence. We conclude that any rule violating IIA necessarily violates decomposability.

Having established this result, we turn to a more general setting of a richer space of outcomes. For example, suppose that each action generates a stochastic reward, a stream of payoffs, or a reward that depends on a state variable that the decision-maker is ambiguous about. All these cases can be captured by allowing outcomes to be elements of an abstract outcome space endowed with a binary operation, which corresponds to combining outcomes of unrelated actions. For example, combining stochastic rewards corresponds to the convolution of reward distributions. The notion of decomposability of a choice rule extends naturally by replacing the summation of

[2] By our theorem, all random utility models except for multinomial logit violate decomposability. We illustrate this violation for the widely used probit rule.

numerical outcomes with their composition via the binary operation.

In this general setting, our second main result demonstrates that a sophisticated decision maker, whose behavior satisfies decomposability, behaves as if she were driven by a utility over outcomes. Namely, there is a canonical way to assign a utility to each element of the outcome space so that any decomposable rule is multinomial logit with respect to these canonical utilities. We refine the result by characterizing the functional form of the canonical utility representation for particular outcome spaces. For example, the canonical utility for stochastic rewards with bounded second moment is the mean-variance utility commonly used to model risk-aversion. For rewards depending on an ambiguous state, we recover the standard expected utility representation. Hence, also in this more general setting, we obtain utility maximization without any rationality assumptions beyond decomposability.

Decomposability only requires that unrelated decisions are perceived and taken as unrelated and makes no assumptions about related decisions.[3] Despite the weakness of this assumption and its possible intellectual appeal, real people's behavior satisfies it only approximately at best. Indeed, the theoretical and experimental literature on ambiguity aversion describes the hedging phenomenon, which implies that combining unrelated choices may alter behavior; see, e.g., Azrieli, Chambers, and Healy (2018).

Our third main result is a robustness check that demonstrates that any choice rule that is approximately decomposable must be close to multinomial logit. This result is quantitative, i.e., we get an explicit bound on how close the behavior is to a multinomial logit based on how close it is to satisfying decomposability.

1.1 Related literature

Multinomial logit is a ubiquitous model of randomness in a variety of fields: economics, psychology, statistics, machine learning, and statistical mechanics.

Multinomial logit was proposed by Luce (1959) to model discrete choice behavior in experimental psychology and then popularized in economics by McFadden (1974); see McFadden (2001) for a history of ideas behind multinomial logit and Train (2009); Anderson, De Palma, and Thisse (1992) for a general economic perspective on stochastic choice models. The early popularity of multinomial logit in economics was driven

[3]Nevertheless, widely documented narrow choice-bracketing indicates that even related decisions are often treated by decision makers as if they were unrelated (see, e.g., Barberis et al., 2006).

by the fact that it can be micro-founded as a random-utility model with shocks following the Gumbel distribution (Luce and Suppes, 1965), and because of its convenience both for applied econometrics and theory; it gives explicit formulas for choice probabilities and welfare whereas other random utility models require Monte Carlo methods. Multinomial logit is also at the heart of the quantal response equilibrium, a generalization of the Nash equilibrium for error-prone decision-makers (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. 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.

The work of Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2021b) is the closest to ours. They obtain the first characterization of multinomial logit for a given utility-scale. The characterization augments IIA by several other axioms to pin down the exponential dependence. The key insight is to characterize the whole one-parametric family of multinomial logit rules, rather than considering multinomial logit for a particular parameter. The additional axioms relate the behavior of the stochastic choice rule for different noise levels and imply the multiplicative property of the exponent. Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2022) develop these ideas further and characterize a dynamic version of multinomial logit where the noise level can depend on the timing of the decision, thus relating static choice models to neuro-economic drift-diffusion models. IIA underpins both results; its rationality foundations are discussed by Cerreia-Vioglio, Lindberg, Maccheroni, Marinacci, and Rustichini (2021a).

Our characterization of multinomial logit relies on decomposability instead of IIA. One can think of decomposability as *independence of irrelevant sub-decisions* rather than alternatives, intuitively a less demanding and more natural requirement than IIA, which is commonly criticized as unrealistic. Indeed, the red bus/blue bus thought experiment by Debreu (1960) indicates that IIA is problematic, especially if alternatives are substitutes, and this conclusion is supported by vast empirical evidence (e.g., Becker, Degroot, and Marschak, 1963; McFadden, Tye, and Train, 1977). Various generalizations of multinomial logit can capture non-IIA behaviors: e.g., Gul, Natenzon, and Pesendorfer (2014) characterize the so-called attribute rules related to nested logit, Saito (2018) obtains the first axiomatization for mixed logit, and Echenique, Saito, and Tserenjigmid (2018) characterize a version of multinomial logit incorporating alternatives' priorities. Similarly to Luce (1959), the scale of utilities in these results is not fixed.

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.

Multinomial logit is known as the Maxwell-Boltzmann or the Gibbs distribution in statistical mechanics and information theory. The result of Matějka and McKay (2015) is related to the well-known fact that multinomial logit maximizes the Shannon entropy over all distributions with fixed mean utility (Shannon, 1948, 1959). The entropy-based derivation is close to Boltzmann's original informal argument that, for a system in thermodynamic equilibrium, the distribution of micro-states must be as uniform as the law of conservation of energy permits. Other informal derivations of the Gibbs distribution use specific properties of physical systems. For example, Landau and Lifshitz (1951) offer a general argument for Hamiltonian systems relying on Noether's characterization of their continuous symmetries, and Feynman, Leighton, and Sands (2011) discuss ideal gases. The Hammersley-Clifford Theorem (see, e.g., Besag, 1974) characterizes the Gibbs distribution for lattice models of statistical mechanics by the Markov property, which shares some similarity to IIA.

Our notion of decomposability has some similarity to separability in dynamic choice models. Chambers, Masatlioglu, and Turansick (2021) 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.

2 Model

We consider a single decision-maker and model her behavior across various decisions. There is a non-empty set of possible actions $\mathcal{A}$. We assume that this set is closed under the operation of forming ordered pairs. I.e., 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 cereal and a_2 is the action of wearing a certain shirt, then (a_1, a_2) is the action of doing both. Note that this condition implies that $\mathcal{A}$ is infinite.

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.

We first formalize the model and discuss the results for $\mathcal{O} = \mathbb{R}$. In this case, a menu is simply a one-player game. This benchmark outcome space can be used to model a decision-maker who compares actions by a single number, such as their monetary reward or cost. General outcome spaces $\mathcal{O}$ are discussed in §4.

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

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

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

Let $\mathcal{M}$ be the collection of all menus. In other words, $\mathcal{M}$ consists of all pairs (A, o) where A is a finite subset of $\mathcal{A}$ and o is a function from A to $\mathcal{O}$. 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. Moreover, in our setting, the same action can have different outcomes in different menus.

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 $\Phi(A, o)_a$ the probability that $\Phi(A, o)$ assigns to $a \in A$. We think of Φ as describing or predicting the choices of a single decision-maker across different situations. Alternatively, a stochastic choice rule can be thought of as a solution concept for one-player games.

We consider several properties of stochastic choice rules. The first one is neutrality. Neutrality captures a sense in which the decision-maker's choice is driven by the outcomes rather than the names of actions.

Axiom 1 (Neutrality). *A rule Φ is **neutral** if actions in the same menu sharing the same outcome are chosen with the same probability: for any menu (A, o) in $\mathcal{M}$ and any $a, a' \in A$ such that $o(a) = o(a')$ it holds that $\Phi(A, o)_a = \Phi(A, o)_{a'}$.*

Note that this axiom does not impose any constraints across menus, but only within a given menu, and only if there are any actions that share the same outcome.

To introduce our main axiom, decomposability, we will need an operation of combining two unrelated choices into one. Given two menus (A_1, o_1) and (A_2, o_2), we define their product $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ by

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

Indeed, it is intuitive to assume that monetary rewards or costs are additive across unrelated actions.

For example, suppose that (A_1, o_1) with $A_1 = \{a, b\}$ is a choice between two cereals at a supermarket and (A_2, o_2) with $A_2 = \{p, q, r\}$ is a choice between job offers. Then the product menu $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ represents a choice of cereal *and* a job offer. The set of actions $A = \{(a, p), (a, q), (a, r), (b, p), (b, q), (b, r)\}$ consists of all pairs of choices from A_1 and A_2. The outcome defined by $o = o_1 + o_2$ captures no interaction between the two dimensions of the decision: under (A, o), the choice of cereal does not affect the decision maker's rewards for job offers.

Decomposability concerns a rule's prediction for product menus.

Axiom 2 (Decomposability). *A rule Φ is **decomposable** if for all menus $(A_1, o_1), (A_2, o_2) \in \mathcal{M}$ and their product (A, o), it holds that*

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

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

Decomposability means that, for product menus, the predicted distribution is independent across the two dimensions. Moreover, in each dimension, the prediction is the same as when that decision is made in isolation. In the cereal and job offer example above, a rule that satisfies decomposability would predict that the choice of cereal would be independent of the choice of the job offer: Observing one would not change the prediction of the other. Furthermore, the probabilities that different cereals are chosen are the same as they would be if the menu included the cereals only. For example, the decomposability of Φ would imply that $\Phi(A, o)_{(a,p)} = \Phi(A_1, o_1)_a \cdot \Phi(A_2, o_2)_p$: The probability of choosing cereal a and job offer p from the product menu is the product of their probabilities in the two menus.

Note that the product menu is well-defined even for (A_1, o_1) and (A_2, o_2) representing related choices such as choosing a shirt and choosing a tie. In this case, the combined decision does not correspond to the product menu $(A_1, o_1) \otimes (A_2, o_2)$, which represents a hypothetical situation where these two choices are combined as if they were unrelated. The decomposability axiom only restricts the behavior of the individual in such a hypothetical situation and does not constrain the rule's behavior on menus that are not products.

If we interpret outcomes as monetary rewards or costs, decomposability captures a sense in which behavior exhibits no wealth effects. This is a strong assumption, though commonly made. In our model, it originates from the additive way in which outcomes in a product menu are defined (2). We go beyond additivity in §4, and now continue with the baseline model.

Some of our results require a mild regularity assumption. For a fixed set of actions A, we say that a series of menus (A, o_n) converges to (A, o) if $\lim_n o_n(a) = o(a)$ for all $a \in A$.

Axiom 3 (Continuity). *A rule Φ is **continuous** if for any sequence of menus (A, o_n) from $\mathcal{M}$ converging to (A, o), we have $\lim_n \Phi(A, o_n)_a = \Phi(A, o)_a$ for all $a \in A$.*

3 Decomposable Rules for $\mathcal{O} = \mathbb{R}$

Recall that the outcome space $\mathcal{O} = \mathbb{R}$ represents a decision-maker comparing actions by a single number, such as their monetary reward or cost. The rewards and costs are additive across unrelated actions (1).

The **multinomial logit rule** with parameter $\beta \in \mathbb{R}$ is given by

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

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

One can easily verify that multinomial logit satisfies neutrality, decomposability, and continuity. At first glance, these requirements are quite weak, and many other rules must satisfy them. Indeed, decomposability constrains a rule only on product menus, and neutrality does not constrain a rule across menus. For example, the two axioms do not seem to relate the behavior of a rule on the menus

$$(A, o) = \begin{Bmatrix} a_1 & a_2 & a_3 \\ -17 & -17 & 42 \end{Bmatrix} \qquad (B, r) = \begin{Bmatrix} b_0 & b_1 \\ 0 & 1 \end{Bmatrix}.$$

These menus are indecomposable in the sense that they are not (non-trivial) product menus. Intuitively, we can define a rule on indecomposable menus arbitrarily in a neutral way and extend it to all product menus by decomposability. Surprisingly, this intuition is wrong, and we would not get a decomposable neutral rule unless it was defined on indecomposable menus in a very particular way.

Theorem 1. *Let Φ be a neutral, decomposable, continuous stochastic choice rule for the outcome space $\mathcal{O} = \mathbb{R}$. Then Φ coincides with the multinomial logit rule for some $\beta \in \mathbb{R}$.*

The constant β in the multinomial logit rule is pinned down by the choice distribution on any non-trivial menu, i.e., on any menu $(A, o) \in \mathcal{M}$ where o is not constant. In other words, the behavior of a rule on any non-trivial menu pins down the behavior on all menus. After discussing other implications of the theorem, we will explain how two weak axioms—decomposability and neutrality—become strong when combined.

