

Essays on Stochastic Choice

Thesis by
Po Hyun Sung

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended May 6, 2026

© 2026

Po Hyun Sung

ORCID: 0009-0000-5922-6686

All rights reserved

ACKNOWLEDGEMENTS

I am deeply indebted to my advisors, Omer Tamuz and Charles Sprenger, for their guidance, insight and generosity. I am also very grateful to Luciano Pomatto and Kirby Nielsen for their continual mentorship and support.

I have also learned so much from my coauthors Ted O'Donoghue and Fedor Sandomirskiy. I am grateful as well to Marina Agranov, Peter Caradonna, Axel Niemeyer, and Kota Saito for many fruitful conversations.

I feel very lucky to have gone through my PhD with the greatest cohort: Ben, Camila, Egor, Jake, Paulo, and Sean. I also want to thank Yeonsoo for being a wonderful friend throughout.

I am deeply grateful to my parents, my family and my friends for their love and support. Finally, I want to thank Ben for being the best partner in both research and life.

ABSTRACT

This dissertation contains three essays on stochastic choice theory, exploring the theoretical properties of stochastic choice models widely used in empirical economics and examining how these properties inform their use in applications.

In Chapter 1, co-authored with Ben Wincelberg, we characterize when risk aversion reversals—situations in which a more risk-averse individual chooses a riskier asset more frequently—arise in weak utility models of binary stochastic choice, a non-parametric class that encompasses virtually all models of stochastic choice used in practice. Reversals necessarily arise if and only if the ratio of the second derivatives of the individuals’ utilities is unbounded. Reversals are robust to arbitrary heterogeneity and choice of utility representation. Our condition is violated by the most widely used utility specifications including CARA, CRRA, and utilities without constant risk aversion such as expo-power. Under these specifications, reversals are pervasive, occurring for empirically relevant risk coefficients and lottery choices. We thus propose two parametric utility families that do not lead to reversals, offering well-behaved alternatives for stochastic choice. Finally, we show that these proposed functional forms remain well-behaved in multinomial settings.

In Chapter 2, co-authored with Ted O’Donoghue, Charles Sprenger, and Ben Wincelberg, we investigate the apparent inconsistencies between the two principal tools used to measure preferences: binary-choice and valuation tasks. While divergent measurements from these tools have led researchers to question the stability of underlying preferences, we theoretically demonstrate that stable preferences do not generally imply identical measurements, even under strong assumptions about noise and preference heterogeneity. We derive the attainable sets of observations under the null of stable preferences combined with various ancillary assumptions about noise and preference heterogeneity. Using data from the prior literature (375 experiments) and new data from McGranaghan et al. (2024a) and McGranaghan et al. (2024b) (424 experiments), we test whether we can reject the null that experimental results come from each attainable set consistent with stable preferences. We find that, even though these data overwhelmingly show the patterns typically interpreted as evidence of instability, the vast majority are in fact consistent with stable preferences once one recognizes the possibilities of noise and preference heterogeneity.

In Chapter 3, co-authored with Fedor Sandomirskiy, Omer Tamuz, and Ben Wincelberg, we study stochastic choice across decision problems, each represented as a menu of action labels paired with observable outcome vectors. We propose a consistency condition for behavior in decision problems composed of two separable components: choice probabilities must agree with those obtained when each component is considered in isolation. Together with monotonicity and continuity, this separability requirement characterizes the family of random coefficients logit rules.

CONTENTS

Acknowledgements	iii
Abstract	iv
Contents	vi
List of Figures	viii
List of Tables	ix
Chapter I: Risk Aversion Reversals	1
1.1 Introduction	1
1.2 Preliminaries	7
1.3 Fechnerian Models	9
1.4 Weak Utility Models	15
1.5 Monotone Alternatives to CARA and CRRA	20
1.6 Multinomial Choice	22
1.7 Conclusion	26
1.A Proof of Proposition 1	27
1.B Proof of Proposition 3	29
1.C Proof of Theorem 1	30
1.D Comparison with Standard Comparative Risk Notions	33
1.E Expo-power Utilities	34
1.F Sum-ex and Sum-power Utility Functions	35
1.G Multinomial Choice Extensions	37
Bibliography	41
Chapter II: Measuring Economic Preferences in the Presence of Noise: The Connections Between Choices and Valuations	44
2.1 Introduction	44
2.2 Failures of Procedural Invariance and Tests of Preference Stability	48
2.3 Theoretical Considerations	50
2.4 Empirical Application	65
2.5 Conclusion	80
2.A Additional Figures	83
2.B Proofs	84
2.C The McGranaghan et al. (2024a) and McGranaghan et al. (2024b) Datasets	89
2.D Connecting Traditional and Structured Analyses	93
2.E Variability of Measurement and Bounding the Correlation in Prefer- ences	95
Bibliography	97
Chapter III: A Separability Foundation for Random Coefficients Logit	99
3.1 Introduction	99
3.2 Model	103

3.3	IID and Random Coefficients Logit	112
3.4	Correlations, Decomposability, and Linear Multinomial Logit	114
3.5	Probit Rules	117
3.6	Utility Representations for Abstract Outcome Spaces $\mathcal{O}$	119
3.7	Proof of Theorem 1 for $\mathcal{O} = \mathbb{R}$	122
3.8	Conclusion	129
3.A	Monotonicity of Random Coefficients Logit	130
3.B	Proof of Proposition 1	130
3.C	Proof of Theorem 1 for $\mathcal{O} = \mathbb{R}^n$ with $n \geq 2$	132
3.D	Proof of Proposition 2	136
3.E	Proof of Theorem 2	136
3.F	Restrictions of Monotonicity and IID	138
	Bibliography	141

LIST OF FIGURES

<i>Number</i>	<i>Page</i>
2.1 Individual Choice and Valuation Probabilities with Varying Noise Assumptions.	54
2.2 Attainable Sets from Propositions 1 and 2.	57
2.3 Failure of Conditional Calibration from Proposition 5.	63
2.4 Traditional Inferential Approach.	68
2.5 Evaluating Propositions 1, 2, and 3(1).	70
2.6 Evaluating Proposition 3(2).	73
2.7 Evaluating Proposition 5.	75
2.8 Estimated Preferences from Probit and Full ML Analysis.	79
2.9 Evaluating Symmetry of $y_{vi} - \overline{y_{vi}}$	83
2.10 Comparative Statics With Respect to Problem Parameters.	93
2.11 Connecting Traditional and Structured Analyses.	94
2.12 Variability of Elicitation and Correlation in Preferences.	96

LIST OF TABLES

<i>Number</i>	<i>Page</i>
1.1 CRRA Choice Probabilities for S with normal errors.	3
2.1 Probit Estimation	77

Chapter 1

RISK AVERSION REVERSALS

1.1 Introduction

A precise understanding of how humans evaluate risk is central to economics. Since choices are inherently noisy, analysts deploy stochastic choice models to recover individuals' risk preferences and noise from choice data. The most widely used stochastic choice models for this purpose are *Fechnerian models*, a non-parametric class that includes the logit and probit models. In practice, the analyst chooses a functional form for utility and a functional form for noise which maps utility differences to choice probabilities (e.g., Harrison, List, and Towe, 2007; Von Gaudecker, Van Soest, and Wengström, 2011; Holzmeister and Stefan, 2021).

However, predictions from these models under CARA or CRRA—the most commonly used utility functions—suffer from a well-known problem. When two individuals differ in their risk-aversion parameters but face the same noise, the more risk-averse individual exhibits more risk-seeking behavior in some choices (Wilcox, 2008; Wilcox, 2011; Blavatskyy, 2011; Apesteguia and Ballester, 2018). In response to these critiques, subsequent work has argued that non-monotonicity is particularly worrisome under homoskedastic noise and has advocated allowing heteroskedasticity to better calibrate noise to utility (Barseghyan et al., 2018; O'Donoghue and Somerville, 2024; Keffert and Schweizer, 2024). Moreover, it is common to model heterogeneity in both preferences and noise and to estimate them jointly (e.g., Hey and Orme, 1994; Von Gaudecker, Van Soest, and Wengström, 2011; Meissner et al., 2023).

In this paper, we show that the problems noted for Fechnerian models are far more pervasive than previously recognized: they persist even when individuals differ arbitrarily in their Fechnerian noise, when noise varies across menus, and when noise takes a more general, non-Fechnerian form. That is, under CARA and CRRA utilities, none of these generalizations can eliminate the counterintuitive predictions. We then characterize which utility functions, beyond CARA and CRRA, give rise to risk-aversion reversals and advocate for two parametric utility families that do not, under various noise specifications, exhibit reversals: a more risk-averse individual always chooses safer options more frequently. Beyond its intuitive appeal, this

comparative static aligns with empirical findings (Bruner, 2017). Finally, we discuss a generalization of Fechnerian models to multinomial choice and show that our proposed utility families do not exhibit reversals even in this more general setting.

To build intuition, we begin with the baseline Fechnerian framework, in which each individual is associated with a Bernoulli utility function u and a strictly increasing function F . We refer to F as a Fechnerian noise structure, where

$$F(U(X) - U(Y))$$

is the probability that X is chosen over Y , and $U(X) = \mathbb{E}[u(X)]$. The most popular Fechnerian noise structures are CDFs of normal and logistic distributions, corresponding to probit and logit models. Indeed, any additive random utility models with i.i.d. shocks fit the Fechnerian framework.¹

Our first result is that, for some pairwise choices between a safe and a risky lottery,² a more risk-averse individual chooses the risky option more often than a less risk-averse individual, even when the two differ arbitrarily in their Fechnerian noise structures (Proposition 1). Thus, the problems identified for the homoskedastic noise model cannot be resolved by jointly estimating risk and noise, even when noise is estimated non-parametrically.

For example, consider an analyst who observes Anne and Bob choosing from three distinct menus—each menu consists of a safe lottery—a sure payoff of \$8, \$10, or \$14—and a 50–50 risky lottery paying plus or minus \$4, \$6, or \$8 from the sure amount. The choice probabilities for the safe lotteries are shown in the second and third columns of Table 1.1. Assuming CRRA expected utilities for their risk preferences and normal distributions for their noise structures, Anne is estimated to be more risk averse and experience less noise than Bob. Indeed, the choice probabilities in Table 1.1 correspond to Anne having a CRRA coefficient of 0.8 and noise variance of 0.5, compared to Bob’s 0.3 and 1, respectively.³

¹Technically, for F to be strictly increasing, the difference in these shocks must have full support.