Recall that the family of independent additive random utility (IARU) models used to model the behavior of a rational utility-maximizing decision-maker is given by

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

where $o(a)$ is interpreted as the utility of action a and ε_a are independent shocks with a continuous distribution F. For $\beta > 0$, multinomial logit can equivalently be defined as an independent additive random utility model corresponding to the Gumbel distribution $F(x) = \exp(-\exp(-\beta \cdot x))$. Note that—although the outcome $o(a) \in \mathbb{R}$ was interpreted as the reward or cost of an action a—we did not a priori assume that the decision-maker is engaged in any sort of utility maximization. Curiously, the theorem implies that a decision-maker whose choice rule Φ satisfies the assumptions behaves as if she was a stochastic utility-maximizer (the case of $\beta > 0$) or disutility-minimizer ($\beta < 0$). The theorem also justifies the interpretation of decomposability and neutrality as rationality requirements. We see that these requirements imply rationality in the usual sense of utility maximization but are stronger than that since only multinomial logit is decomposable among all IARU.

By Theorem 1, a general IARU violates decomposability by inducing correlation between unrelated choices. For example, consider the probit rule corresponding to the standard Gaussian distribution $F = N(0,1)$. We examine its outcomes for a menu

$$(B,r) = \left\{ \begin{matrix} b_0 & b_1 \\ 0 & 1 \end{matrix} \right\} \tag{3}$$

and its "square"

$$(C,s) = (B,r) \otimes (B,r) = \left\{ \begin{matrix} (b_0,b_0) & (b_0,b_1) & (b_1,b_0) & (b_1,b_1) \\ 0 & 1 & 1 & 2 \end{matrix} \right\}.$$

We get $\mathrm{Probit}(B,r)_{b_1} \simeq 0.760$ and $\mathrm{Probit}(C,s)_{(b_1,b_1)} \simeq 0.617$. Since $0.617 > 0.76^2 = 0.58$, Probit violates decomposability, putting more weight on the maximal-utility action in (C,s) than a decomposable rule coinciding with probit on (B,r) would put. Summing Gaussian shocks results in a shock with twice the variance, and so we might anticipate that keeping the same shock variance in the product menu leads to less randomness than required for decomposability. Theorem 1 implies that even if we made the variance or even the shock distribution a function of the menu, we would not achieve decomposability unless the resulting choice probabilities correspond to Gumbel-distributed shocks with a parameter independent of the menu.

Theorem 1 is proved in Appendix A, which also contains a family of related characterizations.[4] To illustrate the mechanics behind Theorem 1, we show how knowing a decomposable neutral rule on a single menu can pin it down for all other menus. Assume we know $\Phi(B,r)$ for (B,r) from (3), and assume that both b_0 and b_1 are chosen with positive probability. Our goal is to show how this knowledge restricts $\Phi(A,o)$ for

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

By neutrality $\Phi(A,o)_{a_1} = \Phi(A,o)_{a_2}$. We will demonstrate that $\Phi(A,o)_{a_2}$ and $\Phi(A,o)_{a_3}$ satisfy a certain identity. Consider the product of (A,o) with the n-fold product of

[4] We are grateful to Gabriel Carroll and Marcin Pęski who suggested the current proof strategy, simplifying our original proof considerably.

(B, r):

$$(A, o) \otimes \underbrace{(B, r) \otimes (B, r) \otimes \ldots \otimes (B, r)}_{n \text{ times}}, \tag{4}$$

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 neutrality. Therefore, decomposability implies

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

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

We now discuss the role of the technical continuity requirement in Theorem 1. To argue that $\Phi(B, r) = \mathrm{MNL}(B, r)$ implies $\Phi(A, o) = \mathrm{MNL}(A, o)$, we did not invoke continuity. Indeed, if we focus on choice rules defined only on menus (A, o) with integer-valued or rational-valued outcomes o, the continuity assumption can be dropped; see Propositions 2 and 3 in Appendix A. In this setting, in addition to multinomial logit, we also get the two limiting rules obtained by letting β go to $+\infty$ or $-\infty$. These rules denoted by $\mathrm{MNL}^{+\infty}$ and $\mathrm{MNL}^{-\infty}$ output the uniform distribution over the highest-reward or lowest-cost actions

$$\mathrm{MNL}^{+\infty}(A, o)_a = \begin{cases} |\operatorname{argmax}_{b \in A} o(b)|^{-1}, & a \in \operatorname{argmax}_{b \in A} o(b) \\ 0, & \text{otherwise} \end{cases}$$

and

$$\mathrm{MNL}^{-\infty}(A, o)_a = \mathrm{MNL}^{+\infty}(A, -o)_a.$$

These rules do not withstand the continuity test and thus do not appear in Theorem 1. Another way to exclude these rules is to require positivity, which postulates that every action is chosen with positive probability.

Axiom 4 (Positivity)**.** *A rule Φ is* ***positive*** *if $\Phi(A, o)_a > 0$ for all menus $(A, o) \in \mathcal{M}$ and actions $a \in A$.*

Replacing continuity with positivity results in pathological non-measurable everywhere discontinuous rules, in addition to multinomial logit; see discussion in §4.2

below. Characterizing decomposable neutral rules lacking both positivity and continuity remains an open question. We conjecture that no such rule is measurable unless it coincides with some multinomial logit rule.

4 Decomposable Rules for General Outcome Spaces $\mathcal{O}$

So far, we have focused on the outcome space $\mathcal{O} = \mathbb{R}$ capturing a decision-maker choosing actions based on simple numerical outcomes such as monetary rewards or costs. It turns out that neither the one-dimensional structure of the outcome space nor the additive structure of unrelated outcomes is critical for our analysis.

Here, we extend the analysis to a more sophisticated decision-maker whose action choices are driven by outcomes in a general space $\mathcal{O}$. For example, $\mathcal{O} = \mathbb{R}^2$ can be used to model a decision-maker whose choices are affected by two numbers, e.g., reward and the cost of the chosen action, utility today and tomorrow, or utilities in two different states of nature. The set $\mathcal{O} = \{\text{bounded continuous } f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}\}$ can represent a decision-maker who cares about infinite payoff streams. A risk-sensitive decision-maker whose actions result in a lottery, and who bases her decisions on the mean and the standard deviation only, can be modeled with $\mathcal{O} = \mathbb{R} \times \mathbb{R}_{\geqslant 0}$. If the whole distribution matters, $\mathcal{O} = \Delta(\mathbb{R})$.

To capture the aforementioned examples, we will impose minimal assumptions on $\mathcal{O}$. The notions of a menu (A, o), a collection of menus $\mathcal{M}$, and a stochastic choice rule Φ extended straightforwardly. Similarly, the axioms—neutrality, positivity, and continuity (for $\mathcal{O}$ endowed with a topology)—need no modifications.

To motivate the extension of decomposability to a general outcome space, we consider the following example. Let $\mathcal{O} = \mathbb{R} \times \mathbb{R}_{\geqslant 0}$, where the first component is interpreted as the mean and the second component as the standard deviation of a stochastic monetary reward. In the following menu, the decision-maker compares two investment decisions that differ substantially by expected rewards and, even more dramatically, by their standard deviation

$$(A, o) = \left\{ \begin{matrix} \text{bonds} & \text{crypto} \\ \binom{5}{2} & \binom{20}{100} \end{matrix} \right\}. \tag{5}$$

We assume that $\mathcal{O}$ is endowed with a binary operation "$*$" corresponding to combining outcomes of actions with stochastically independent outcomes. For example, consider (A, o) from (5) and suppose that there is another investment opportunity with rewards that are independent of those from A, say, a lottery ticket with outcome $\binom{-0.01}{0.05}$. We can construct a new menu where the decision-maker compares two options: (i) buying the lottery ticket and investing in bonds or (ii) buying the ticket and investing in crypto:

$$\left\{ \begin{array}{cc} \text{(lottery, bonds)} & \text{(lottery, crypto)} \\ \binom{-0.01+5}{\sqrt{0.05^2+0.1^2}} & \binom{-0.01+20}{\sqrt{0.05^2+100^2}} \end{array} \right\}.$$

The outcomes are defined this way since the sum of two independent random variables with mean m_1 and m_2 and standard deviations σ_1 and σ_2 has mean $m_1 + m_2$ and standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$. Accordingly, for $\mathcal{O}$ representing the mean/standard-deviation pairs, the binary operation is naturally defined as follows

$$\begin{pmatrix} m_1 \\ \sigma_1 \end{pmatrix} * \begin{pmatrix} m_2 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} m_1 + m_2 \\ \sqrt{\sigma_1^2 + \sigma_2^2} \end{pmatrix}. \tag{6}$$

For a general outcome space $\mathcal{O}$, we assume that it is endowed with a binary operation $*$ corresponding to combining outcomes of unrelated actions. The existence of an operation $*$ is justified by the interpretation of the space of outcomes $\mathcal{O}$. Indeed, an outcome of an action captures all information about it relevant to the decision-maker. Every action in every menu, including combined ones, must be assigned an outcome. An outcome assigned to a combination of unrelated actions must be a function of their outcomes as the latter capture all relevant information. This function $\mathcal{O} \times \mathcal{O} \to \mathcal{O}$ is the operation $*$.

We impose the following mild requirements on $*$.

1. **Existence of an irrelevant outcome.** There exists $e \in \mathcal{O}$ such that

   $$e * x = x * e$$

   for any $x \in \mathcal{O}$. The decision-maker does not care about an action a having outcome e in the sense that combining a with any other action b with outcome x does not change the decision-maker's perception of b.

2. **Possibility of compensation.** For any pair of outcomes $x, x' \in \mathcal{O}$,

   $$\text{either} \quad \exists s \in \mathcal{O}: \quad x' = s * x \qquad \text{or} \quad \exists s' \in \mathcal{O}: \quad x = s' * x'.$$

   In other words, any two actions with outcomes x and x' can be bridged by combining one of them with an outcome of some other action.

Abusing notation, we call the pair $(\mathcal{O}, *)$ an outcome space in what follows.

Our analysis from §3 pertains to the outcome space $(\mathcal{O}, *) = (\mathbb{R}, +)$, which satisfies these requirements. More generally, our requirements on $*$ are satisfied whenever $(\mathcal{O}, *)$ is a group. The operation $*$ defined in (6) also satisfies the above requirements (even though in this case $(\mathcal{O}, *)$ is not a group), and is also commutative and associative. However, we assume neither commutativity nor associativity. In particular, hearing "good news" and then "bad news" may not be the same as "bad news" and then "good news" from the decision-maker's perspective.

Since $*$ corresponds to combining outcomes of unrelated actions, the definition of a product menu extends straightforwardly. Given two menus (A_1, o_1) and (A_2, o_2), we define their product $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$ by

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

As in the case of $(\mathcal{O}, *) = (\mathbb{R}, +)$, the product menu corresponds to combining the two unrelated choices (or as if they were unrelated).

Once the product menus are defined, the requirement of decomposability (Axiom 2) applies to choice rule with arbitrary outcome spaces $(\mathcal{O}, *)$. Indeed, consider a rule Φ defined on a collection $\mathcal{M}$ of menus. This rule is decomposable if

$$\Phi\big((A_1, o_1) \otimes (A_2, o_2)\big)_{(a_1,a_2)} = \Phi(A_1, o_1)_{a_1} \cdot \Phi(A_2, o_2)_{a_2}$$

for all $(A_1, o_1), (A_2, o_2) \in \mathcal{M}$.

In the context of the example of $\mathcal{O} = \mathbb{R} \times \mathbb{R}_{\geqslant 0}$ above with the operation (6), decomposability means that the choice between bonds and crypto is unaffected by the presence of a lottery, and thus captures a certain indifference to wealth effects or background risk. Of course, whether this assumption is justified depends on the context.

As another example of $(\mathcal{O}, *)$ consider the following setting of choice under ambiguity. Let Θ be a finite set of states. An outcome x is a function $x\colon \Theta \to \mathbb{R}$, specifying reward or cost for each of the states, so that x is a Savage act and $\mathcal{O} = \mathbb{R}^\Theta$.