²We say that lottery X is safer than lottery Y if X dominates Y in the concave order. Equivalently, every expected utility maximizer with a concave utility function must prefer X to Y .

³I.e., the analyst deploys the probit model for Anne and Bob

$$F_A(U_A(X) - U_A(Y)) \quad \text{and} \quad F_B(U_B(X) - U_B(Y)),$$

respectively, where U_A and U_B are CRRA expected utility functions with relative risk aversions a and b , and F_A and F_B are the CDFs of normal distributions with zero means and respective variances σ_A^2 and σ_B^2 .

Base Lottery (S vs R)	Probability of S		Probability of $3S$	
	Anne	Bob	Anne	Bob
8 vs 4, 12	0.60	0.57	0.62	0.65
10 vs 4, 16	0.65	0.62	0.69	0.74
14 vs 6, 22	0.64	0.63	0.68	0.77

Table 1.1: CRRA Choice Probabilities for S with normal errors.

In the last two columns of Table 1.1, we report the choice probabilities of safe options whose outcomes are tripled from the base lotteries under the estimated parameters. Note that when the stakes are increased, Bob chooses safer options more frequently than Anne from all menus. Thus, despite Anne being estimated as more risk-averse and more precise, the model predicts that she will behave more risk-seeking than Bob once the stakes are scaled up.

This reversal pattern does not hinge on these specific CRRA risk coefficients and noise structures, nor on the specific choice of lotteries. Indeed, given lotteries X and Y with X safer than Y , we show how to construct lotteries X' and Y' offering higher potential rewards, such that X' remains safer than Y' yet leads to a reversal (Proposition 3). In the special case that the noise structures satisfy $F_A(t/\sigma_A) = F_B(t/\sigma_B)$ for some $\sigma_A, \sigma_B > 0$ (e.g., when both are probit, with different variances as the example above), a reversal occurs if both lotteries are scaled sufficiently: for all large enough scales k , Bob chooses kX over kY more frequently than Anne does (Proposition 2). For CARA utilities, reversals occur after shifting the lotteries by the same sufficiently large outcome under arbitrary noise structures for Anne and Bob (Proposition 3). Thus, CARA and CRRA Fechnerian models with arbitrary noise structures make the perverse and systematic prediction that those most inclined toward choosing safer options when stakes are small must become the least inclined when stakes are larger.

Assuming that this reversal pattern does not always bear out in reality—and indeed Bruner (2017) finds that it does not—our results indicate that the estimates yielded by these models will be sensitive to the level of stakes that are analyzed, making them unreliable.⁴ Indeed, if the analyst instead observed the choice probabilities of the scaled lotteries in Table 1.1, the analyst would estimate that Bob is more risk-

⁴This echoes the critique of expected utility in Rabin (2000), which also reveals paradoxes when stakes are increased. While Rabin's result is about the curvature of utility functions in deterministic models, our result is about the interaction between noise structures and CARA/CRRA utilities.

tolerant than Anne, even if he exhibited more risk-averse behavior. As illustrated in that table, the stake increases are often modest, indicating that such reversals can emerge even in low-stakes experimental settings. Moreover, these models give counterintuitive, and perhaps counterfactual, out-of-sample predictions for choices made at larger or smaller stakes.

We next show that the paradoxical reversals of CARA and CRRA Fechnerian models are not artifacts of overly restrictive Fechnerian noise structures. To establish this, we explore *weak expected utility* (WEU) models. A WEU model consists of a vNM utility function U , a Fechnerian noise structure F and a menu-dependent scale parameter $\sigma: \mathcal{L} \times \mathcal{L} \rightarrow (0, \infty)$ such that the probability of choosing X over Y is given by

$$F\left(\frac{U(X) - U(Y)}{\sigma(X, Y)}\right).^5$$

WEU models generalize the standard Fechnerian framework by allowing choice probabilities to depend on factors beyond strict utility differences. An important subclass of WEU models arises in the additive random utility framework, when the random shocks to utility that are associated with each lottery are independent but not identically distributed. Economically, this accommodates context-dependent noise; for example, large-stakes lotteries may induce increased attention, resulting in less variable shocks.

Without any restrictions on how noise varies across menus, the only implication of WEU models is that individuals will choose their preferred lottery more than half the time. Thus, to derive meaningful comparisons between individuals with different risk attitudes, we must impose additional structure. Our next results (Theorem 1 and Corollary 1) show that, under a mild continuity requirement on the menu-dependent scale parameters, WEU models based on CARA or CRRA utilities are still guaranteed to produce reversals. Thus, even when allowing for arbitrary noise that varies across both individuals and menus, the CARA and CRRA expected utility functional forms lead to untenable predictions.

One might suspect, at this point, that these reversals arise because F and σ are applied to vNM utilities, which are just representations of preferences and carry no cardinal significance. A natural alternative would be to apply F and σ to certainty equivalents $u^{-1}(\mathbb{E}[u(X)])$, or more generally to $f(\mathbb{E}[u(X)])$ for some strictly increasing transformation f . However, Corollary 2 establishes that this

⁵We require $\sigma(X, Y) = \sigma(Y, X)$ so that the probabilities sum up to 1.

transformation does not resolve the problem. Ultimately, CARA and CRRA utilities prove to be fundamentally incompatible with stochastic choice consistency.

We next turn our attention toward the characterization of families of Bernoulli utilities that interact well with noise. Our main result is that when u_A is more risk averse than u_B , there exist reversals for every weak utility model if and only if the ratio

$$u_B''(x)/u_A''(x)$$

is unbounded (Theorem 1). Indeed, for CARA and CRRA utilities, this condition holds, leading to reversals in every model of noise we have considered. Moreover, this condition is met by a class of utility functions that generalizes CARA and CRRA (Corollary 3).

We thus advocate for two families of utility functions for which there are no reversals under homoskedastic noise (see Section 1.5). These families are obtained by taking a linear combination of two distinct CARA (CRRA) utilities and are parameterized by the coefficient on the more risk-averse utility function. For example, for any $0 \leq r < s$, let crra_r and crra_s be CRRA Bernoulli utility functions with respective relative risk coefficients r and s . Summing $\text{crra}_r + a \cdot \text{crra}_s$ yields a family of utility functions parameterized by $a \geq 0$, with greater a corresponding to greater risk aversion. These simple families of utility functions offer well-behaved alternatives for empirical analysis. We discuss the properties of these utility families and the interpretation of the parameter a in Section 1.F.

We next study reversals in the more general multinomial-choice setting. We introduce two competing definitions that extend the notion of reversals from the binary-choice setting to multinomial choice. We focus on the *additive perturbed utility* (APU) model of Fudenberg, Iijima, and Strzalecki (2015), a multinomial extension of the Fechnerian model. Our main result is that the utility families that avoid reversals in the binary setting also avoid both notions of reversals in the multinomial setting (Theorem 2). Finally, we show that the random parameter models considered by Apesteguia and Ballester (2018) lead to one notion of reversal in the multinomial setting, but not the other.

Related Literature

We contribute to a large body of literature on risky choices in the presence of noise (e.g., Becker, DeGroot, and Marschak, 1963; Harless and Camerer, 1994; Hey and Orme, 1994; Loomes, Moffatt, and Sugden, 2002; Blavatskyy, 2007).

See Wilcox (2021) for a recent survey. In particular, we study Fechnerian noise structures and their generalizations. Axiomatic investigations of these models have been undertaken by Debreu (1958), Tversky and Russo (1969), and Fudenberg, Iijima, and Strzalecki (2015), among others. See Strzalecki (2025) for an extensive review.

The papers most closely related to ours are Wilcox (2011) and Apesteguia and Ballester (2018), which demonstrate that CARA and CRRA utilities coupled with identical Fechnerian noise are problematic: a more risk-averse individual exhibits more risk-seeking behavior in some pairwise choices.⁶ Our work builds on their insights in several ways. First, we allow for heterogeneous noise structures and show that the paradox persists without any parametric assumptions on how noise differs across individuals (Proposition 1). Second, we show that the reversals occur under WEU models, in which noise varies across menus (Corollary 1).

WEU models recognize that some comparisons may involve greater cognitive effort or uncertainty. Menu-dependent noise in risky choice has been extensively studied. Hey (1995), Buschena and Zilberman (2000), and Loomes (2005) model Fechnerian noise that depends on, for example, value differences between lotteries and question difficulty. More recently, He and Natenzon (2024) and Shubatt and Yang (2024) characterize Fechnerian noise that depends on a measure of distance between alternatives.

Finally, we characterize which families of utility functions, beyond CARA and CRRA, suffer from these reversals (Theorem 1) and propose alternative parametric families of utilities that do not suffer from the problem. This is in contrast with the solutions proposed by Wilcox (2011) and Apesteguia and Ballester (2018) who retain CARA and CRRA utilities.

Wilcox (2011) considers a WEU model and shows it is monotone for lotteries over three fixed outcomes. Nevertheless, Apesteguia and Ballester (2018) show that this model is no longer monotone for more than three outcomes and propose instead the random parameter model, where each individual is associated with a distribution over CARA or CRRA preferences and chooses probabilistically as if they draw a random preference from their distribution. Random parameter models do not give rise to reversals in binary choice, but they are based on a different foundation than

⁶Wilcox (2011) shows that other parameterizations of CARA and CRRA utilities also lead to reversals. This corresponds to two individuals having noise structures that are related by a scale factor. We study this special case in Proposition 2.

traditional models of noise. In particular, in random parameter models there is no longer a core preference over alternatives such that an individual usually chooses the more preferred outcome. This is in contrast with our proposal to use alternative vNM utility functions which maintains the idea that individuals choose according to a core preference under noise. Moreover, we show that the random parameter models advocated for by Apesteguia and Ballester (2018) can lead to what we call *compound reversals* when applied to menus with three or more alternatives. This is in contrast with the models we advocate for, which do not generate compound reversals in multinomial choice contexts.

Another paper that studies risk aversion in stochastic choice is Blavatsky (2011), which examines when one individual is always more likely to choose a sure thing over a lottery than another individual. A main result of that paper is that this notion of stochastic comparative risk aversion is too strong to be compatible with any Fechnerian expected utility models.

We compare the riskiness of lotteries using the concave order of Rothschild and Stiglitz (1970). In the deterministic setting, Kihlstrom, Romer, and Williams (1981) and Ross (1981) have noted that insurance and risk premia are not increasing in the Arrow-Pratt order. Ross (1981) thus proposes a stronger notion of comparative risk preferences than Arrow-Pratt that leads to monotonicity. Our characterization result (Theorem 1) shows that this stronger notion implies our condition, and thus is sufficient to eliminate reversals under some noise structures.

Finally, we contribute to a literature examining failures of expected utility theory in predicting choices across varying ranges of stakes. The most prominent critique in this area is by Rabin (2000), who demonstrates that within the expected utility framework, even modest levels of risk aversion exhibited over small-stakes, imply absurd, counterfactual levels of risk aversion for large stakes. While this result holds for deterministic choice, Propositions 2 and 3 echo Rabin (2000) in the stochastic setting: out of sample predictions made from small-stakes observations are counterintuitive.

1.2 Preliminaries

We denote by $\mathcal{L}$ the set of all bounded real random variables defined over a non-atomic probability space $(\Omega, \Sigma, \mathbb{P})$. We use the term *lotteries* to refer to elements of $\mathcal{L}$. A stochastic choice rule is a map $\rho: \mathcal{L} \times \mathcal{L} \rightarrow \mathbb{R}$ where $\rho(X, Y) + \rho(Y, X) = 1$ and $\rho(X, Y)$ is the probability of choosing X over Y . A vNM *utility function* is a

map $U: \mathcal{L} \rightarrow \mathbb{R}$ given by $U(X) = \mathbb{E}[u(X)]$ for some strictly increasing, concave, and continuous Bernoulli utility $u: \mathbb{R} \rightarrow \mathbb{R}$.

We consider an individual that evaluates lotteries according to a vNM utility U but faces noise and chooses stochastically according to ρ . We think of ρ as a noisy expression of the underlying utility U . For example, an individual with utility U whose noise takes a logit form has choice probabilities

$$\rho(X, Y) = \frac{e^{\beta U(X)}}{e^{\beta U(X)} + e^{\beta U(Y)}},$$

for some $\beta > 0$. When β is fixed across all menus, ρ belongs to the class of Fechnerian models, where choice probabilities are a function of utility differences. We discuss Fechnerian models in depth in Section 1.3. More generally, the logit scale parameter β may depend on (X, Y) . We study these general models in Section 1.4.

We say that X dominates Y in the *concave order*, denoted $X \succeq_c Y$ if $\mathbb{E}[g(X)] \geq \mathbb{E}[g(Y)]$ for all concave functions $g: \mathbb{R} \rightarrow \mathbb{R}$. We denote the strict part of $\succeq_c$ by $>_c$. Recall that $X \succeq_c Y$ if and only if Y is a mean-preserving spread of X . If $X \succeq_c Y$, then $U(X) \geq U(Y)$ for all vNM utilities U .⁷ We will often use the variables S and R when referring to lotteries where $S >_c R$, in order to highlight that S is *safer* while R is *riskier*.⁸ Finally, we denote by C_+^k the set of k -times continuously differentiable functions defined on an open interval of $\mathbb{R}$ with strictly positive k th derivatives.

Stochastic Comparative Risk Aversion

We are interested in comparing the choices of individual A with those of B . Each individual i is associated with a utility function U_i and stochastic choice model ρ_i . We will often refer to agents and their associated stochastic choice models interchangeably. We say A is *weakly more risk-averse than* B if $U_A(X) \geq U_A(y)$ implies $U_B(X) \geq U_B(y)$ for any $X \in \mathcal{L}$ and degenerate lottery y .⁹ We say A is *more risk-averse than* B if A is weakly more risk averse than B and B is not weakly more risk averse than A .

⁷Since we defined vNM utilities as expectations of concave Bernoulli utilities, they are increasing in $\succeq_c$.

⁸While we focus on lotteries that can be ranked in the concave order, analogous results hold throughout for lotteries comparable by second-order stochastic dominance. See Section 1.5.

⁹This is Yaari (1969)'s notion of comparative risk, which coincides with the Arrow-Pratt notion of comparative risk since U_A and U_B are vNM utility functions. I.e., U_A being more risk-averse than U_B is equivalent to the concavity of $u_A \circ u_B^{-1}$, when u_A and u_B are the corresponding Bernoulli utility functions to U_A and U_B . Indeed, CARA and CRRA utilities are totally ordered by this order.

Definition 1. When A is more risk-averse than B , we say that there is a reversal if there are lotteries $S \succ_c R$, where

$$\rho_A(S, R) < \rho_B(S, R).$$

Moreover, we say that S and R generate a reversal.

This means that A , the more risk-averse individual, exhibits more risk-tolerant behavior by choosing the safer lottery S less frequently than B . A 's behavior can also be seen as more risk-tolerant in the following sense: Let s_A and s_B denote A 's and B 's respective choice probabilities of S . The ex-ante lottery that A receives is the compound lottery resulting in S with probability s_A and in R with probability $1 - s_A$, and likewise for B . Since $s_A < s_B$, A 's compound lottery is riskier than B 's, so A evidently makes riskier choices than B . These two ways of interpreting reversals in binary choice naturally lead to different extensions of the notion of reversal to multinomial choice; see Section 1.6.

In the following sections, we show that many models of noise lead to reversals under the Constant Absolute Risk Aversion (CARA) and Constant Relative Risk Aversion (CRRA) utility specifications, the most commonly used parametric families in both theory and practice. CARA and CRRA utilities are defined by $\text{CARA}_a(X) = \mathbb{E}[\text{cara}_a(X)]$ and $\text{CRRA}_a(X) = \mathbb{E}[\text{crra}_a(X)]$, where $a \geq 0$ is the coefficient of absolute/relative risk aversion and

$$\text{cara}_a(x) = \begin{cases} x & a = 0 \\ \frac{1 - e^{-ax}}{a} & a > 0, \end{cases} \quad \text{crra}_a(x) = \begin{cases} \frac{x^{1-a} - 1}{1-a} & a \neq 1 \\ \ln(x) & a = 1. \end{cases}$$

The coefficients reflect varying levels of risk aversion, with higher coefficients corresponding to greater aversion to risk.¹⁰ To ensure that $\text{CRRA}_a(X)$ is well defined, we restrict our attention to X that is non-negative and bounded away from 0 whenever referring to CRRA.

1.3 Fechnerian Models

We first consider models of the form

$$\rho(X, Y) = F(U(X) - U(Y)), \tag{1.1}$$

¹⁰Note that there are many possible ways of parameterizing CARA and CRRA preferences, since applying a positive affine transformation to each Bernoulli utility does not change the underlying preference/coefficient. Nevertheless, we will show that our results hold for every possible parameterization of these families.

where $F: \mathbb{R} \rightarrow [0, 1]$ is a C_+^1 function satisfying $F(t) + F(-t) = 1$. We denote by $\mathcal{F}$ the set of all such functions and refer to a model of the form (U, F) , where U is a vNM utility and $F \in \mathcal{F}$, as a *Fechnerian expected utility* (FEU) model.¹¹ In these models, the larger the difference between $U(X)$ and $U(Y)$, the higher the probability that X is chosen.

FEU models are among the most widely used stochastic choice models (Becker, DeGroot, and Marschak, 1963; Loomes, Moffatt, and Sugden, 2002). An important subclass of FEU models yields choice probabilities

$$\mathbb{P}(U(X) + \varepsilon \geq U(Y) + \varepsilon') = F(U(X) - U(Y)), \quad (1.2)$$

where ε and ε' are i.i.d. continuous random variables and F is the CDF of $\varepsilon - \varepsilon'$. This subclass includes many widely used discrete choice models such as the logit and probit models, where the random shocks are Gumbel and Gaussian, respectively.

In this section, we focus on FEU models and show that the CARA and CRRA specifications display a paradox: they predict that a more risk-averse individual will sometimes choose a riskier asset more frequently than a more risk-tolerant individual under any Fechnerian noise specifications.

Proposition 1. *For any $F_A, F_B \in \mathcal{F}$ and distinct CARA (CRRA) utilities U_A, U_B , there exist reversals.*

Since CARA (CRRA) preferences are totally ordered by risk-aversion, without loss of generality suppose A is more risk averse. Proposition 1 says that there nevertheless exist lotteries $S >_c R$ such that $\rho_A(S, R) = F_A(U_A(S) - U_A(R)) < F_B(U_B(S) - U_B(R)) = \rho_B(S, R)$.

Proposition 1 is an impossibility result for the commonly used CARA/CRRA-expected-utility Fechnerian models since reversals arise under arbitrary pairs of noise structures and risk coefficients. In the case of $F_A = F_B$, i.e., when individuals face the same noise structures, it was shown by Wilcox (2011) and Apesteguia and Ballester (2018) that the probability of choosing the safer option from a pair is non-monotone in the CARA/CRRA coefficients.¹² In practice, however, it is

¹¹The assumptions that F must be continuously differentiable with positive derivative can be relaxed without affecting our results, provided these assumptions hold at 0. Note that we do not assume that $\lim_{t \rightarrow \infty} F(t) = 1$, nor that $\lim_{t \rightarrow -\infty} F(t) = 0$, although this will be the case for the examples we study.

¹²When $F_A = F_B$, Proposition 1 follows from Proposition 3 of Apesteguia and Ballester (2018).

common to model heterogeneity in both preferences and noise and to estimate them jointly (Hey and Orme, 1994; Von Gaudecker, Van Soest, and Wengström, 2011). Indeed, subsequent work has argued that non-monotonicity is particularly worrisome under homoskedastic noise and has advocated allowing heteroskedasticity for better calibration of each individual's noise to their utility scale (Barseghyan et al., 2018; O'Donoghue and Somerville, 2024; Keffert and Schweizer, 2024).

Yet, perhaps surprisingly, Proposition 1 demonstrates that non-monotonicity persists even when noise structures vary arbitrarily across individuals. Thus, no matter how we jointly estimate risk coefficients and noise structures for two individuals, the resulting estimated models (CARA_a, F_A) and (CARA_b, F_B) will predict reversals. Because reversals are inherent to these models, we must be cautious when using these estimates to rank individuals by risk aversion.