The decision-maker is ambiguous about the state $\theta \in \Theta$ and so she may take into account all the possible values x_θ. For a given state, similarly to §3, monetary rewards or costs are assumed to be additive over unrelated actions. Thus the operation $*$ is component-wise addition, or simply addition in $\mathbb{R}^\Theta$. Here, decomposability captures the idea that there is a true (but unknown) state, which is fixed across all decision problem, and so the decision maker compares rewards state by state.

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

$$u(s * t) = u(s) + u(t) \quad \text{for all} \quad s, t \in \mathcal{O}. \tag{7}$$

In other words, a utility representation assigns a numerical value to each outcome so that combining outcomes of unrelated actions corresponds to summing their utilities. For example, if $(\mathcal{O}, *) = (\mathbb{R}, +)$, a linear function $u(t) = \beta \cdot t$ is a utility representation; if $(\mathcal{O}, *)$ is the mean-variance outcome space (6), a function $u(m, \sigma) = \beta \cdot m + \gamma \cdot \sigma^2$ provides a utility representation; and if $(\mathcal{O}, *) = (\mathbb{R}^\Theta, +)$ as in the choice under ambiguity example, $u(x) = \sum_\theta q_\theta \cdot x_\theta$ is a utility representation. As discussed below, these are all continuous utility representations for these outcome spaces. Our second main result relates decomposability and utility representations.

Theorem 2. *Let Φ be a neutral, decomposable, positive stochastic choice rule for an outcome space $(\mathcal{O}, *)$. Then, there exists a utility representation u of $(\mathcal{O}, *)$ such that*

$$\Phi(A, o)_a = \frac{\exp\Big(u\big(o(a)\big)\Big)}{\sum_{b \in A} \exp\Big(u\big(o(b)\big)\Big)} \tag{8}$$

for any menu $(A, o) \in \mathcal{M}$. The utility representation u is continuous if and only if Φ is continuous.

Informally, Theorem 2 says that there is a canonical way to assign utilities to elements of the outcome space so that choices are governed by a multinomial logit rule with respect to these utilities and summing utilities corresponds to combining unrelated actions. The constant β in the multinomial logit is normalized to one, as any other constant can always be absorbed by u.

We conclude that, no matter how sophisticated the decision-maker is, she behaves like a rational stochastic utility maximizer with a very particular form of stochasticity leading to the multinomial logit distribution. This conclusion can be surprising as we assumed neither that the decision maker is a utility maximizer nor that her choices are driven by a numerical characteristic of actions.

Theorem 2 also implies that a sophisticated decision-maker using $\mathcal{O}$ with non-commutative or non-associative operation $*$ behaves as if the composition operation was commutative and associative. Indeed, any utility representation u of $(\mathcal{O}, *)$ satisfies $u(x * y) = u(x) + u(y) = u(y * x)$ and similarly $u(x * (y * z)) = u((x * y) * z))$. Hence, the non-commutative or non-associative dimensions of $(\mathcal{O}, *)$ will all be in the kernel of the utility representation and will not affect decision-making.

Theorem 2 is proved in Appendix B. The idea is to extract u from Φ as follows. For each element $x \in \mathcal{O}$, consider a binary menu $(A, o_x) \in \mathcal{M}$ with $A = \{a, b\}$, $o_x(a) = e$ and $o(b) = x$. We define u by

$$u(x) = \ln\left(\frac{\Phi(A, o_x)_b}{1 - \Phi(A, o_x)_b}\right) \tag{9}$$

so that Φ is given by the multinomial logit formula (8) on binary menus (A, o_x). By decomposability of Φ on binary menus, u satisfies the generalized Cauchy equation (7), i.e., u is a utility representation. The multinomial logit formula (8) is extended from binary menus to all menus by decomposability and neutrality. Positivity of Φ in Theorem 2 ensures that u given by (9) is well-defined. Dropping positivity is an open problem requiring a new proof technique already for the case of $(\mathcal{O}, *) = (\mathbb{R}^2, +)$.

4.1 Applications

Theorem 2 can be refined for those outcome spaces $\mathcal{O}$, where the set of all utility representations u admits a simple characterization. Describing all u boils down to understanding solutions to the generalized Cauchy equation (7)

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

We will see a family of economically-relevant examples below.

The first few examples correspond to the case of a linear space $\mathcal{O}$, equipped with the operation of addition, e.g., $(\mathcal{O}, *) = (\mathbb{R}^d, +)$. For such $\mathcal{O}$, the Cauchy equation becomes

$$u(x + y) = u(x) + u(y). \tag{10}$$

Lemma 1. *Suppose that the outcome space $(\mathcal{O}, *)$ is a Banach space equipped with the operation of addition. Then, any utility representation $u\colon \mathcal{O} \to \mathbb{R}$ continuous at some $x_0 \in \mathcal{O}$ is a linear map continuous at all $x \in \mathcal{O}$. Any utility representation discontinuous at some x_0 is discontinuous everywhere and non-measurable.*

Measurability in Lemma 1 is understood in the usual sense, i.e., with respect to the Borel σ-algebra of $\mathcal{O}$. This lemma is a folk result in the theory of functional equations. For $(\mathcal{O}, *) = (\mathbb{R}, +)$, it dates back to Cauchy. For a general Banach space $(\mathcal{O}, *)$ see, e.g., (Kuczma, 2009; Jung, 2011).

4.2 Choice driven by monetary rewards

As a first simple example, consider the familiar case of a decision maker comparing actions by their monetary rewards or costs, i.e., $(\mathcal{O}, *) = (\mathbb{R}, +)$ as in §3. By Lemma 1, any continuous utility representation $u\colon \mathbb{R} \to \mathbb{R}$ is linear, i.e., $u(x) = \beta \cdot x$ for some $\beta \in \mathbb{R}$. Moreover, it suffices to require continuity at a single point x_0.

Combining this insight with Theorem 2, we get that any continuous decomposable positive rule Φ is a multinomial logit rule MNL^{β}. In other words, we obtain the conclusion of Theorem 1 under an additional assumption of positivity. While the result is weaker because of the redundant positivity assumption, this proof technique highlights that we only need to require continuity at a single menu (with at least two actions).

By Lemma 1, utility representations discontinuous at one point are necessarily discontinuous everywhere and are moreover non-measurable. We conclude that so are discontinuous decomposable positive choice rules. Such rules can be obtained using discontinuous solutions to the Cauchy equation. Similarly to the existence of a discontinuous solution to (10), this construction relies on Hamel bases and thus requires an explicit use of the axiom of choice. To conclude, discontinuous rules are not more than a technical curiosity.

4.3 Choice under ambiguity

Recall again the example in which Θ is a finite set of states, $(\mathcal{O}, *) = (\mathbb{R}^{\Theta}, +)$, and each $x \in \mathcal{O}$ is an act assigning a monetary reward in each state.

By Lemma 1, any continuous utility representation $u\colon \mathbb{R}^{\Theta} \to \mathbb{R}$ is linear, and so can be written as

$$u(x) = \beta \cdot \sum_{\theta \in \Theta} p_\theta \cdot (-1)^{\sigma_\theta} \cdot x_\theta$$

for some $\beta \geqslant 0$, $\sigma\colon \Theta \to \{0, 1\}$, and $p \in \Delta(\Theta)$, where $\Delta(\Theta)$ denotes the set of probability distributions over Θ. The distribution p can be interpreted as the decision-maker's prior over the states, and σ_θ determines whether the component x_θ is treated as a reward or a cost.

By Theorem 2, any decision maker whose choices follow a continuous decomposable positive rule Φ, behaves as a stochastic expected utility maximizer with some prior p over the set of states and Gumbel-distributed shocks:

$$\Phi(A, o)_a = \frac{\exp\left(\beta \cdot \sum_{\theta\in\Theta} p_\theta \cdot (-1)^{\sigma_\theta} \cdot o(a)_\theta\right)}{\sum_{b\in A} \exp\left(\beta \cdot \sum_{\theta\in\Theta} p_\theta \cdot (-1)^{\sigma_\theta} \cdot o(b)_\theta\right)}. \tag{11}$$

To rule out the situation where some components of x are treated as rewards and some as costs, one can impose a simple monotonicity requirement. If the choice probability is non-decreasing in the θ-component of the outcome, one can assume $\sigma_\theta = 0$ in (11). Moreover, it suffices to require this monotonicity at a single menu with at least two actions.

4.4 Intertemporal choice

Consider a decision-maker who chooses an action taking into account the stream of payoffs $x\colon T \to \mathbb{R}$ that it generates, where $T = \mathbb{R}_{\geqslant 0}$ is set of time periods. We assume that these payoffs x are continuous and stop after some point in time. Hence, the space of outcomes $\mathcal{O}$ is the set of all continuous and compactly-supported functions $x\colon T \to \mathbb{R}$. We equip $\mathcal{O}$ with the operation of addition and the topology induced by the sup-norm. By the Riesz representation theorem (see Folland, 1999, Theorem 7.17), any continuous linear functional on $\mathcal{O}$ can be represented as integration against a finite signed measure μ on T. Lemma 1 implies that any continuous utility representation u of $(\mathcal{O}, +)$ has the following form

$$u(x) = \int_T x(t)\, \mathrm{d}\mu(t)$$

for some finite signed measure μ on T. Thus, by Theorem 2, any decision maker whose choices follow a continuous decomposable positive rule Φ is given by

$$\Phi(A, o)_a = \frac{\exp\left(\int_T o(a)_t\, \mathrm{d}\mu(t)\right)}{\sum_{b\in A} \exp\left(\int_T o(b)_t\, \mathrm{d}\mu(t)\right)}. \tag{12}$$

As above, an additional monotonicity assumption yields that μ is a positive measure.

4.5 Risk-sensitive choice

Now, suppose that the actions generate stochastic payoffs. In the simplest case discussed above, the decision-maker cares only about the mean m and the standard deviation $\sigma \geqslant 0$ of the payoff and thus the outcome space is $\mathcal{O} = \mathbb{R} \times \mathbb{R}_{\geqslant 0}$ with the operation $*$ defined in (6). It is easy to see that any continuous utility representation $u\colon \mathbb{R} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is a mean-variance utility

$$u(m, \sigma) = \gamma_1 \cdot m + \gamma_2 \cdot \sigma^2. \tag{13}$$

Indeed, any continuous utility representation u defines a continuous solution v to the additive Cauchy equation (10) by

$$v(y_1, y_2) = u\big(y_1, \sqrt{y_2 + |y_2|}\big) - u\big(0, \sqrt{|y_2|}\big).$$

By Lemma 1, v is linear, and thus u has the desired form.

Theorem 2, implies that any continuous decomposable positive rule Φ is given by

$$\Phi(A, o)_a = \frac{\exp\Big(\gamma_1 \cdot m(a) + \gamma_2 \cdot \sigma(a)^2\Big)}{\sum_{b \in A} \exp\Big(\gamma_1 \cdot m(b) + \gamma_2 \cdot \sigma(b)^2\Big)}. \tag{14}$$

More generally, consider a decision-maker who takes into account the entire payoff distribution, and thus an outcome x is the element of $\Delta(\mathbb{R})$. Let $\mathcal{O}$ be the set of all probability distributions on $\mathbb{R}$, with finite n moments for some $n \geqslant 0$. The operation $*$ is the convolution, which corresponds to summing independent random variables. The topology is defined so that the n moments are continuous functionals on $\mathcal{O}$: the distance between two distributions x and y is given by the weighted total variation distance

$$d(x, y) = \int_{\mathbb{R}} (1 + |t|)^n \mathrm{d}|x - y|(t) \tag{15}$$

so that the k moments are continuous functionals on $\mathcal{O}$. The following result is a direct corollary of Mattner (2004).

Lemma 2. *Let the outcome space $(\mathcal{O}, *)$ be the set of all probability distributions with $n \geqslant 0$ finite moments, the operation of convolution, and topology induced by (15). Then, any continuous utility representation $u\colon \mathcal{O} \to \mathbb{R}$ has the following form*

$$u(x) = \sum_{l=1}^{n} \gamma_l \kappa_l(x), \tag{16}$$

where κ_l is the l-th cumulant of x and $\gamma_1, \ldots, \gamma_n$ are some fixed real numbers.