Note that choice probabilities are not invariant to reparameterizations of CARA and CRRA utilities. For example, consider the model (U, F) , where U is a CARA or CRRA utility and $F \in \mathcal{F}$. Let $V(X) = \frac{1}{c}U(X) + d$ for some $c > 0$. While V represents the same preference as U , replacing U with V has the same effect on choice probabilities as changing the noise structure from F to G which is defined by $G(t) = F(t/c)$. However, since G is also a member of $\mathcal{F}$, it follows from Proposition 1 that reparameterizations of the CARA or CRRA utility family cannot resolve the non-monotonicity.

Scale-Family Heteroskedasticity

To develop intuition for Proposition 1, we first consider individuals with distinct CARA or CRRA utilities U_A and U_B with respective coefficients a and b , where $a > b$ and whose noise structures are related by a scale factor. That is, $F_A(t) = F(t/\sigma_A)$ and $F_B(t) = F(t/\sigma_B)$ for some $\sigma_A, \sigma_B > 0$ and common $F \in \mathcal{F}$.¹³ The choice probabilities of S over R are then given by

$$F\left(\frac{U_A(S) - U_A(R)}{\sigma_A}\right) \tag{1.3}$$

$$F\left(\frac{U_B(S) - U_B(R)}{\sigma_B}\right). \tag{1.4}$$

Since F is strictly increasing, the pair (S, R) generates a reversal (i.e., (1.4) is strictly larger than (1.3)) if and only if

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)} > \frac{\sigma_B}{\sigma_A}. \tag{1.5}$$

¹³This is the case, for example, when F_A and F_B are both normal or both logistic CDFs.

Note that $\frac{\sigma_B}{\sigma_A}$, which is the relative noise level, does not depend on the lotteries. Given this, the next lemma immediately implies that to generate a reversal, we can start with any $S >_c R$, and get a reversal by adding to both a large enough constant x (in the CARA case) or multiplying both by a large enough constant (in the CRRA case).

Lemma 1. *For any $S >_c R$, $a, b > 0$, and all $x \in \mathbb{R}$ and $k > 0$,*

$$\begin{aligned} 1. \quad & \frac{\text{CARA}_b(S+x) - \text{CARA}_b(R+x)}{\text{CARA}_a(S+x) - \text{CARA}_a(R+x)} = C_1 \cdot e^{(a-b)x} \\ 2. \quad & \frac{\text{CRRA}_b(k \cdot S) - \text{CRRA}_b(k \cdot R)}{\text{CRRA}_a(k \cdot S) - \text{CRRA}_a(k \cdot R)} = C_2 \cdot k^{a-b}, \end{aligned}$$

for some positive constants C_1 and C_2 .

Lemma 1 is proved in Section 1.A. Note that when $a > b$, the ratios are strictly increasing in x and k and tend to infinity. Moreover, these properties hold for any $S >_c R$. We thus have the following result which strengthens the existence of reversal pairs established in Proposition 1 when noise structures are related by a scale factor.

Proposition 2. *Let $\sigma_A, \sigma_B > 0$ and $F \in \mathcal{F}$. Let $F_A(t) = F(\frac{t}{\sigma_A})$, $F_B(t) = F(\frac{t}{\sigma_B})$, and $S >_c R$. Let U_A, U_B be distinct CARA utilities and let V_A, V_B be distinct CRRA utilities. Then there exist unique x_0 and k_0 such that*

1. $S+x$ and $R+x$ generate a reversal for (U_A, F_A) and (U_B, F_B) if and only if $x > x_0$;
2. kS and kR generate a reversal for (V_A, F_A) and (V_B, F_B) if and only if $k > k_0$.

Note that the critical values x_0 and k_0 given in Proposition 2 depend on the noise scale parameters σ_A, σ_B , risk coefficients, and the baseline lotteries. The values can be derived from (1.5) and the positive constants C_1 and C_2 given in Lemma 1. I.e., critical values x_0 and k_0 solve for $C_1 \cdot e^{(a-b)x} = \frac{\sigma_B}{\sigma_A}$ and $C_2 \cdot k^{(a-b)} = \frac{\sigma_B}{\sigma_A}$.¹⁴ For

¹⁴Exact expressions for C_1 and C_2 are given in (1.14). We obtain

$$x_0 = \frac{1}{a-b} \ln \left(\frac{\sigma_B}{\sigma_A} \frac{b}{a} \left(\frac{\mathbb{E}[e^{-aR}] - \mathbb{E}[e^{-aS}]}{\mathbb{E}[e^{-bR}] - \mathbb{E}[e^{-bS}]} \right) \right), \quad k_0 = \left(\frac{\sigma_B}{\sigma_A} \frac{1-b}{1-a} \left(\frac{\mathbb{E}[S^{1-a}] - \mathbb{E}[R^{1-a}]}{\mathbb{E}[S^{1-b}] - \mathbb{E}[R^{1-b}]} \right) \right)^{\frac{1}{a-b}}.$$

example, when the CRRA coefficients a and b are 0.8 and 0.3, respectively, and the noise variances are $\sigma_A = 0.5$ and $\sigma_B = 1$, the critical values that generate reversals for lotteries given in Table 1.1 are all less than 3, which illustrates that reversals often occur under empirically relevant risk coefficients and lotteries.

Proposition 2 highlights two problems with the CARA and CRRA Fechnerian models. First, there is a disconnect between the standard Arrow-Pratt notion of comparative risk-aversion developed in the deterministic framework with the resulting stochastic choice predictions. Namely, more risk-averse individuals are not more likely to make risk-averse choices. This counterintuitive prediction is contradicted by empirical evidence that risk aversion is negatively correlated with decision noise, as measured by the frequency with which individuals choose the concave-order dominated option (Bruner, 2017). In terms of estimation, this means that even if Bob chooses safer lotteries more frequently than Anne, we may conclude that Bob is more risk-tolerant than Anne.¹⁵ This is the case, for example, if we only observe the choice probabilities of the scaled lotteries in Table 1.1.

Second, these models make the systematic prediction that those most inclined to choose safer options when stakes are small must become the least inclined once the stakes grow modestly larger. Even without a specific model of risk aversion in mind, this prediction is implausible.

Arbitrary Fechnerian Noise Structures

So far we have considered the special case that F_A and F_B are related by a scale factor. To prove the more general case considered in Proposition 1, we establish the following lemma, which shows that the properties of CARA and CRRA utilities given in Lemma 1 give rise to non-monotonicity for any $F_A, F_B \in \mathcal{F}$.

Lemma 2. *Let $F_A, F_B \in \mathcal{F}$. Let U_A and U_B be vNM utility functions with U_A more risk averse. Suppose that for each $M \in \mathbb{R}$ there exist lotteries $S >_c R$ such that*

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)} \geq M. \quad (1.6)$$

Then there are reversals for (U_A, F_A) and (U_B, F_B) .

Lemma 2 identifies a sufficient condition on the vNM utility functions to yield the negativity result in Proposition 1. By Lemma 1, CARA and CRRA utilities satisfy this condition, and thus Proposition 1 follows immediately from these lemmas.

¹⁵This holds under any consistent estimator.

Let $S >_c R$ satisfying (1.6) for some $M \in \mathbb{R}$. The key step of the proof of Lemma 2 is to construct $S' >_c R'$ such that

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)} = \frac{U_B(S') - U_B(R')}{U_A(S') - U_A(R')},$$

while making the utility differences arbitrarily small. To this end, we consider the menu $(S, R_\lambda S)$, where $\lambda \in (0, 1)$ and $R_\lambda S$ is distributed as a compound lottery that yields R with probability λ and S with probability $1 - \lambda$. As $\lambda \rightarrow 0$, the utility difference between S and $R_\lambda S$ vanishes. Lemma 2 exploits the linearity of vNM utilities, which ensures that the ratio of utility differences is independent of λ . It also relies on the differentiability of F_A and F_B to approximate them around 0 by affine functions. Lemma 2 is proved in Section 1.A.

Note that by Lemma 1, (1.6) is satisfied for all concave-ordered lotteries under sufficient shifting/scaling. Moreover, for CARA, the utility difference $U(S + x) - U(R + x)$ tends to 0 as x tends to infinity (see Lemma 4 in Section 1.B). For CRRA, on the other hand, $U(kS) - U(kR)$ increases with k for relative risk coefficients less than unity. We therefore apply an additional transformation of scaling down the probability of receiving the risky prospect so that the utility difference is arbitrarily small. Based on these observations, the following proposition extends the non-monotonicity result of Proposition 2 to all $F_A, F_B \in \mathcal{F}$.

Proposition 3. *Let $F_A, F_B \in \mathcal{F}$ and $S >_c R$. Let U_A and U_B be distinct CARA utilities and V_A and V_B be distinct CRRA utilities. Then there exist $x_0, k_0 > 0$ such that*

1. $S + x$ and $R + x$ generate a reversal for (U_A, F_A) and (U_B, F_B) for $x > x_0$.
2. $k \cdot S$ and $k \cdot (R_\lambda S)$ generate a reversal for (V_A, F_A) and (V_B, F_B) for $k > k_0$ and λ small enough.

Proposition 3 is proved in Section 1.B. While CARA preferences are invariant to changes in background wealth, the above proposition shows that non-monotonicity occurs in FEU models under CARA utilities for any concave-ordered lotteries under sufficient background wealth. Likewise, while CRRA preferences are invariant to the scaling of stakes and to mixing with the safe lottery, these transformations can always generate a non-monotonicity.

Proposition 3 demonstrates that non-monotonicity is a pervasive problem. It holds for all pairs of risk parameters, for all pairs of Fechnerian noise specifications, and

for all concave-ordered lotteries after a transformation that preserves the concave ordering as well as the family of preferences being measured.

1.4 Weak Utility Models

In the previous section, we studied models in which noise varies across individuals but not across menus for the same individual. In this section, we study a more flexible model that incorporates menu-dependent noise and show that the reversals under CARA/CRRA utilities persist.

Formally, ρ has a *weak expected-utility* (WEU) representation (U, F, σ) if

$$\rho(X, Y) = F\left(\frac{U(X) - U(Y)}{\sigma(X, Y)}\right), \quad (1.7)$$

where U is a vNM utility function, $F \in \mathcal{F}$ is a Fechnerian noise structure, and $\sigma: \mathcal{L} \times \mathcal{L} \rightarrow (0, \infty)$ is a menu-dependent scale parameter satisfying $\sigma(X, Y) = \sigma(Y, X)$. Symmetry of σ is required so that $\rho(X, Y) + \rho(Y, X) = 1$.

An example of WEU is the *menu-dependent probit model*. In this model, the shock terms ε and ε' in (1.2) follow normal distributions, but the identity assumption is relaxed to allow their variances to depend on the lotteries. That is,

$$\rho(X, Y) = \mathbb{P}(U(X) + \varepsilon_X \geq U(Y) + \varepsilon_Y),$$

where ε_X and ε_Y are independent and normal random variables with means zero and variances σ_X and σ_Y .¹⁶ For example, σ_X may depend on the variance of lottery X in the menu. This can capture the intuitive idea that more variable lotteries may generate more noise, allowing for mistake probabilities to remain high even when stakes are high.

Another example is the *strict utility model* of Luce and Suppes (1965) in which the choice probabilities are determined by the utility ratio

$$\rho(X, Y) = \frac{U(X)}{U(X) + U(Y)},$$

rather than by the utility difference as in FEU models (see (1.1)) of the previous section. Indeed, any *simply scalable expected utility* (SSEU) representation, which takes the form

$$\rho(X, Y) = H(U(X), U(Y)), \quad (1.8)$$

¹⁶In this case, ρ can be written as (1.7) where F is the CDF of the standard normal distribution and $\sigma(X, Y) = \sqrt{\sigma_X^2 + \sigma_Y^2}$.

where $H : \mathbb{R}^2 \rightarrow [0, 1]$ satisfies $H(s, t) + H(t, s) = 1$ and is strictly increasing in its first argument, has a WEU representation.¹⁷

As the examples above suggest, the menu-dependent scale parameter σ allows for a highly flexible stochastic choice model. In fact, without imposing restrictions on σ , the sole behavioral implication of WEU is that individuals select their preferred lottery with probability greater than one-half.¹⁸ In order to rule out pathological noise assignments, we require that menus with similar lotteries are associated with similar menu-dependent scale parameters. We say that a sequence of lotteries $X_1, X_2, \dots$ converges to a lottery X if for all continuous functions u it holds that

$$\mathbb{E}[|u(X_n) - u(X)|] \rightarrow 0.$$

In other words, $X_1, X_2, \dots$ converges to X if the sequence of random utilities $u(X_1), u(X_2), \dots$ converges to $u(X)$ in L^1 for every continuous Bernoulli u . We say that σ is *continuous* if whenever $X_n \rightarrow X$ and $Y_n \rightarrow Y$, it holds that $\sigma(X_n, Y_n) \rightarrow \sigma(X, Y)$. We say a WEU representation (U, F, σ) is *continuous* when σ is continuous.¹⁹

Importantly, vNM utilities are continuous, i.e., $U(X_n) \rightarrow U(X)$ whenever $X_n \rightarrow X$, so we can allow σ to depend continuously on U . Moreover, σ is not required to depend only on the marginal distributions of the lotteries in the menu; it may, for example, depend on their joint distribution.

Characterization

Our next theorem characterizes when two vNM utilities necessarily lead to reversals under any continuous WEU models. Given two increasing and concave Bernoulli utility functions u_A and u_B , we say that u_A is *absolutely more concave than* u_B , and write $u_A \geq_{\text{abs}} u_B$ if $u_A - u_B$ is concave. Moreover, we say that u_A is *relatively more*

¹⁷To see that the simply scalable model is a special case of the weak utility model, take

$$\sigma(X, Y) = \begin{cases} \frac{U(X) - U(Y)}{F^{-1}(H(U(X), U(Y)))} & U(X) \neq U(Y) \\ s > 0 & U(X) = U(Y) \end{cases}$$

in (1.7) to derive $\rho(X, Y) = H(U(X), U(Y))$.

¹⁸Since $F(t/\sigma(X, Y)) \in \mathcal{F}$ for all $(X, Y) \in \mathcal{L} \times \mathcal{L}$, if $U(X) - U(Y) \geq 0$, then $F\left(\frac{U(X) - U(Y)}{\sigma(X, Y)}\right) \geq \frac{1}{2}$ with equality if and only if $U(X) = U(Y)$. Conversely, if $U(X) > U(Y)$, then for any $p \in (\frac{1}{2}, 1)$, letting $\sigma(X, Y) = \frac{U(X) - U(Y)}{F^{-1}(p)}$ yields the desired probability.

¹⁹Formally, we require σ to be sequentially continuous in the product topology on $\mathcal{L} \times \mathcal{L}$, when $\mathcal{L}$ is equipped with the coarsest topology making each $X \mapsto u(X)$ continuous as an L^1 -valued map. This topology is very fine (finer than any L^p topology for $1 \leq p < \infty$), making many functions on $\mathcal{L}$ continuous.

concave than u_B , and write $u_A \geq_{\text{rel}} u_B$ if there exists $k > 0$ such that $k \cdot u_A(x) - u_B(x)$ is concave.²⁰ We extend these notions to vNM utilities $U_A = \mathbb{E}[u_A]$ and $U_B = \mathbb{E}[u_B]$, writing $U_A \geq_{\text{abs}} U_B$ if $u_A \geq_{\text{abs}} u_B$ and $U_A \geq_{\text{rel}} U_B$ if $u_A \geq_{\text{rel}} u_B$. For twice differentiable functions with $u''_A < 0$, the condition that $u_A \geq_{\text{rel}} u_B$ is equivalent to the boundedness of $u''_B(x)/u''_A(x)$, i.e., that $u''_B = O(u''_A)$ on $\mathbb{R}$.

Theorem 1. *Let U_A and U_B be vNM utilities with U_A more risk averse and let $F_A, F_B \in \mathcal{F}$. There exist continuous menu-dependent scale parameters σ_A and σ_B for which no reversals occur if and only if $U_A \geq_{\text{rel}} U_B$.*

The necessity of $U_A \geq_{\text{rel}} U_B$ to avoid reversals is proved in Section 1.C. A key idea is to show that if U_A is not relatively more concave than U_B , then the ratio

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)}$$

is unbounded over the set of lottery pairs $S >_c R$. Under continuity of σ_A and σ_B , we show that this unboundedness leads to reversals using an argument similar to that of Proposition 1. This argument relies on the fact that continuity imposes a bound on the relative noisiness of A and B for some lottery comparisons. We note that limitations on relative noisiness lead to reversals even under some discontinuous menu-dependent scale parameters. For example, for (U_A, F, σ_A) and (U_B, F, σ_B) , if $\sigma_B(S, R)/\sigma_A(S, R)$ is bounded, there will necessarily be reversals even if σ_A and σ_B are not continuous.

We prove sufficiency of $U_A \geq_{\text{rel}} U_B$ here by showing that even under a simple heteroskedastic probit model, without menu dependence, this condition suffices. Indeed, suppose that $ku_A - u_B$ is concave for some $k > 0$. Let F be the CDF of the standard normal distribution. Let $\sigma_A(X, Y) = 1/k$ and $\sigma_B(X, Y) = 1$ for all X, Y . Then for any lotteries $S \geq_c R$,

$$\rho_A(S, R) = F(k(U_A(S) - U_A(R))) \quad (1.9)$$

$$\rho_B(S, R) = F(U_B(S) - U_B(R)). \quad (1.10)$$

Since $ku_A - u_B$ is concave, $kU_A(S) - U_B(S) \geq kU_A(R) - U_B(R)$. Hence, since F is strictly increasing, (1.9) exceeds (1.10), meaning that there are no reversals.²¹

²⁰Recall that u_A is more risk averse than u_B in the Arrow-Pratt sense if $u_A = f \circ u_B$ for some increasing and concave function f . This notion of comparative concavity is logically independent of both $\geq_{\text{abs}}$ and $\geq_{\text{rel}}$. In Section 1.D, we show that the stronger comparative risk notion of Ross (1981) implies $\geq_{\text{rel}}$ and discuss an interpretation of his stronger notion in the stochastic framework.

²¹Note that for this direction, we can relax the requirement that $F \in C^1_+$.

Theorem 1 establishes a simple test for identifying whether a pair of vNM utilities generates reversals. Indeed, for CARA and CRRA utilities, $\text{cara}_b''(x)/\text{cara}_a''(x) = (b/a)e^{-(b-a)x}$ and $\text{crra}_b''(x)/\text{crra}_a''(x) = (b/a)x^{a-b}$ are unbounded for $a > b$. Thus, it follows from Theorem 1 that reversals of Fechnerian CARA and CRRA models cannot be eliminated by allowing continuous menu-dependent scale parameters.

Corollary 1. *Let $F_A, F_B \in \mathcal{F}$ and let σ_A and σ_B be continuous. Then for any distinct CARA (CRRA) utilities U_A and U_B there are reversals.*

Corollary 1 highlights that the non-monotonicity result of the menu-independent FEU models (Proposition 1) is not merely a byproduct of asymptotic properties of choice probabilities when stakes are increased. Indeed, in the following example scale parameters are engineered so that choice probabilities are scale-invariant, yet monotonicity does not obtain.

Example 1. Let $a > b > 0$, $U_A = \text{CRRA}_a$, and $U_B = \text{CRRA}_b$. Let F be the CDF of the standard normal distribution. Let

$$\sigma_A(X, Y) = \mathbb{E}[X^{a-1} + Y^{a-1}]^{-1}, \quad \sigma_B(X, Y) = \mathbb{E}[X^{b-1} + Y^{b-1}]^{-1}.$$

Then for all $k > 0$ and for $i \in \{A, B\}$, we have

$$F\left(\frac{U_i(kX) - U_i(kY)}{\sigma_i(kX, kY)}\right) = F\left(\frac{U_i(X) - U_i(Y)}{\sigma_i(X, Y)}\right).$$

When $a \geq 1$, σ_A decreases as X and Y are scaled up, capturing the behavioral property that individuals pay more attention when facing higher stakes. Moreover, the standard deviations are chosen so that each individual's probability of choosing βX over βY does not depend on $\beta > 0$, as $\sigma_A(\beta X, \beta Y) = \beta^{1-a}\sigma_A(X, Y)$ and

$$U_A(\beta X) - U_A(\beta Y) = \beta^{1-a}(U_A(X) - U_A(Y)).$$

Since $\sigma_A(X_n, Y_n) \rightarrow \sigma_A(X, Y)$ whenever $X_n \rightarrow X$ and $Y_n \rightarrow Y$, the model is continuous. Thus, it follows from Corollary 1 that even these calibrated models cannot avoid the paradox.