Cumulants are additive, i.e., $\kappa_l(x * y) = \kappa_l(x) + \kappa_l(y)$, which is why representation (16) satisfies the generalized Cauchy equation (7). The first two cumulants are the mean and the variance, and so (16) extends (13).[5]

By Theorem 2, any continuous decomposable positive rule Φ has the following form

$$\Phi(A, o)_a = \frac{\exp\left(\sum_{l=1}^{n} \gamma_l \kappa_l(o(a))\right)}{\sum_{b \in A} \exp\left(\sum_{l=1}^{n} \gamma_l \kappa_l(o(b))\right)}.$$

An interesting corollary of Lemma 2 is that in the case $n = 0$, in which $\mathcal{O} = \Delta(\mathbb{R})$ endowed with the total-variation distance, there are no continuous utility representations $u \colon \mathcal{O} \to \mathbb{R}$ except for $u \equiv 0$. Thus, the only continuous decomposable positive rules are those that pick an action uniformly at random.

4.6 Finite streams of prizes

We next consider an example of an outcome space with a non-commutative operation $*$. Let P be a set of prize types, and let $\mathcal{O}$ be the set of finite streams of prizes. Formally, $\mathcal{O}$ is the set of finite sequences $x = (x_1, \ldots, x_n)$ with elements in P. The operation $*$ is concatenation:

$$(x_1, \ldots, x_m) * (y_1, \ldots, y_n) = (x_1, \ldots, x_m, y_1, \ldots, y_n).$$

Note that this operation is not commutative: receiving the stream x and then y is not the same as receiving y and then x.

In the context of this operation, exploring decomposability corresponds to studying decision-makers whose choice probabilities are invariant to previously received streams. A standard argument shows that every utility representation $u \colon \mathcal{O} \to \mathbb{R}$ is of the form

$$u(z_1, \ldots, z_k) = \sum_{p \in P} w(p) \cdot |\{l \,:\, z_l = p\}|.$$

[5]Recall that the cumulants of a distribution x are defined by the following formula

$$\kappa_l(x) = \left(\frac{1}{i^l} \cdot \frac{d^l}{d\alpha^l} \log \int_{\mathbb{R}} \exp\left(i\alpha t\right) dx(t)\right)\Big|_{\alpha=0},$$

where i is the imaginary unit.

where $w\colon P \to \mathbb{R}$ is any function. In other words, we assign a value $w(p)$ to each prize type p and define the utility of a stream as the sum of values for each of the prizes in the stream.

By Theorem 2, any decomposable positive rule Φ has the form

$$\Phi(A, o)_a = \frac{\exp\left(\sum_{p\in P} w(p)\cdot |\{l \,:\, o(a)_l = p\}|\right)}{\sum_{b\in A}\exp\left(\sum_{p\in P} w(p)\cdot |\{l \,:\, o(b)_l = p\}|\right)}$$

for some $w\colon P \to \mathbb{R}$. This example highlights that in non-commutative cases, the choice probabilities can only be driven by commutative components of $\mathcal{O}$, which in this case correspond to counting the number of appearances of each prize type in a stream.

4.7 Matrices as outcomes

In this section, we discuss a mathematical example which does not admit a natural economic interpretation but nevertheless highlights an interesting application of Theorem 2.

Suppose $\mathcal{O}$ is the set of all non-degenerate (i.e., invertible) $n \times n$ matrices with the operation $*$ given by the matrix product. The next lemma follows from Chamberlin and Wolfe (1953).

Lemma 3. *Let the outcome space $(\mathcal{O}, *)$ be the set of non-degenerate $n \times n$ matrices with the operation of multiplication. Then, any continuous utility representation $u\colon \mathcal{O} \to \mathbb{R}$ has the following form*

$$u(x) = \beta \cdot \ln\left|\det x\right| \tag{17}$$

for some $\beta \in \mathbb{R}$.

The determinant is multiplicative, i.e., $\det(x * y) = \det(x)\cdot\det(y)$, and thus the utility representation (17) satisfies (7). We conclude that any continuous decomposable positive rule Φ takes the form

$$\Phi(A, o)_a = \frac{\left|\det o(a)\right|^{\beta}}{\sum_{b\in A}\left|\det o(b)\right|^{\beta}}.$$

5 Framing effects and approximate neutrality

Our neutrality axiom captures a weak sense in which outcomes, rather than action labels, are what drives choice. In this section, we relax this assumption, allowing for framing effects which affect choice probabilities even when there are no differences in outcomes. We show that under an appropriate notion of approximate neutrality, our results remain the same.

Fix an arbitrary outcome space $\mathcal{O}$ and consider a rule Φ defined on a collection of menus $\mathcal{M}$. We now formulate the approximate versions of our main axioms.

Axiom 5 (Approximate Neutrality)**.** *A rule Φ is **approximately neutral** with parameter $\varepsilon_{\mathrm{neut}} \geqslant 0$ if for any menu (A, o) in $\mathcal{M}$ and any $a, a' \in A$ such that $o(a) = o(a')$ it holds that $\Phi(A, o)_a \leqslant (1 + \varepsilon_{\mathrm{neut}})\Phi(A, o)_{a'}$.*

Approximate neutrality is a natural relaxation of neutrality which allows actions with the same outcome to be chosen with different probabilities, but limits the ratio between them. For zero values of $\varepsilon_{\mathrm{neut}}$ we obtain the familiar requirement of exact neutrality (Axiom 1).

The main result of this section shows that under decomposability, approximate neutrality is the same as exact neutrality.

Proposition 1. *For any outcome space $(\mathcal{O}, *)$, every decomposable rule Φ satisfying approximate neutrality with any parameter $\varepsilon_{\mathrm{neut}} > 0$ is neutral.*

In other words, decomposability and approximate neutrality imply exact neutrality.[6] As a corollary, we can relax neutrality to approximate neutrality in Theorems 1 and 2 without altering the conclusion. Note that in the hypothesis of Proposition 1 there is no assumption that $\varepsilon_{\mathrm{neut}}$ is small. Moreover, the proof contained in Appendix C demonstrates that the same conclusion holds even for $\varepsilon_{\mathrm{neut}}$ that is allowed to depend on the menu, as long as it grows sub-linearly with the size of the menu.

6 Approximate decomposability

Real decision-makers may violate exact decomposability. For example, the choice of cereal and shirt may be negatively correlated because the cognitive effort required

[6] A similar phenomenon in the theory of functional equations is known as superstability (see, e.g., Jung, 2011, Chapter 10). An equation is superstable if an approximate solution is always exact.

to choose a cereal reduces the quality of the choice of shirt. In this section, we relax decomposability in a similar way to our relaxation of neutrality in the previous section. We explore the robustness of our results to approximate decomposability, and show that when approximate decomposability holds, then so do approximate versions of Theorems 1 and 2. The conclusions allow for approximate decomposability and approximate neutrality simultaneously.

Axiom 6 (Approximate Decomposability). *A rule Φ is **approximately decomposable** with parameter $\varepsilon_{\text{decomp}} \geqslant 0$ if for all menus $(A_1, o_1), (A_2, o_2) \in \mathcal{M}$ and their product (A, o), it holds that*

$$\Phi(A,o)_{(a_1,a_2)} \leqslant (1+\varepsilon_{\text{decomp}}) \cdot \Phi(A_1,o_1)_{a_1} \cdot \Phi(A_2,o_2)_{a_2}$$
$$\Phi(A_1,o_1)_{a_1} \cdot \Phi(A_2,o_2)_{a_2} \leqslant (1+\varepsilon_{\text{decomp}}) \cdot \Phi(A,o)_{(a_1,a_2)}.$$

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

6.1 Robustness for $\mathcal{O} = \mathbb{R}$

Consider the benchmark case of §3, where the outcome space is $\mathcal{O} = \mathbb{R}$ with the operation of addition.

A rule Φ on $\mathcal{M}$ is δ-close to a multinomial logit with parameter $\beta \in \mathbb{R}$ if for any menu $(A, o) \in \mathcal{M}$ there is a function $s\colon A \to [-\delta, \delta]$ such that

$$\Phi(A,o)_a = \frac{\exp(\beta \cdot o(a) + s(a))}{\sum_{b \in A} \exp(\beta \cdot o(b) + s(b))}.$$

The utility shock s may depend on the menu (A, o) and can capture framing effects.

Theorem 3. *Let Φ be a continuous stochastic choice rule for the outcome space $\mathcal{O} = \mathbb{R}$ satisfying approximate neutrality and approximate decomposability with parameters $\varepsilon_{\text{neut}} \geqslant 0$ and $\varepsilon_{\text{decomp}} \geqslant 0$, respectively. Then there exists a unique $\beta \in \mathbb{R}$ such that Φ is $\varepsilon_{\text{decomp}}$-close to a multinomial logit rule with parameter $\beta \in \mathbb{R}$.*

We see that rules satisfying our approximate axioms are close to those that satisfy the exact ones, i.e., to multinomial logit. We note that the closeness of Φ to logit does not depend on the approximation parameter in the neutrality axiom. This effect has the same origin as in Proposition 1.

The theorem is proved in Appendix D. The idea is to consider an auxiliary rule Υ defined by

$$\Upsilon(A, o)_a = \lim_{n\to\infty} \sqrt[n]{\Phi(A, o)^{\otimes n}_{(a,\ldots,a)}},$$

where $(A, o)^{\otimes n}$ denotes the product of (A, o) with itself n times. It turns out that Υ satisfies exact decomposability and neutrality and is close to Φ by construction. Thus Theorem 3 follows from applying the characterization of decomposable neutral rules to Υ.

6.2 Robusteness for general outcome spaces $\mathcal{O}$

Theorem 2 for general outcome space $\mathcal{O}$ shows that decomposable positive neutral Φ coincides with multinomial logit for some additive utility function $u\colon \mathcal{O} \to \mathbb{R}$. The possibility to replace neutrality and decomposability with their approximate versions is tightly related to the so-called Ulam stability of the Cauchy functional equation.

Consider the Cauchy equation for the outcome space $(\mathcal{O}, *)$

$$u(x * y) = u(x) + u(y) \quad \text{for all} \quad x, y \in \mathcal{O}. \tag{18}$$

Informally, the equation is stable if any approximate solution is close to the exact one. Formally, consider $\varepsilon > 0$ and suppose that there is $\delta > 0$ such that for any function w satisfying

$$\big|w(x * y) - w(x) + w(y)\big| \leqslant \varepsilon \quad \text{for all} \quad x, y \in \mathcal{O}. \tag{19}$$

there is a solution u of (18) such that

$$|u(x) - w(x)| \leqslant \delta \quad \text{for all} \quad x \in \mathcal{O}.$$

If such a pair (ε, δ) exists, equation (18) is called (ε, δ)-stable. If, for every $\delta > 0$, there is an $\varepsilon > 0$ such that the equation is (ε, δ)-stable, then we say that the equation is Ulam stable. This is an upper-hemicontinuity-type condition for the set of solutions to (19) as ε tends to zero.

We note that whether or not the Cauchy equation is stable is a property of the space $(\mathcal{O}, *)$. For example, $\mathcal{O} = \mathbb{R}$ with addition and, more generally, any Banach space $\mathcal{O}$ are $(\varepsilon, \varepsilon)$-stable for any $\varepsilon > 0$ as conjectured by Ulam and proved by Hyers (1941); see a survey by Hyers and Rassias (1992) for other examples.

Similarly to the previous section, we say that a rule Φ is δ-close to multinomial logit if there is a utility representation $u\colon \mathcal{O} \to \mathbb{R}$ and, for any menu $(A, o) \in \mathcal{M}$, there is a function $s\colon A \to [-\delta, \delta]$ such that

$$\Phi(A, o)_a = \frac{\exp\big(u\big(o(a)\big) + s(a)\big)}{\sum_{b \in A} \exp\big(u\big(o(b)\big) + s(b)\big)}.$$

We stress that u in this definition solves the exact Cauchy equation (18), not the approximate one (19).

Theorem 4. *Consider an outcome space $(\mathcal{O}, *)$ such that the Cauchy equation (18) is Ulam stable. Then, for every $\delta > 0$ there exists an $\varepsilon_{\text{decomp}} > 0$ such that every positive, $\varepsilon_{\text{neut}}$-neutral, $\varepsilon_{\text{decomp}}$-decomposable, stochastic choice rule Φ (with any $\varepsilon_{\text{neut}} \geqslant 0$) is δ-close to multinomial logit.*

Theorem 4 is a particular case of a more general result relating δ to $\varepsilon_{\text{decomp}}$; see Theorem 5 in Appendix E.

Recall that the examples discussed in §4—ambiguous monetary rewards (11), inter-temporal choice (12), and risk sensitive choice driven by mean and variance of rewards (14)—correspond to a Banach space $(\mathcal{O}, *)$. Combining Theorem 5 with the result of Hyers (1941), we get that for any Banach space $\mathcal{O}$, any positive $\varepsilon_{\text{neut}}$-neutral $\varepsilon_{\text{decomp}}$-decomposable Φ is $\left(10\varepsilon_{\text{decomp}} + 2\varepsilon_{\text{decomp}}^2\right)$-close to multinomial logit. Similarly to Theorem 3, this bound does not depend on the value of $\varepsilon_{\text{neut}}$. Hyers (1941) also demonstrates that if an approximate solution w to the Cauchy equation is continuous for at least one point, then it is close to a globally continuous exact solution u. Thus, under an additional assumption of continuity of Φ, the utility representation u is a continuous linear functional on $\mathcal{O}$ since such functionals exhaust continuous solutions of the Cauchy equation on a Banach space. Consequently, the functional forms for a rule Φ established in (11), (12), and (14) are robust to replacing neutrality and decomposability with their approximate versions.

7 Conclusion

This paper explores a novel approach to stochastic choice and study decomposability as our main assumption. We show that in very general settings, decision-makers who satisfy decomposability choose using multinomial logit applied to a utility function.

An interesting direction for future research is to relax decomposability and only require that in a product menu, the choice distribution has the same marginals as in the component menus, but to drop the independence requirement; e.g., allow the choice of shirts and cereals to be correlated, but require that the probability that a particular shirt is chosen does not change when this is considered in conjunction with a choice of cereal (see the marginality condition proposed by Chambers, Masatlioglu, and Turansick, 2021). This relaxation allows for choice rules beyond multinomial logit, such as mixed logit rule obtained via averaging multinomial logits with different parameters. We conjecture that there are no other neutral continuous rules satisfying this condition.

References

M. Agranov and P. Ortoleva. Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68, 2017.

M. Agranov and P. Ortoleva. Revealed preferences for randomization: An overview. In *AEA Papers and Proceedings*, volume 112, pages 426–430. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2022.

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

Y. Azrieli, C. P. Chambers, and P. J. Healy. Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503, 2018.

N. Barberis, M. Huang, and R. H. Thaler. Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American economic review*, 96(4):1069–1090, 2006.

G. M. Becker, M. H. Degroot, and J. Marschak. Probabilities of choices among very similar objects: An experiment to decide between two models. *Behavioral science*, 8(4):306–311, 1963.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

H. D. Block and J. Marschak. Random orderings and stochastic theories of response. 1959.

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

S. Cerreia-Vioglio, D. Dillenberger, P. Ortoleva, and G. Riella. Deliberately stochastic. *American Economic Review*, 109(7):2425–2445, 2019.

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. doi: 10.1093/restud/rdac046. URL https://doi.org/10.1093/restud/rdac046. rdac046.

E. Chamberlin and J. Wolfe. Multiplicative homomorphisms of matrices. *Proceedings of the American Mathematical Society*, 4(1):37–42, 1953.

C. P. Chambers, Y. Masatlioglu, and C. Turansick. Correlated choice. *arXiv preprint arXiv:2103.05084*, 2021.

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

F. Echenique, K. Saito, and G. Tserenjigmid. The perception-adjusted luce model. *Mathematical Social Sciences*, 93:67–76, 2018.

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

R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat*, volume 1. Basic books, 2011.

G. B. Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

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.

F. Gul, P. Natenzon, and W. Pesendorfer. Random choice as behavioral optimization. *Econometrica*, 82(5):1873–1912, 2014.

D. H. Hyers. On the stability of the linear functional equation. *Proceedings of the National Academy of Sciences*, 27(4):222–224, 1941.

D. H. Hyers and T. M. Rassias. Approximate homomorphisms. *Aequationes mathematicae*, 44(2-3):125–153, 1992.

S.-M. Jung. *Hyers-Ulam-Rassias stability of functional equations in nonlinear analysis*, volume 48. Springer Science & Business Media, 2011.

N. Kashaev, C. Gauthier, and V. H. Aguiar. Dynamic and stochastic rational behavior. *arXiv preprint arXiv:2302.04417*, 2023.

M. Kuczma. An introduction to the theory of functional equations and inequalities. *(No Title)*, 2009.

L. D. Landau and E. M. Lifshitz. *Statistical Physics: Volume 5*, volume 5. Elsevier, 1951.

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.

M. J. Machina. Stochastic choice functions generated from deterministic preferences over lotteries. *The economic journal*, 95(379):575–594, 1985.

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. Mattner. Cumulants are universal homomorphisms into hausdorff groups. *Probability theory and related fields*, 130:151–166, 2004.

L.-G. Mattsson and J. W. Weibull. Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior*, 41(1):61–78, 2002.

D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142, 1974.

D. McFadden. Economic choices. *American economic review*, 91(3):351–378, 2001.

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, W. B. Tye, and K. Train. *An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model*. Institute of Transportation Studies, University of California Berkeley, 1977.

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

K. Saito. Axiomatizations of the mixed logit model. 2018.

C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec*, 4(142-163):1, 1959.

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

R. Webb. The (neural) dynamics of stochastic choice. *Management Science*, 65(1): 230–255, 2019.

M. Woodford. Stochastic choice: An optimizing neuroeconomic model. *American Economic Review*, 104(5):495–500, 2014.

A Proof of Theorem 1 and Related Results for $\mathcal{O} \subset \mathbb{R}$

Recall that $\mathcal{A}$ is a fixed non-empty set of possible actions that is closed under the operation of forming ordered pairs.[7] For a subset of reals $R \subset \mathbb{R}$, denote by $\mathcal{M}_R$ the collection of all menus (A, u) with $A \subset \mathcal{A}$ and $u\colon A \to R$. Apart from $R = \mathbb{R}$, we will consider R equal to the set of integers $\mathbb{Z}$ or rational numbers $\mathbb{Q}$. In other words, $\mathcal{M}_R$ is the set of all menus with the set of outcomes $\mathcal{O} = R$. We will say that Φ is a stochastic choice rule with outcomes in R if it is defined for menus $(A, u) \in \mathcal{M}_R$ but may not be defined beyond.

Theorem 1 characterizes choice rules with $\mathcal{O} = \mathbb{R}$ that satisfy neutrality, decomposability, and continuity. We prove it in three steps. First, we formulate and prove a version of the characterization for rules with $\mathcal{O} = \mathbb{Z}$ without the continuity assumption, then deduce the result for $\mathcal{O} = \mathbb{Q}$, and finally derive the theorem from the result for $\mathbb{Q}$ by applying continuity.

Recall that the multinomial logit rule with parameter $\beta \in \mathbb{R}$ is denoted by MNL^β and is given by

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

We also consider the limiting cases for $\beta \to \pm\infty$ denoted by $\mathrm{MNL}^{+\infty}$ and $\mathrm{MNL}^{-\infty}$; these are the rules that output the uniform distribution over the highest-outcome and the lowest-outcome actions, respectively. We will refer to MNL^β with $\beta \in \mathbb{R} \cup \{\pm\infty\}$ as the generalized multinomial logit rule.

[7]Note that $\mathcal{A}$ must be infinite, since if $a \in \mathcal{A}$ then so are (a, a), $(a, (a, a))$, etc.

Proposition 2. *Let Φ be a neutral, decomposable stochastic choice rule for the outcome space $\mathcal{O} = \mathbb{Z}$. Then Φ coincides with the generalized multinomial logit rule* MNL^{β} *for some $\beta \in \mathbb{R} \cup \{\pm\infty\}$.*

Proof of Proposition 2. Consider a menu

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

and for $i \in \{0, 1\}$ denote $p_i = \Phi(B, r)_{b_i}$. Since $p_0 + p_1 = 1$, at least one of these probabilities is non-zero.

We now show that $\Phi(B, r)$ determines the outcome of $\Phi(A, o)$ for any other menu $(A, o) \in \mathcal{M}_{\mathbb{Z}}$. If all actions in A have the same outcome, then $\Phi(A, o)_a = 1/|A|$ for all $a \in A$ by neutrality. Henceforth, we focus on menus where not all outcomes are the same. We show that for any pair of actions a and a' such that $o(a) > o(a')$, the following identity holds:

$$\Phi(A, o)_a \cdot {p_0}^{o(a)-o(a')} = \Phi(A, o)_{a'} \cdot {p_1}^{o(a)-o(a')}. \tag{20}$$

Denote $n = o(a) - o(a') > 0$ and consider an auxiliary menu

$$(A, o) \otimes \underbrace{(B, r) \otimes (B, r) \otimes \ldots \otimes (B, r)}_{n \text{ times}}.$$

The actions $(a, b_0, \ldots, b_0)$ and $(a', b_1, \ldots, b_1)$ have the same outcome, $o(a)$. By neutrality, these actions are assigned the same probability by Φ. Expressing these probabilities via decomposability, we get (20).

With the help of identity (20), we obtain the following answer for $\Phi(A, o)$, depending on whether p_0 and p_1 are positive or zero.

If $p_0 = 0$, then choosing a to be an action with the highest outcome, we conclude from (20) that $\Phi(A, o)_{a'} = 0$ for any a' with $o(a') < o(a)$. Thus only the actions with the highest outcomes can be assigned a non-zero probability. By neutrality, we conclude that $\Phi(A, o)$ is the uniform distribution over actions with the highest outcomes, and thus $\Phi = \mathrm{MNL}^{+\infty}$.

Similarly, if $p_1 = 0$, we conclude that $\Phi(A, o)$ is the uniform distribution over actions with the lowest outcomes, i.e., $\Phi = \mathrm{MNL}^{-\infty}$.

Finally, consider the case where both p_0 and p_1 are non-zero. Denote $\beta = \ln(p_1/p_0)$. Let a' be an action with the lowest outcome. Denote $\gamma_{(A,o)} = \Phi(A, o)_{a'} \cdot$

$\exp(-\beta \cdot o(a'))$. Hence, for any action $a \in A$, identity (20) can be rewritten as follows

$$\Phi(A, o)_a = \gamma_{(A,o)} \cdot \exp(\beta \cdot o(a)).$$

Since $\sum_b \Phi(A, o)_b = 1$, we obtain $1 = \gamma_{(A,o)} \cdot \sum_{b \in A} \exp(\beta \cdot o(b))$. Thus

$$\Phi(A, o)_a = \frac{\exp(\beta \cdot o(a))}{\sum_{b \in A} \exp(\beta \cdot o(b))},$$

i.e., Φ is the multinomial logit rule with parameter β. □

Proposition 2 implies an analogous result for rational utilities.

Proposition 3. *Let Φ be a neutral, decomposable stochastic choice rule for the outcome space $\mathcal{O} = \mathbb{Q}$. Then Φ coincides with the generalized multinomial logit rule MNL^β for some $\beta \in \mathbb{R} \cup \{\pm\infty\}$.*

Proof. Given Φ with $\mathcal{O} = \mathbb{Q}$, consider a family of rules Φ^k, $k = 1, 2, \ldots$ with $\mathcal{O} = \mathbb{Z}$ by

$$\Phi^k(A, o) = \Phi\left(A,\ \frac{1}{k} \cdot o\right). \tag{21}$$

Each Φ^k is a neutral, decomposable rule. Hence, Proposition 2 implies that there is $\beta_k \in \mathbb{R} \cup \{\pm\infty\}$ such that Φ^k coincides with the generalized multinomial logit MNL^{β_k}. By (21),

$$\Phi^1(A, o) = \Phi^k(A,\ k \cdot o)$$

for any $(A, o) \in \mathcal{M}_\mathbb{Z}$. Thus $\beta_1 = k \cdot \beta_k$.