Another example of continuous WEU is when ρ admits an SSEU representation (U, H) given in (1.8), above, where H is further assumed to be differentiable and its first partial derivative, H_1 , is continuous and positive. Indeed, for

$$\sigma(X, Y) = \begin{cases} \frac{U(X) - U(Y)}{F^{-1}(H(U(X), U(Y)))} & U(X) \neq U(Y) \\ \frac{F'(F^{-1}(H(U(Y), U(Y))))}{H_1(U(Y), U(Y))} & U(X) = U(Y), \end{cases}$$

the WEU model (U, F, σ) coincides with the SSEU model (U, H) . Moreover, σ is continuous since U, F^{-1}, F', H , and H_1 are continuous and the two cases coincide in the limit as $U(X_n)$ and $U(Y_n)$ tend to the same value.

One may guess, at this point, that reversals arise in WEU models under CARA and CRRA utilities because the Fechnerian noise and menu-dependent scale parameter are applied to the vNM utility U , which is merely a representation of the preference and carries no cardinal significance. A natural alternative would be to calculate, for each lottery X , its certainty equivalent $CE(X) = u^{-1}(\mathbb{E}[u(X)])$, and let

$$\rho(X, Y) = F\left(\frac{CE(X) - CE(Y)}{\sigma(X, Y)}\right).$$

However, the next result demonstrates that this adjustment does not resolve the issue. In fact, reversals persist even if we more generally consider utility representations of the form $f \circ U$ for an arbitrary $f \in C_+^1$.

Corollary 2. *Let $F_A, F_B \in \mathcal{F}$, $f_A, f_B \in C_+^1$ and let σ_A and σ_B be continuous. For any distinct CARA (CRRA) utilities U_A and U_B , there are reversals for $(f_A \circ U_A, F_A, \sigma_A)$ and $(f_B \circ U_B, F_B, \sigma_B)$.*

Corollary 2 shows that reversals are a fundamental feature of CARA and CRRA preferences, not an artifact of the utility representation. The underlying mechanism is straightforward: for any transformation $f \in C_+^1$, the WEU model $(f \circ U, F, \sigma)$ has a WEU representation (U, F, τ) , where the new menu-dependent scale parameter τ remains continuous.²² Given this, the result follows directly from Corollary 1. Since inverses of cara and crra belong to C_+^1 , replacing CARA and CRRA expected utilities with their certainty equivalents cannot eliminate reversals.

Corollaries 1 and 2 demonstrate that CARA and CRRA utilities generate reversals whenever choices are noisy, even if we use very general utility representations and allow noise structures to vary across individuals and menus. Thus, if we hope to model the stochastic choice behavior of A and B , where A consistently chooses safer lotteries more frequently than B , we must forgo CARA and CRRA preferences.

²²Here,

$$\tau(X, Y) = \begin{cases} \frac{U(X) - U(Y)}{f(U(X)) - f(U(Y))} \sigma(X, Y) & U(X) \neq U(Y) \\ \frac{\sigma(X, Y)}{f'(U(Y))} & U(X) = U(Y), \end{cases}$$

which is continuous since U, f, f' , and σ are continuous and the two cases coincide in the limit.

Moreover, the boundedness of u_B''/u_A'' is violated by many parametric families of utilities that are used to model risk aversion. In particular, we show in Section 1.E that expo-power utility—a popular generalization of the CARA/CRRA families—violates the condition, leading to reversals under any continuous WEU models. We take on these challenges in the next section, where we suggest alternative vNM utilities that can accommodate consistent comparative statics under standard models of noise.

1.5 Monotone Alternatives to CARA and CRRA

In the previous section, we showed that when two vNM utilities have an unbounded ratio of second derivatives, there are necessarily reversals under very general noise structures. Conversely, when the ratio of second derivatives is bounded, there exist heteroskedastic noise structures such that there are no reversals. In this section, we demonstrate when two vNM utilities will not lead to reversals under any homoskedastic Fechnerian models and, more generally, under homoskedastic WEU models. We provide two parametric families of utilities such that any two utilities within either family satisfy this condition. We then discuss the economic properties of these utility families.

Recall that $u_A \geq_{\text{abs}} u_B$ if $u_A - u_B$ is concave. The next proposition states that this absolute comparative concavity notion leads to no reversals under identical Fechnerian noise.

Proposition 4. *Let $F \in \mathcal{F}$ and fix U_A more risk-averse than U_B . Then there are no reversals under (U_A, F) and (U_B, F) if and only if $U_A \geq_{\text{abs}} U_B$.*

Proof of Proposition 4. Since Fechnerian noise structures $F \in \mathcal{F}$ are strictly increasing, the choice probability of S is higher for A than B if and only if the utility difference is higher for A than B , i.e.,

$$U_A(S) - U_A(R) \geq U_B(S) - U_B(R),$$

or equivalently,

$$U_A(S) - U_B(S) \geq U_A(R) - U_B(R). \quad (1.11)$$

We now show that this inequality holds for all lotteries $S \geq_c R$ if and only if $U_A \geq_{\text{abs}} U_B$.

Let $h = u_A - u_B$. If h is concave, then (1.11) is satisfied by the definition of the concave order. Conversely, for the sake of contradiction, suppose that h is not concave. Then there are $x, y \in \mathbb{R}$ and $\lambda \in (0, 1)$ such that

$$h(\lambda x + (1 - \lambda)y) < \lambda h(x) + (1 - \lambda)h(y).$$

Let S denote a degenerate lottery that always pays $\lambda x + (1 - \lambda)y$, and let R denote a lottery that pays x with probability λ and y with complementary probability. Then

$$U_A(S) - U_B(S) = h(\lambda x + (1 - \lambda)y) < \lambda h(x) + (1 - \lambda)h(y) = U_A(R) - U_B(R),$$

contradicting (1.11). Thus $u_A - u_B$ is concave, meaning $U_A \geq_{\text{abs}} U_B$. $\square$

In the case where A and B have different noise structures, i.e., $F_A \neq F_B$, a sufficient condition for no reversals is if $U_A \geq_{\text{abs}} U_B$ and F_A is *more precise than* F_B , in the sense that $F_A(t) \geq F_B(t)$ for $t \geq 0$. A more precise noise structure translates the same utility difference into a higher probability of choosing a more preferred alternative. Since A is risk averse, increasing A 's precision will only increase the probability that A chooses S . In particular,

$$F_A(U_A(S) - U_A(R)) \geq F_A(U_B(S) - U_B(R)) \geq F_B(U_B(S) - U_B(R)),$$

where the first inequality follows from Proposition 4.²³ Of course, A may experience much more noise than B , in which case A will make more mistakes and may choose riskier options more frequently than B , even with $U_A \geq_{\text{abs}} U_B$.

Parametric Families of Utility Functions

We conclude by proposing parametric utility families ordered by both risk aversion and $\geq_{\text{abs}}$, leading to monotone comparative statics under identical (menu-dependent) Fechnerian noise. Recall that for strictly increasing concave functions f and g , f is *more concave* than g if $f \circ g^{-1}$ is increasing and concave. The following proposition provides a simple way to construct a parametric utility family that yields consistent comparative statics.

Proposition 5. *Let f and g be strictly increasing concave functions such that f is more concave than g . Let $u_A(x) = af(x) + g(x)$ and $u_B(x) = bf(x) + g(x)$, where $a > b \geq 0$. Then $U_A = \mathbb{E}[u_A]$ is more risk-averse than $U_B = \mathbb{E}[u_B]$ and $U_A \geq_{\text{abs}} U_B$.*

²³The same conclusion holds under WEU models when A and B face the same menu-dependent scale parameters or when $\sigma_A(S, R) \leq \sigma_B(S, R)$ for all $S \geq_c R$.

These utilities are parametrized by the coefficient on the more risk-averse Bernoulli utility function f , i.e., the higher the coefficient, the greater the risk aversion.²⁴ Moreover, $u_A - u_B = (a - b)f$ is concave since f is concave, so $u_A \geq_{\text{abs}} u_B$, and increasing, since f is increasing and $a > b$. Thus, monotonicity of choice probabilities in the risk coefficient also holds for all lotteries comparable by second-order stochastic dominance (SOSD). We write $X \geq_{\text{ssd}} Y$ if X dominates Y in the SOSD order.

Remark 1. Let f and g be strictly increasing concave functions such that f is more concave than g . Let $u_A(x) = af(x) + g(x)$ and $u_B(x) = bf(x) + g(x)$, where $a > b \geq 0$. Then for any $F \in \mathcal{F}$,

$$F(U_A(X) - U_A(Y)) \geq F(U_B(X) - U_B(Y)) \quad (1.12)$$

for all $X \geq_{\text{ssd}} Y$.

The contrast between the utility families in Proposition 5 and standard functional forms, such as CARA and CRRA, is stark. Under Fechnerian noise, the latter admit reversals under the concave order, whereas our proposed families preclude reversals even under the more general criterion of second-order stochastic dominance.

When f and g are CARA (CRRA) utilities with different coefficients, u_A and u_B correspond to sum-ex (sum-power) utilities (see, e.g., Farquhar and Nakamura, 1987; Bell, 1988; Pedersen and Satchell, 2001). Proposition 5 thus suggests a replacement of the CARA utility family or CRRA utility family for estimation. One specifies an upper and lower bound on absolute risk aversion, corresponding to two CARA utilities, and then estimates the weight on each utility. The same exercise can be done with bounds on relative risk aversion and CRRA utilities. In Section 1.F, we discuss properties of sum-ex and sum-power utility functions and how to interpret the parameters.

1.6 Multinomial Choice

In this section, we extend our results to multinomial choice. Given a finite menu $M \subset \mathcal{L}$, we denote by $\rho^M \in \Delta(M)$ the choice probabilities for the menu M .

²⁴Indeed,

$$-\frac{u_A''(x)}{u_A'(x)} \geq -\frac{u_B''(x)}{u_B'(x)} \iff (a - b) \left(-\frac{f''(x)}{f'(x)} + \frac{g''(x)}{g'(x)} \right) \geq 0.$$

Since f is more concave than g , it follows that $a \geq b$ is equivalent to U_A being more risk-averse than U_B .

There are a number of ways to extend the notions of a reversal to multinomial choice. We discuss two notions that seem natural.

Definition 2. When A is more risk-averse than B , we say that there is a compound reversal if there is a menu M such that for all concave functions $u: \mathbb{R} \rightarrow \mathbb{R}$

$$\sum_{X \in M} \mathbb{E}[u(X)] \rho_B^M(X) \geq \sum_{X \in M} \mathbb{E}[u(X)] \rho_A^M(X),$$

with strict inequality for at least one such u .

When a compound reversal occurs at M , the compound lottery resulting from A 's stochastic choice is dominated by that of B in the concave order. The following alternative notion requires that for each set $N \subset M$ that is upward-closed in the concave order,²⁵ B is at least as likely as A to choose a lottery in N and strictly more likely for at least one such N .

Definition 3. When A is more risk-averse than B , we say that there is a FOSD reversal if there is a menu M such that for all functions $f: \mathcal{L} \rightarrow \mathbb{R}$ that are increasing in $\geq_c$

$$\sum_{X \in M} f(X) \rho_B^M(X) \geq \sum_{X \in M} f(X) \rho_A^M(X),$$

with strict inequality for at least one such f .

By choosing functions f that are expectations of strictly concave real functions in Definition 3, we see that an FOSD reversal is always a compound reversal. The converse holds only for menus of size two, in which case both notions coincide with Definition 1.

Additive Perturbed Utility Models

In this section, we study multinomial choice reversals in the *additive perturbed utility* (APU) framework of Fudenberg, Iijima, and Strzalecki (2015), wherein

$$\rho^M = \operatorname{argmax}_{p \in \Delta(M)} \sum_{X \in M} U(X)p(X) - c(p(X)), \quad (1.13)$$

where U is a vNM utility function and $c: [0, 1] \rightarrow \mathbb{R} \cup \{\infty\}$ is strictly convex and continuously differentiable on $(0, 1)$ and satisfies $\lim_{p \rightarrow 0} c'(p) = -\infty$. Any function c with these properties is referred to as a *cost function*. Both U and c

²⁵ N is upward closed in the concave order if for all $S \geq_c R$, $R \in N$ implies $S \in N$.

are fixed across menus. We say that ρ has an *APU representation* (U, c) if ρ coincides with (1.13). Fudenberg, Iijima, and Strzalecki (2015) show that APU is a generalization of Fechnerian choice to the multinomial setting; see also Section 1.G.

Our next result shows that reversals occur in APU models with arbitrarily heterogeneous cost functions unless the $\geq_{\text{rel}}$ order obtains.

Theorem 2. *Let U_A and U_B be vNM utilities with U_A more risk averse. The following are equivalent:*

1. *There exist cost functions $c_A, c_B \in C_+^2$ for which no compound reversals occur;*
2. *There exist cost functions $c_A, c_B \in C_+^2$ for which no FOSD reversals occur;*
3. $U_A \geq_{\text{rel}} U_B$.

Theorem 2 is proved in Section 1.G, in which we further show that when ρ_A and ρ_B have APU representations with the same cost function, reversals of both kinds will occur in some menus unless $U_A \geq_{\text{abs}} U_B$, extending Proposition 4 to the multinomial setting (Proposition 10). Given that APU is an extension of the FEU (see Lemma 7), we already know from the earlier sections that reversals will occur under heterogeneous (homogeneous) costs if the $\geq_{\text{rel}}$ ($\geq_{\text{abs}}$) order does not obtain. The message of this section is thus that when these orders do obtain, reversals can be avoided even for multinomial choice.

Random Parameter Models

As discussed above, Apesteguia and Ballester (2018) raise the issue of reversals for CARA and CRRA models of binary choice under homogeneous Fechnerian noise. As an alternative, they advocate for CARA- and CRRA-based *random parameter* models, where there is a continuous random variable $\varepsilon \geq 0$ such that an individual with CARA or CRRA risk coefficient $a \geq 0$ chooses the lottery X over Y with probability²⁶

$$\rho(X, Y) = \mathbb{P}(U_{a+\varepsilon}(X) \geq U_{a+\varepsilon}(Y)).$$

Clearly if $X >_c Y$ then every realization of $U_{a+\varepsilon}$ would evaluate X greater than Y , since every possible realized utility function exhibits risk aversion. Since lotteries

²⁶Technically, this is not a well-defined probabilistic choice rule when the probability of ties is non-zero. We will ignore the question of how to deal with ties, which has no bearing on our analysis.

dominated in the concave order are never chosen in these models, there cannot, of course, be reversals.²⁷

The random parameter model is often considered in multinomial settings with the obvious extension:

$$\rho^M(X) = \mathbb{P}(U_{a+\varepsilon}(X) = \operatorname{argmax}_{Y \in M} U_{a+\varepsilon}(Y)).$$

While these models preclude reversals in binary menus, our next result shows that the situation is more subtle for larger menus.

Proposition 6. *The CARA- and CRRA-based random parameter models can lead to compound reversals but not FOSD reversals.*

The reason that random parameter models do not lead to FOSD reversals is much the same as the reason that they do not admit reversals in binary menus. To see this formally, suppose that individual A has random utility $U_{a+\varepsilon}$ and B has random utility $U_{b+\varepsilon}$, where $a > b$, so that A is more risk averse than B . Fix a menu M and let $N \subset M$ denote the set of lotteries that A chooses with positive probability in the random parameter model. Since A never chooses dominated lotteries, for $Y \in N$, the indicator function $\mathbb{1}_Y$ is increasing in $\succeq_c$. Thus, for there to be an FOSD reversal,

$$\rho_B^M(Y) = \sum_{X \in M} \mathbb{1}_Y \rho_B^M(X) \geq \sum_{X \in M} \mathbb{1}_Y \rho_A^M(X) = \rho_A^M(Y).$$

Since this must hold for all $Y \in N$, we conclude that $\rho_B^M = \rho_A^M$, which precludes an FOSD reversal since the defining inequality holds with equality for all increasing f .

In Section 1.G, we provide examples of compound reversals in these models, proving Proposition 6. The key step is to construct a menu $\{X, Y, Z\}$, where Y dominates $X \frac{1}{2} Z$ in the concave order, and the most preferred alternative shifts from Y to X to Z as the risk coefficient increases. Let A have random parameters supported on the lower region of risk coefficients, where Y is the most preferred alternative. Moreover, let the random parameters of B be supported on the higher range of coefficients, where the most preferred alternatives are X and Z , assigning equal and strictly positive probability mass to the regions where each is optimal. The resulting compound lottery from A 's choice distribution then dominates that of B in the concave order, yielding a compound reversal.

²⁷Indeed, Apesteguia and Ballester (2018) point out that random parameter models have limitations when applied to stochastic-dominance related lotteries. They suggest adding an additional trembling stage as a way to generate stochasticity such pairs.

Interestingly, random parameter models based on sum-ex and sum-power utilities—and more generally, utilities of the form in Proposition 5—do not yield compound reversals. This follows from the fact that, for these families, the preference between any two lotteries can reverse at most once as the risk coefficient increases. Thus, if the most preferred alternative shifts from Y to X to Z as in the previous example, then Y must eventually become the least preferred alternative among the three for high risk coefficients. It follows that the corresponding utility functions with high coefficients also evaluate mixtures of X and Z over Y . Since these utility functions are concave, we conclude that Y does not dominate any mixture of X and Z in the concave order, precluding the possibility of compound reversal. This argument generalizes to menus of arbitrary size.

1.7 Conclusion

This paper establishes that the paradox long noted for CARA and CRRA utilities under homoskedastic Fechnerian noise is not an artifact of restrictive parametric assumptions, but instead reflects a deeper incompatibility between these utility forms and noisy choice. We show that the same non-monotone comparative statics arise under highly flexible noise specifications—including heterogeneous noise across individuals and menu-dependent noise.

We obtain a simple characterization of whether two utility functions can yield sensible predictions in the presence of noise. This condition rules out even the more general expo-power utilities, and points to new parametric families that restore intuitive comparative statics. These families, which include sum-ex and sum-power utilities, provide empirically tractable, well-behaved alternatives for measuring risk preferences in noisy environments. We leave to future work an empirical examination of how risk aversion and decision noise co-vary across different stake levels, as well as systematic tests of these proposed utility families under a range of noise specifications.

We conclude by noting that the continuity requirement on noise assignments may be restrictive in some settings. For example, discontinuous noise assignments may capture interesting features of binary comparisons, such as salience or complexity. Understanding which discontinuous models generate reversals thus remains an interesting direction for future research.

1.A Proof of Proposition 1

We start by proving Lemma 1, which shows that

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)}$$

is unbounded over the set of pairs $S >_c R$.

Proof of Lemma 1. Let $S >_c R$ and $a, b > 0$. For CARA, we have

$$\begin{aligned} \frac{\text{CARA}_b(S+x) - \text{CARA}_b(R+x)}{\text{CARA}_a(S+x) - \text{CARA}_a(R+x)} &= \frac{a \mathbb{E}[e^{-b(R+x)}] - \mathbb{E}[e^{-b(S+x)}]}{b \mathbb{E}[e^{-a(R+x)}] - \mathbb{E}[e^{-a(S+x)}]} \\ &= \frac{a}{b} \left(\frac{\mathbb{E}[e^{-bR}] - \mathbb{E}[e^{-bS}]}{\mathbb{E}[e^{-aR}] - \mathbb{E}[e^{-aS}]} \right) e^{(a-b)x}. \end{aligned}$$

For CRRA, when a and b are not equal to 1, we have

$$\begin{aligned} \frac{\text{CRRA}_b(k \cdot S) - \text{CRRA}_b(k \cdot R)}{\text{CRRA}_a(k \cdot S) - \text{CRRA}_a(k \cdot R)} &= \frac{1-a}{1-b} \left(\frac{\mathbb{E}[S^{1-b}] - \mathbb{E}[R^{1-b}]}{\mathbb{E}[S^{1-a}] - \mathbb{E}[R^{1-a}]} \right) \frac{k^{1-b}}{k^{1-a}} \\ &= \frac{1-a}{1-b} \left(\frac{\mathbb{E}[S^{1-b}] - \mathbb{E}[R^{1-b}]}{\mathbb{E}[S^{1-a}] - \mathbb{E}[R^{1-a}]} \right) k^{a-b}. \end{aligned}$$

Since positive risk coefficients a and b for CARA/CRRA utilities correspond to risk-averse utilities,

$$C_1 = \frac{a}{b} \left(\frac{\mathbb{E}[e^{-bR}] - \mathbb{E}[e^{-bS}]}{\mathbb{E}[e^{-aR}] - \mathbb{E}[e^{-aS}]} \right), \quad C_2 = \frac{1-a}{1-b} \left(\frac{\mathbb{E}[S^{1-b}] - \mathbb{E}[R^{1-b}]}{\mathbb{E}[S^{1-a}] - \mathbb{E}[R^{1-a}]} \right) \quad (1.14)$$

are positive. For a or b equal to 1, the expression for C_2 is given by the corresponding limit as a or b tends to 1. $\square$

Lemma 1 provides sequences of lottery pairs (S_n, R_n) along which the ratio of expected-utility differences $(U_B(S_n) - U_B(R_n))/(U_A(S_n) - U_A(R_n))$ diverges, for CARA and CRRA utilities. Lemma 2 establishes that when such a sequence exists there will be reversals. An important subtlety of Lemma 2 is that the reversal-generating lotteries may not be part of this sequence. Indeed, Lemma 2 relies on the following lemma, which establishes that under the additional condition that utility differences tend to zero, there will be lotteries in the sequence that generate reversals.