Consider now an arbitrary menu $(A, o) \in \mathcal{M}_\mathbb{Q}$ and let k be such that $k \cdot o$ is integer-valued. We can express $\Phi(A, o)$ as follows:

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

Multinomial logit has the following property $\mathrm{MNL}^\beta(A,\ \alpha \cdot o) = \mathrm{MNL}^{\alpha \cdot \beta}(A, o)$. Using this property and that $\beta_1 = k \cdot \beta_k$, we obtain

$$\Phi(A, o) = \mathrm{MNL}^{\beta_k}\left(A,\ k \cdot o\right) = \mathrm{MNL}^{k \cdot \beta_k}\left(A,\ o\right) = \mathrm{MNL}^{\beta_1}\left(A,\ o\right).$$

Thus Φ equals the generalized multinomial logit with parameter $\beta = \beta_1$ for every menu $(A, o) \in \mathcal{M}_\mathbb{Q}$. □

We now turn to rules for outcomes $\mathcal{O} = \mathbb{R}$ and are ready to prove Theorem 1 characterizing neutral, decomposable, continuous rules. Recall that (A, o_n) converges to (A, o) if $\lim_n o_n(a) = u(a)$ for all $a \in A$. A rule Φ is continuous if $\lim_n \Phi(A, o_n)_a = \Phi(A, u)_a$ for all $a \in A$ when $\lim_n (A, o_n) = (A, o)$.

Proof of Theorem 1. We aim to show that any stochastic choice rules Φ with $\mathcal{O} = \mathbb{R}$ satisfying neutrality, decomposability, and continuity is the multinomial logit rule MNL^β for some $\beta \in \mathbb{R}$.

By Proposition 3, we know that there exists $\beta \in \mathbb{R} \cup \{\pm\infty\}$ such that Φ coincides with MNL^β on the set of menus with rational utilities $\mathcal{M}_\mathbb{Q}$. We now demonstrate that β cannot equal $\pm\infty$ by checking that $\mathrm{MNL}^{\pm\infty}$ does not admit a continuous extension from $\mathcal{M}_\mathbb{Q}$ to $\mathcal{M}_\mathbb{R}$. We first focus on MNL^∞ and consider a sequence of menus $(A, o_n) \in M_\mathbb{Q}$ with $A = \{a_0, a_1\}$, $o_n(a_0) = 0$ and $o_n(a_1) = 1/n$. The limit menu (A, o) has zero utility for both actions, and thus $\mathrm{MNL}^{+\infty}(A, o)$ is the uniform distribution over A. However, $\mathrm{MNL}^{+\infty}(A, o_n)$ puts the whole mass on a_1 and thus $\lim_n \mathrm{MNL}^{+\infty}(A, o_n) \neq \mathrm{MNL}^{+\infty}(A, o)$. Since $\mathrm{MNL}^{-\beta}(B, r) = \mathrm{MNL}^{\beta}(B, -r)$ for any menu (B, r), discontinuity of $\mathrm{MNL}^{-\infty}$ follows from that of $\mathrm{MNL}^{+\infty}$.

We conclude that $\Phi = \mathrm{MNL}^\beta$ with $\beta \in \mathbb{R}$ for menus from the dense set $\mathcal{M}_\mathbb{Q} \subset \mathcal{M}_\mathbb{R}$. Both rules Φ and MNL^β are continuous and thus coincide on $\mathcal{M}_\mathbb{R}$. $\square$

B Proof of Theorem 2

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

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

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

Define the function $u\colon \mathcal{O} \to \mathbb{R}$ by

$$u(x) = \ln \frac{p_x}{1 - p_x}. \tag{22}$$

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

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

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

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

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

Computing their outcomes, we get

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

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

$$\Phi(B, o)_b = \Phi(B, o)_{b'}.$$

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

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

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

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

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

We now consider an arbitrary menu (A, o) and demonstrate that $\Phi(A, o)$ is given by the multinomial logit with constructed utility function (8). Let $a, b \in A$ be two distinct actions. By the assumption 2 on the operation $*$, there is $x \in \mathcal{O}$ such that $o(a) = x * o(b)$, or there is $y \in \mathcal{O}$ such that $o(b) = y * o(a)$, or both. Without loss of generality, we assume that $o(a) = x * o(b)$. Consider the product menu

$$(A_x, o_x) \otimes (A, o).$$

This menu contains two actions

$$(a_e, a) \qquad \text{and} \qquad (a_x, b)$$

with equal outcomes

$$e * o(a) = o(a) \qquad \text{and} \qquad x * o(a') = o(a).$$

By neutrality, Φ assigns equal probabilities to (a_e, a) and (a_x, b). Applying decomposability, we obtain the following identity

$$(1 - p_x) \cdot \Phi(A, o)_a = p_x \cdot \Phi(A, o)_b.$$

Thus

$$\frac{\Phi(A, o)_a}{\Phi(A, o)_b} = \frac{p_x}{1 - p_x} = \exp\big(u(x)\big), \tag{23}$$

where the denominators are non-zero by positivity of Φ.

As $o(a) = x * o(b)$, we get $u(o(a)) = u(x) + u(o(b))$. Expressing $u(x) = u(o(a)) - u(o(b))$ and plugging it into (23), we get

$$\frac{\Phi(A, o)_a}{\Phi(A, o)_b} = \frac{\exp\Big(u\big(o(a)\big)\Big)}{\exp\Big(u\big(o(b)\big)\Big)}.$$

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

$$\Phi(A, o)_a = \frac{\exp\Big(u\big(o(a)\big)\Big)}{\sum_{b \in A} \exp\Big(u\big(o(b)\big)\Big)}, \tag{24}$$

i.e., Φ is multinomial logit.

It remains to check the equivalence between the continuity of Φ and that of u. If $u \colon \mathcal{O} \to \mathbb{R}$ is a continuous utility representation, then Φ given by (24) is continuous since the right-hand side of (24) is a continuous function of the profile of outcomes. In the opposite direction, we suppose that Φ is continuous and show that $u(x_n) \to u(x)$ for any sequence $x_n \to x$ in $\mathcal{O}$. Fix a binary set of actions $A = \{a, b\}$ and consider outcome functions o_n and o given by $o_n(a) = o(a) = x$, $o_n(b) = x_n$, and $o(b) = x$. Thus (A, o_n) converges to the menu with identical outcomes (A, o). By continuity, $\Phi(A, o_n)$ converges to the uniform distribution $\Phi(A, o)$. By (24), we get

$$\frac{\exp\Big(u\big(x_n\big)\Big)}{\exp\Big(u\big(x\big)\Big)} = \frac{\Phi(A, o_n)_b}{\Phi(A, o_n)_a} \to \frac{\Phi(A, o)_b}{\Phi(A, o)_a} = 1$$

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

$\square$

C Proof of Proposition 1

We prove the proposition and then discuss various extensions.

Proof of Proposition 1. Let Φ be a decomposable rule satisfying approximate neutrality with parameter $\varepsilon_{\text{neut}} > 0$. Our goal is to demonstrate that it is neutral. In other words, we need to show that, for any menus (A, o) and actions $a, a' \in A$ with the same outcome $o(a) = o(a')$, the probabilities assigned by Φ to a and a' are the same.

If one of a or a' has zero probability, then the other also has zero probability by approximate neutrality. Hence, we can focus on the case where $\Phi(A, o)_a > 0$ and $\Phi(A', o')_{a'} > 0$.

Consider a menu $(A, o)^{\otimes n}$ equal to the n-fold product of (A, o) with itself. The two actions actions $(a, \ldots, a)$ and $(a', \ldots, a')$ have the same outcomes. Thus, by approximate neutrality,

$$\left| \ln \frac{\Phi\Big((A, o)^{\otimes n}\Big)_{(a,\ldots,a)}}{\Phi\Big((A, o)^{\otimes n}\Big)_{(a',\ldots,a')}} \right| \leqslant \ln(1 + \varepsilon_{\text{neut}}) \tag{25}$$

On the other hand, by decomposability,

$$\Phi\Big((A, o)^{\otimes n}\Big)_{(a,\ldots,a)} = \Big(\Phi(A, o)_a\Big)^n, \quad \text{and} \quad \Phi\Big((A, o)^{\otimes n}\Big)_{(a',\ldots,a')} = \Big(\Phi(A, o)_{a'}\Big)^n. \tag{26}$$

Plugging these identities in (25), we get

$$\left| \ln \frac{\Phi(A, o)_a}{\Phi(A, o)_{a'}} \right| \leqslant \frac{1}{n} \cdot \ln(1 + \varepsilon_{\text{neut}}). \tag{27}$$

Since n is arbitrary, $\Phi(A, o)_a = \Phi(A, o)_{a'}$ and thus Φ is neutral. $\square$

The proof suggests two straightforward generalizations of Proposition 1. First, we can allow $\varepsilon_{\text{neut}}$ to depend on the menu, i.e., $\varepsilon_{\text{neut}} = \varepsilon_{\text{neut}}[(B, r)]$. Indeed, by (27), the conclusion of the proposition holds as long as

$$\sqrt[n]{1 + \varepsilon_{\text{neut}}[(A, o)^{\otimes n}]} \to 1, \quad \text{as} \quad n \to \infty.$$

In particular, if $\varepsilon_{\text{neut}}[(B, r)]$ grows sub-linearly with the size of the menu $|B|$, then $\varepsilon_{\text{neut}}$-neutrality implies neutrality.

Proposition 1 also extends to rules Φ that are both approximately neutral and approximately decomposable with parameters $\varepsilon_{\text{neut}} \geqslant 0$ and $\varepsilon_{\text{decomp}} \geqslant 0$. For such Φ, identities (26) hold up to a multiplicative factor $(1+\varepsilon_{\text{decomp}})^n$ and we get

$$\left|\ln \frac{\Phi(A,o)_a}{\Phi(A,o)_{a'}}\right| \leqslant \frac{1}{n}\Big(\ln(1+\varepsilon_{\text{neut}}) + 2\cdot \ln\big((1+\varepsilon_{\text{decomp}})^n\big)\Big)$$

instead of (27). Letting n go to infinity, we conclude that

$$\Phi(A,o)_a \leqslant (1+\varepsilon_{\text{decomp}})^2 \cdot \Phi(A,o)_{a'}.$$

Thus any $\varepsilon_{\text{neut}}$-neutral $\varepsilon_{\text{decomp}}$-decomposable Φ is also $\varepsilon'_{\text{neut}}$-neutral with

$$\varepsilon'_{\text{neut}} = \min\Big\{\varepsilon_{\text{neut}},\ 2\varepsilon_{\text{decomp}} + \varepsilon_{\text{decomp}}^2\Big\}. \tag{28}$$

D Proof of Theorem 3

Recall that the outcome space is $\mathcal{O} = \mathbb{R}$ and Φ is a continuous rule satisfying approximate neutrality and approximate decomposability with parameters $\varepsilon_{\text{decomp}} \geqslant 0$ and $\varepsilon_{\text{neut}} \geqslant 0$. Our goal is to demonstrate that Φ is $\varepsilon_{\text{decomp}}$-close to multinational logit. The proof is split into several lemmas. The first one demonstrates that any rule Φ from the statement of the theorem satisfies a stronger neutrality notion that we are about to define.

We call two menus (A,o) and (A',o') equivalent if there exists a bijection $\sigma\colon A \to A'$ such that $o(a) = o'(a')$ for all $a \in A$ and $a' = \sigma(a)$. Equivalence means that the menus are the same up to renaming the actions. A rule Ψ is **strongly neutral** if for equivalent (A,o) and (A',o') we have

$$\Psi(A,o)_a = \Psi(A',o')_{a'} \quad \text{when} \quad o(a) = o'(a').$$

Strong neutrality means that the profile of outcomes is a sufficient statistic for choice probabilities. Similarly, Ψ is **approximately strongly neutral** with a parameter $\varepsilon_{s-\text{neut}} \geqslant 0$ if

$$\Psi(A,o)_a \leqslant (1+\varepsilon_{s-\text{neut}})\Psi(A',o')_{a'} \quad \text{when} \quad o(a) = o'(a').$$

It is easy to see that approximate strong neutrality implies approximate neutrality with the same parameter. By Theorem 2, exact neutrality and decomposability imply strong neutrality. The following lemma shows that this implication extends to the approximate axioms.

Lemma 4. *Let Ψ be a rule satisfying approximate neutrality and decomposability with parameters $\varepsilon_{\text{neut}} \geqslant 0$ and $\varepsilon_{\text{decomp}} \geqslant 0$, respectively. Then Ψ is approximately strongly neutral with parameter $\varepsilon_{s-\text{neut}}$ such that $1 + \varepsilon_{s-\text{neut}} = (1 + \varepsilon_{\text{neut}})(1 + \varepsilon_{\text{decomp}})^2$.*

The lemma implies that the rule Φ from the statement of the theorem is approximately strongly neutral.

Proof of Lemma 4. Consider equivalent menus (A, o) and (A', o') and let $\sigma\colon A \to A'$ be the bijection such that $o(a) = o'(a')$ for all $a \in A$ and $a' = \sigma(a)$. Fix a and $a' = \sigma(a)$. In a product menu $(B, r) = (A, o) \times (A', o')$, actions (a, b') and $(\sigma^{-1}(b'), a')$ have the same outcomes. Hence, approximate neutrality implies

$$\Psi(B, r)_{(a,b')} \leqslant (1 + \varepsilon_{\text{neut}})\Psi(B, r)_{(\sigma^{-1}(b'),\, a')}.$$

Expanding both sides via approximate decomposability, we get

$$\frac{1}{1 + \varepsilon_{\text{decomp}}}\Psi(A, o)_a \cdot \Psi(A', o')_{b'} \leqslant (1 + \varepsilon_{\text{neut}})(1 + \varepsilon_{\text{decomp}})\Psi(A, o)_{\sigma^{-1}(b')} \cdot \Psi(A', o')_{a'}.$$

Summing both sides over $b' \in A'$ and using the fact that probabilities sum up to one, we obtain

$$\Psi(A, o)_a \leqslant (1 + \varepsilon_{\text{neut}})(1 + \varepsilon_{\text{decomp}})^2 \cdot \Psi(A', o')_{a'}$$

and conclude that Ψ is approximately strongly neutral with parameter $\varepsilon_{s-\text{neut}}$. $\square$

For a menu (A, o), we denote (as above) by $(A, o)^{\otimes n}$ the product of (A, o) with itself n times. Given a rule Φ from the statement of the theorem, we define an auxiliary rule $\Upsilon(A, o)$ as follows:

$$\Upsilon(A, o)_a = \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A, o)^{\otimes n}\Big)_{(a,\ldots,a)}}. \tag{29}$$

We will demonstrate that this limit exists and that Υ is close to Φ. Moreover, we will see that Υ satisfies exact decomposability and neutrality.

Lemma 5. *The limit in (29) exists and*

$$\begin{gathered}\Phi\Big((A, o)^{\otimes n}\Big)_{(a,\ldots,a)} \leqslant (1 + \varepsilon_{\text{decomp}})\Big(\Upsilon(A, o)_a\Big)^n \\ \Big(\Upsilon(A, o)_a\Big)^n \leqslant (1 + \varepsilon_{\text{decomp}})\Phi\Big((A, o)^{\otimes n}\Big)_{(a,\ldots,a)}\end{gathered} \tag{30}$$

for any $n \geqslant 1$.

Proof. We first consider the case $\Phi(A,o)_a = 0$. By approximate decomposability, $\Phi\big((A,o)^{\otimes n}\big)_{(a,\ldots,a)}$ is also zero. Thus the limit (29) exists, equals zero, and inequalities (30) are satisfied trivially.

We now assume $\Phi(A,o)_a > 0$. Denote

$$f_n = \ln\left(\Phi\Big((A,o)^{\otimes n}\Big)_{(a,\ldots,a)}\right).$$

By the approximate decomposability,

$$|f_{n+m} - f_n - f_m| \leqslant \ln(1+\varepsilon_{\text{decomp}}) \tag{31}$$

for any $n, m \geqslant 1$. We need the following standard result about subadditive sequences.

Lemma (Fekete's subadditive lemma)**.** *Consider a sequence of real numbers g_n, $n \geqslant 1$, with the subadditivity property: $g_{n+m} \leqslant g_n + g_m$. Then, there exists a limit $\gamma = \lim_{n\to\infty} \frac{g_n}{n} \in \mathbb{R} \cup \{-\infty\}$ and g_n satisfies the lower bound $g_n \geqslant \gamma \cdot n$ for any $n \geqslant 1$.*

Inequality (31) implies that the two sequences $g_n = f_n + \ln(1+\varepsilon_{\text{decomp}})$ and $g_n' = -f_n + \ln(1+\varepsilon_{\text{decomp}})$ are both subadditive. By Fekete's lemma, there exists a limit

$$\gamma = \lim_{n\to\infty} \frac{f_n}{n} \in \mathbb{R}, \qquad \text{and} \qquad |f_n - \gamma \cdot n| \leqslant \ln(1+\varepsilon_{\text{decomp}}).$$

Expressing the limit in the definition (29) of Υ through f_n, we get

$$\Upsilon(A,o)_a = \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A,o)^{\otimes n}\Big)_{(a,\ldots,a)}} = \exp\left(\lim_{n\to\infty} \frac{f_n}{n}\right) = \exp(\gamma).$$

Thus the limit in (29) exists. Moreover, the inequalities (30) hold as they are equivalent to $|f_n - \gamma \cdot n| \leqslant \ln(1+\varepsilon_{\text{decomp}})$. $\square$

Lemma 6. *Υ is neutral.*

To prove this and other statements below, we will use the notation $\Omega(1)$ to denote a quantity bounded away from zero and infinity. Formally, a sequence $h_n \geqslant 0$, $n \geqslant 1$, satisfies $h_n = \Omega(1)$ if there exist constants $\alpha > 0$ and N_0 such that $\alpha \leqslant h_n \leqslant 1/\alpha$ for any $n \geqslant N_0$.

Proof. Consider a menu (A,o) and a pair of actions $a, a' \in A$ with $o(a) = o(a')$. Our goal is to show that $\Upsilon(A,o)_a = \Upsilon(A,o)_{a'}$. By the approximate neutrality of Φ,

$$\Phi\Big((A,o)^{\otimes n}\Big)_{(a,\ldots,a)} = \Omega(1) \cdot \Phi\Big((A,o)^{\otimes n}\Big)_{(a',\ldots,a)'}.$$

By the definition of Υ, we get

$$\begin{aligned}
\Upsilon(A,o)_a &= \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A,o)^{\otimes n}\Big)_{(a,\ldots,a)}} \\
&= \lim_{n\to\infty} \sqrt[n]{\Omega(1)\cdot\Phi\Big((A,o)^{\otimes n}\Big)_{(a',\ldots,a')}} \\
&= \lim_{n\to\infty} \sqrt[n]{\Omega(1)}\cdot \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A,o)^{\otimes n}\Big)_{(a',\ldots,a')}} \\
&= \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A,o)^{\otimes n}\Big)_{(a',\ldots,a')}} \\
&= \Upsilon(A,o)_{a'}.
\end{aligned}$$

Thus $\Upsilon(A,o)_a = \Upsilon(A,o)_{a'}$ and so Υ is neutral. □

Lemma 7. Υ *is decomposable.*

Proof. Consider a pair of menus (A_1,o_1) and (A_2,o_2) and let $(A,o) = (A_1,o_1)\otimes(A_2,o_2)$. Our goal is to show that $\Upsilon(A,o)_{(a,b)} = \Upsilon(A_1,o_1)_a \cdot \Upsilon(A_2,o_2)_b$. By the definition of Υ,

$$\Upsilon(A,o)_{(a,b)} = \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A,o)^{\otimes n}\Big)_{((a,b),\ldots,(a,b))}}.$$

Menus $(A,o)^{\otimes n}$ and $(A_1,o_1)^{\otimes n}\otimes(A_2,o_2)^{\otimes n}$ are equivalent and the actions $((a,b),\ldots,(a,b))$ and $\big((a,\ldots,a),\,(b,\ldots,b)\big)$ have the same outcomes. By Lemma 4 and approximate decomposability, we obtain

$$\Phi\Big((A,o)^{\otimes n}\Big)_{((a,b),\ldots,(a,b))} = \Omega(1)\cdot\Phi\Big((A_1,o_1)^{\otimes n}\Big)_{(a\ldots,a)}\cdot\Phi\Big((A_2,o_2)^{\otimes n}\Big)_{(b\ldots,b)}.$$

Thus

$$\begin{aligned}
\Upsilon(A,o)_{(a,b)} &= \lim_{n\to\infty} \sqrt[n]{\Omega(1)\cdot\Phi\Big((A_1,o_1)^{\otimes n}\Big)_{(a\ldots,a)}\cdot\Phi\Big((A_2,o_2)^{\otimes n}\Big)_{(b\ldots,b)}} \\
&= \lim_{n\to\infty} \sqrt[n]{\Omega(1)}\cdot \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A_1,o_1)^{\otimes n}\Big)_{(a\ldots,a)}}\cdot \lim_{n\to\infty} \sqrt[n]{\Phi\Big((A_2,o_2)^{\otimes n}\Big)_{(b\ldots,b)}} \\
&= \Upsilon(A_1,o_1)_a\cdot\Upsilon(A_2,o_2)_b.
\end{aligned}$$

We conclude that Υ is decomposable. □

From the definition of Υ, it is not apparent that the probabilities of all the actions sum up to one. The following lemma verifies this.

Lemma 8. *For any menu (A,o), we have*

$$\sum_{a\in A}\Upsilon(A,o)_a = 1.$$

Proof. Consider a menu (A,o) and its n-fold product $(A,o)^{\otimes n}$.

Since $\Phi(A,o)^{\otimes n}$ is a probability measure, we get

$$1 = \sum_{t\in A\times\ldots\times A}\Phi\Big((A,o)^{\otimes n}\Big)_t \tag{32}$$

Assuming that $n\to\infty$, we approximately express each of the terms $\Phi(A,u)_t^{\otimes n}$ in this sum through Υ. For each $a\in A$, denote by $n_a(t)$ the number of times a enters $t\in A\times\ldots\times A$. By approximate decomposability and approximate strong neutrality (Lemma 4),

$$\Phi\Big((A,o)^{\otimes n}\Big)_t = \Omega(1)\cdot\prod_{a\in A}\Phi\Big((A,o)^{\otimes n_a(t)}\Big)_{(a,\ldots,a)}$$

Using inequalities (30), we obtain

$$\Phi\Big((A,o)^{\otimes n}\Big)_t = \Omega(1)\cdot\prod_{a\in A}\big(\Upsilon(A,o)_a\big)^{n_a(t)}.$$

Plugging this expression into (32) gives

$$1 = \sum_{t\in A\times\ldots\times A}\Omega(1)\cdot\prod_{a\in A}\big(\Upsilon(A,o)_a\big)^{n_a(t)} = \Omega(1)\cdot\left(\sum_{a\in A}\Upsilon(A,o)_a\right)^n.$$

Thus

$$\sum_{a\in A}\Upsilon(A,o)_a = \lim_{n\to\infty}\sqrt[n]{\Omega(1)} = 1.$$

We conclude that $\Upsilon(A,o)$ is indeed a probability distribution over A. $\square$

Proof of Theorem 3. Consider a continuous rule Φ satisfying approximate neutrality and decomposability with parameters $\varepsilon_{\text{neut}}$ and $\varepsilon_{\text{decomp}}$.

Let Υ be defined by formula (29). As we established in Lemmas 5, 6, 7, and 8, Υ is a decomposable neutral rule that is close to Φ in the following sense

$$\begin{aligned}\Phi(A,o)_a &\leqslant (1+\varepsilon_{\text{decomp}})\Upsilon(A,o)_a \\ \Upsilon(A,o)_a &\leqslant (1+\varepsilon_{\text{decomp}})\Phi(A,o)_a\end{aligned} \tag{33}$$

Consider a restriction of Υ to the collection $\mathcal{M}_{\mathbb{Q}}$ of menus (A, o) with rational-valued o. By Proposition 3, there is $\beta \in \mathbb{R} \cup \{\pm\infty\}$ such that Υ coincides with the generalized logit MNL^{β} for any $(A, o) \in \mathcal{M}_{\mathbb{Q}}$. We now show that the case of infinite β is ruled out by continuity of Φ. Towards contradiction, suppose that $\beta = +\infty$ and so $\Upsilon = \mathrm{MNL}^{\beta}(A, o)$ is the uniform distribution over the highest-outcome actions $A^* = \arg\max o(a)$. By (33), $\Phi(A, o)$ places non-zero weight on actions $a \in A^*$ only and $\Phi(A, o)_a \geqslant 1/(|A^*| \cdot (1 + \varepsilon_{\mathrm{decomp}}))$. Consider a sequence of menus (A, o_n) with $A = \{a_0, a_1\}$ and $o_n(a_0) = 0$, $o_n(a_1) = 1/n$ as in the proof of Theorem 1. We obtain that $\lim_n \Phi(A, o_n)_{a_0} = 0$ but $\Phi(A, o)_{a_0} \geqslant 1/(2 \cdot (1 + \varepsilon_{\mathrm{decomp}}))$ for the limiting menu $(A, o) = \lim_n (A, o_n)$. This contradiction with continuity of Φ implies that β cannot be equal $+\infty$. The case of $\beta = -\infty$ is also ruled out as $\mathrm{MNL}^{-\infty}(A, o) = \mathrm{MNL}^{+\infty}(A, -o)$.

We conclude that there exists $\beta \in \mathbb{R}$ such that, for any menu with rational outcomes $(A, o) \in \mathcal{M}_{\mathbb{Q}}$,

$$\begin{aligned} \Phi(A, o)_a &\leqslant (1 + \varepsilon_{\mathrm{decomp}})\mathrm{MNL}^{\beta}(A, o)_a \\ \mathrm{MNL}^{\beta}(A, o)_a &\leqslant (1 + \varepsilon_{\mathrm{decomp}})\Phi(A, o)_a \end{aligned} \tag{34}$$

Since $\mathcal{M}_{\mathbb{Q}}$ is dense in the set of all menus and the rules Φ and MNL^{β} are continuous, the inequalities hold for all menus (A, o).

Since $\mathrm{MNL}(A, o)_a > 0$ for any action and menu, the inequalities (34) imply that $\Phi(A, o)_a > 0$. Define

$$s(a) = \ln\left(\frac{\Phi(A, o)_a}{\mathrm{MNL}^{\beta}(A, o)_a}\right).$$

By (34), we get $|s(a)| \leqslant \ln(1 + \varepsilon_{\mathrm{decomp}})$. Since $\ln(1 + x) \leqslant x$,

$$|s(a)| \leqslant \varepsilon_{\mathrm{decomp}}.$$

We conclude that $\Phi(A, o)_a$ is proportional to $\exp(\beta \cdot o(a) + s(a))$, where $s \colon A \to [-\varepsilon_{\mathrm{decomp}}, \varepsilon_{\mathrm{decomp}}]$. Thus Φ is $\varepsilon_{\mathrm{decomp}}$-close to MNL^{β}.

Finally, we check that $\beta \in \mathbb{R}$ such that Φ is $\varepsilon_{\mathrm{decomp}}$-close to MNL^{β} is unique. Indeed, suppose that Φ is $\varepsilon_{\mathrm{decomp}}$-close both to MNL^{β} and $\mathrm{MNL}^{\beta'}$. Therefore, for any menu (A, o) and $a \in A$,

$$\frac{1}{(1 + \varepsilon_{\mathrm{decomp}})^2} \leqslant \frac{\mathrm{MNL}^{\beta}(A, o)_a}{\mathrm{MNL}^{\beta'}(A, o)_a} \leqslant (1 + \varepsilon_{\mathrm{decomp}})^2. \tag{35}$$

Consider a binary menu (B, r) with $B = \{b_0, b_1\}$ and outcomes $r(b_i) = i$ and take (A, o) in (35) equal to the n-fold product $(A, o) = (B, r)^{\otimes n}$. Picking $a = (b_1, \ldots, b_1)$

and letting n go infinity, we get that $\beta = \beta'$. Thus β is unique and the proof is completed. $\square$

E Proof of Theorem 4

Theorem 4 claims that a positive $\varepsilon_{\text{neut}}$-neutral and $\varepsilon_{\text{decomp}}$-decomposable rule Φ for a general outcome space $(\mathcal{O}, *)$ that is Ulam stable must be close to multinomial logit. We prove here an extension of this theorem providing an explicit bound on the distance in terms of $(\varepsilon_{\text{neut}}, \varepsilon_{\text{decomp}})$. Recall that a generalized Cauchy equation

$$u(x * y) = u(x) + u(y) \quad \text{for all} \quad x, y \in \mathcal{O}. \tag{36}$$

is (ε, δ)-stable if for any function w solving the ε-approximate equation

$$\big|w(x * y) - w(x) + w(y)\big| \leqslant \varepsilon \quad \text{for all} \quad x, y \in \mathcal{O} \tag{37}$$

there is a solution u of (36) such that

$$|u(x) - w(x)| \leqslant \delta \quad \text{for all} \quad x \in \mathcal{O}.$$

Theorem 5. *Consider an outcome space $\mathcal{O}$ such that the Cauchy equation (36) is (ε, δ)-stable with $\delta = d(\varepsilon)$ and any $\varepsilon > 0$. Let Φ be a positive $\varepsilon_{\text{neut}}$-neutral $\varepsilon_{\text{decomp}}$-decomposable stochastic choice rule with some $\varepsilon_{\text{neut}} \geqslant 0$ and $\varepsilon_{\text{decomp}} \geqslant 0$. Then Φ is $\Big(2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}} + d(4\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}})\Big)$-close to multinomial logit.*

As established in (28), any $\varepsilon_{\text{neut}}$-neutral $\varepsilon_{\text{decomp}}$-decomposable rule is also $(2\varepsilon_{\text{decomp}} + \varepsilon_{\text{decomp}}^2)$-neutral. Combining this insight with Theorem 5, we obtain that Φ from the statement is $\Big(4\varepsilon_{\text{decomp}} + \varepsilon_{\text{decomp}}^2 + d(6\varepsilon_{\text{decomp}} + \varepsilon_{\text{decomp}}^2)\Big)$-close to multinomial logit, where the bound no longer depends on $\varepsilon_{\text{neut}}$. This completes the proof of Theorem 4.

Proof of Theorem 5. The proof resembles that of Theorem 2 with the exception that exact equalities are replaced with approximate ones.

As in the proof of Theorem 2, for each element $x \in \mathcal{O}$ of the outcome space, fix a menu

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

and let $p_x = \Phi(A_x, o_x)_{a_x}$.

We define
$$w(x) = \ln \frac{p_x}{1-p_x}.$$
By positivity, w is finite.

We now demonstrate that w solves the approximate Cauchy equation (37) with $\varepsilon = 4\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}}$, i.e.,
$$\left|w(x*y) - w(x) - w(y)\right| \leqslant \varepsilon$$
for all $x, y \in \mathcal{O}$. Consider a product menu
$$(B,o) = \Big((A_x, o_x) \otimes (A_y, o_y)\Big) \otimes (A_{x*y}, o_{x*y}).$$
The constructed product menu contains actions
$$b = \big((a_e, a_e), a_{x*y}\big) \qquad \text{and} \qquad b' = \big((a_x, a_y), a_e\big).$$
Computing their outcomes, we get
$$o(b) = (e*e)*(x*y) = x*y \qquad \text{and} \qquad o(b') = (x*y)*e = x*y,$$
where we used the fact that e is both a left and a right identity. Since the outcomes of b and b' are the same, approximate neutrality of Φ implies
$$\left|\ln \frac{\Phi(B,o)_b}{\Phi(B,o)_{b'}}\right| \leqslant \ln(1+\varepsilon_{\text{neut}}).$$
By approximate decomposability applied twice,
$$\left|\ln \frac{\Phi(B,o)_b}{(1-p_x)\cdot(1-p_y)\cdot p_{x*y}}\right| \leqslant \ln((1+\varepsilon_{\text{decomp}})^2)$$
and
$$\left|\ln \frac{\Phi(B,o)_{b'}}{p_x \cdot p_y \cdot (1-p_{x*y})}\right| \leqslant \ln((1+\varepsilon_{\text{decomp}})^2).$$
Combining these inequalities, we get
$$\left|\ln \frac{(1-p_x)\cdot(1-p_y)\cdot p_{x*y}}{p_x \cdot p_y \cdot (1-p_{x*y})}\right| \leqslant 2\ln((1+\varepsilon_{\text{decomp}})^2) + \ln(1+\varepsilon_{\text{neut}}).$$
Since $\ln(1+t) \leqslant t$ for any $t > -1$, the right-hand side does not exceed $4\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}}$. Expressing the left-hand side through the function w, we get
$$\left|w(x*y) - w(x) - w(y)\right| \leqslant 4\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}}.$$

Since the Cauchy equation is assumed to be $(\varepsilon, d(\varepsilon))$-stable, we conclude that there is a utility representation $u\colon \mathcal{O} \to \mathbb{R}$ solving the exact Cauchy equation

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

and such that

$$|u(x) - w(x)| \leqslant d(\varepsilon) \qquad \text{with} \quad \varepsilon = 4\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}}.$$

We now consider an arbitrary menu (A, o). Let $a, b \in A$ be two distinct actions. By the assumption 2 on the operation $*$, there is $x \in \mathcal{O}$ such that $o(a) = x * o(a')$, or there is $y \in \mathcal{O}$ such that $o(a') = y * o(a)$, or both. Without loss of generality, we assume that $o(a) = x * o(a')$. Consider the product menu

$$(C, s) = (A_x, o_x) \otimes (A, o).$$

This menu contains two actions

$$c = (a_e, a) \qquad \text{and} \qquad c' = (a_x, b)$$

with equal outcomes

$$s(c) = e * o(a) = o(a) \qquad \text{and} \qquad s(c') = x * o(b) = o(b).$$

By approximate neutrality,

$$\left| \ln \frac{\Phi(C, s)_c}{\Phi(C, s)_{c'}} \right| \leqslant \ln(1 + \varepsilon_{\text{neut}}) \leqslant \varepsilon_{\text{neut}}.$$

Approximate decomposability implies

$$\left| \ln \frac{\Phi(C, s)_c}{(1 - p_x) \cdot \Phi(A, o)_a} \right| \leqslant \ln(1 + \varepsilon_{\text{decomp}}) \leqslant \varepsilon_{\text{decomp}}$$

and

$$\left| \ln \frac{\Phi(C, s')_{c'}}{p_x \cdot \Phi(A, o)_b} \right| \leqslant \ln(1 + \varepsilon_{\text{decomp}}) \leqslant \varepsilon_{\text{decomp}}.$$

Thus

$$\left| \ln \frac{(1 - p_x) \cdot \Phi(A, o)_a}{p_x \cdot \Phi(A, o)_b} \right| \leqslant 2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}}.$$

We conclude that

$$\left| \ln \frac{\Phi(A, o)_a}{\Phi(A, o)_b} - w(x) \right| \leqslant 2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}}.$$

Expressing the approximate solution w through the exact solution u, we obtain

$$\left|\ln \frac{\Phi(A,o)_a}{\Phi(A,o)_b} - u(x)\right| \leqslant 2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}} + d(\varepsilon)$$

Since, $o(a) = x * o(b)$ and u solves the Cauchy equation, we get $u(o(a)) = u(x) + u(o(b))$. Expressing $u(x) = u(o(a)) - u(o(b))$ and plugging it back we obtain

$$\left|\ln \frac{\Phi(A,o)_a}{\Phi(A,o)_b} - u(o(a)) + u(o(b))\right| \leqslant 2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}} + d(\varepsilon).$$

Fix some $b' \in A$ and define

$$s(a) = \ln \Phi(A,o)_a - u(a) + C,$$

where the constant C is selected so that $s(b') = 0$. We conclude that

$$|s(a)| \leqslant 2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}} + d(\varepsilon) \qquad \text{for any} \quad a \in A.$$

On the other hand,

$$\frac{\Phi(A,o)_a}{\Phi(A,o)_b} = \frac{\exp\Big(u\big(o(a) + s(a)\big)\Big)}{\exp\Big(u\big(o(b) + s(b)\big)\Big)}$$

and so

$$\Phi(A,o)_a = \frac{\exp\left(u\big(o(a)\big) + s(a)\right)}{\sum_{b\in A} \exp\left(u\big(o(b)\big) + s(b)\right)}.$$

Thus Φ is $\Big(2\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}} + d(4\varepsilon_{\text{decomp}} + \varepsilon_{\text{neut}})\Big)$-close to multinomial logit. $\square$