Lemma 3. *Let $F_A, F_B \in \mathcal{F}$ and let U_A and U_B be risk-averse vNM utility functions with U_A more risk averse. There exist $M \in \mathbb{R}$ and $\delta > 0$ such that all lotteries*

$S >_c R$ satisfying

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)} \geq M \quad \text{and} \quad U_i(S) - U_i(R) < \delta$$

for $i = A, B$ generate a reversal.

Proof of Lemma 3. Let $f_A(t) = \frac{d}{dt}F_A(t)$ and $f_B(t) = \frac{d}{dt}F_B(t)$. Since F_A is continuously differentiable, f_A is continuous at 0 so that for each $\varepsilon_A > f_A(0)$ there is $\delta_A > 0$ such that for each $t \in (0, \delta_A)$, $f_A(t) < \varepsilon_A$ and $F_A(t) < \frac{1}{2} + t\varepsilon_A$. Since F_B is continuously differentiable and $f_B > 0$, for each $0 < \varepsilon_B < f_B(0)$, there exists $\delta_B > 0$ such that for any $t \in (0, \delta_B)$, $f_B(t) > \varepsilon_B$ and $F_B(t) > \frac{1}{2} + t\varepsilon_B$. Let $\delta = \min\{\delta_A, \delta_B\}$ and let $M = \varepsilon_A/\varepsilon_B$.

Let S and R as in the statement of the lemma. Then

$$\begin{aligned} F_B(U_B(S) - U_B(R)) &> \frac{1}{2} + (U_B(S) - U_B(R))\varepsilon_B \\ &\geq \frac{1}{2} + (U_A(S) - U_A(R))\varepsilon_A \\ &> F_A(U_A(S) - U_A(R)). \end{aligned}$$

□

In light of Lemma 3, Lemma 2 follows from the observation that for any lotteries $S >_c R$ and $\varepsilon > 0$, there exist lotteries $S' >_c R'$ with the same utility difference ratio such that the each utility difference is less than ε .

Proof of Lemma 2. Let $F_A, F_B \in \mathcal{F}$ and let U_A and U_B be risk-averse vNM utility functions with U_A more risk averse. By Lemma 3, there exist $M \in \mathbb{R}$ and $\delta > 0$ such that all lotteries $S >_c R$ satisfying

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)} \geq M \quad \text{and} \quad U_i(S) - U_i(R) < \delta$$

for $i = A, B$ generate a reversal. By hypothesis, there exist lotteries $S >_c R$ satisfying the first inequality.

For $\lambda \in (0, 1)$, let $R_\lambda S$ denote a lottery distributed as a compound lottery that yields R with probability λ and yields S with probability $1 - \lambda$.

Then $S >_c R_\lambda S$ and

$$\frac{U_B(S) - U_B(R_\lambda S)}{U_A(S) - U_A(R_\lambda S)} = \frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)},$$

by the linearity of expected utility, i.e., $U(R_\lambda S) = \lambda U(R) + (1 - \lambda)U(S)$ for any vNM utility U . Moreover, as λ tends to zero, $U_i(S) - U_i(R_\lambda S)$ for $i = A, B$ tend to zero. Hence, S and $R_\lambda S$ generate a reversal for all λ small enough.

□

1.B Proof of Proposition 3

Proposition 3 shows how to construct the reversal-generating lotteries of Proposition 1 by generalizing Proposition 2 to arbitrarily different noise structures. The CARA case relies on the following lemma about diminishing utility differences as background wealth increases.

Lemma 4. *Let $S \geq_c R$. Then*

$$\lim_{x \rightarrow \infty} \text{CARA}_a(S + x) - \text{CARA}_a(R + x) = 0$$

for all $a > 0$.

Proof. Since S and R have equal means, we suppose, without loss of generality, that $\mathbb{E}[S] = \mathbb{E}[R] = 0$. Recall that $\text{cara}_a(x) = \frac{1 - e^{-ax}}{a}$ denotes the CARA Bernoulli utility under coefficient a , and let m denote the essential infimum of S .²⁸ By monotonicity, $\text{cara}_a(x) - \text{cara}_a(m + x) \geq \text{cara}_a(x) - \text{CARA}_a(S + x)$ and by concavity of u_a , since $\mathbb{E}[S] = 0$, $\text{cara}_a(x) - \text{CARA}_a(S + x) \geq 0$. Moreover, since $m \leq 0$ and cara_a is concave,

$$\frac{d}{dx} \text{cara}_a(m + x) \cdot |m| \geq \text{cara}_a(x) - \text{cara}_a(m + x).$$

Note that $\frac{d}{dx} \text{cara}_a(m + x) = e^{-a(m+x)}$. Thus

$$e^{-a(m+x)} \cdot |m| \geq \text{cara}_a(x) - \text{cara}_a(m + x) \geq \text{cara}_a(x) - \text{CARA}_a(S + x) \geq 0.$$

Taking the limit as $x \rightarrow \infty$, we see that $\text{cara}_a(x) - \text{CARA}_a(S + x) \rightarrow 0$. Since S was an arbitrary mean-zero lottery and $\mathbb{E}[R] = 0$, we have

$$\lim_{x \rightarrow \infty} \text{cara}_a(x) - \text{CARA}_a(R + x) = 0$$

as well, concluding the proof. □

Proof of Proposition 3. For the case of CARA utilities, Lemma 1 establishes that for any pair of lotteries $S >_c R$, the ratio

$$\frac{U_B(S + x) - U_B(R + x)}{U_A(S + x) - U_A(R + x)}$$

²⁸This is the largest value that S exceeds with probability 1.

tends to infinity with x , while Lemma 4 establishes that the numerator and denominator go to zero. Hence, the result follows from Lemma 3.

For the CRRA case, Lemma 1 establishes that for any pair of lotteries $S >_c R$, the ratio

$$\frac{U_B(k \cdot S) - U_B(k \cdot R)}{U_A(k \cdot S) - U_A(k \cdot R)}$$

tends to infinity with k . As in the proof of Lemma 2,

$$\frac{U_B(k \cdot S) - U_B(k \cdot R_\lambda S)}{U_A(k \cdot S) - U_A(k \cdot R_\lambda S)} = \frac{U_B(k \cdot S) - U_B(k \cdot R)}{U_A(k \cdot S) - U_A(k \cdot R)},$$

and as λ tends to zero, $U_i(k \cdot S) - U_i(k \cdot R_\lambda S)$ tends to zero for $i = A, B$, and the result follows from Lemma 3. $\square$

1.C Proof of Theorem 1

Theorem 1 provides a characterization of vNM utilities that lead to consistent predictions under some noise structures. The proof of Theorem 1 makes use of Proposition 4 along with the following two lemmas.

Recall that in a WEU model (U, F, σ) , the probability of choosing X over Y is given by

$$F\left(\frac{U(X) - U(Y)}{\sigma(X, Y)}\right).$$

Let $f(t) = \frac{d}{dt}F(t)$. Note that f is continuous since $F \in \mathcal{F}$. In the following lemma, we establish compact convergence of $(\frac{1}{\sigma_n}f(\frac{t}{\sigma_n}))_n$ when $\sigma_n \rightarrow \sigma$. For notational convenience, we let

$$f_\sigma(t) = \frac{1}{\sigma} f\left(\frac{t}{\sigma}\right).$$

Lemma 5. *If $\sigma_n \rightarrow \sigma$, then $(f_{\sigma_n})_n$ converges compactly to f_σ .*

Proof of Lemma 5. Let $K \subset \mathbb{R}$ be compact.

$$\begin{aligned} & \sup_{t \in K} |f_{\sigma_n}(t) - f_\sigma(t)| \\ &= \sup_{t \in K} \left| \frac{1}{\sigma_n} f\left(\frac{t}{\sigma_n}\right) - \frac{1}{\sigma} f\left(\frac{t}{\sigma}\right) \right| \\ &= \sup_{t \in K} \left| \frac{1}{\sigma_n} f\left(\frac{t}{\sigma_n}\right) - \frac{1}{\sigma} f\left(\frac{t}{\sigma_n}\right) + \frac{1}{\sigma} f\left(\frac{t}{\sigma_n}\right) - \frac{1}{\sigma} f\left(\frac{t}{\sigma}\right) \right| \\ &\leq \left| \frac{1}{\sigma_n} - \frac{1}{\sigma} \right| \sup_{t \in K} \left| f\left(\frac{t}{\sigma_n}\right) \right| + \frac{1}{\sigma} \sup_{t \in K} \left| f\left(\frac{t}{\sigma_n}\right) - f\left(\frac{t}{\sigma}\right) \right| \end{aligned}$$

Note that since $\sigma_n \rightarrow \sigma$, there is a compact set S such that $\frac{t}{\sigma_n} \in S$ for all n high enough and $t \in K$. Since f is bounded on S , the first term tends to 0, and since $\frac{t}{\sigma_n}$ converges to $\frac{t}{\sigma}$ uniformly on K , and f is uniformly continuous on S , the second term tends to 0 as well. $\square$

Lemma 6. *Let $F_A, F_B \in \mathcal{F}$ and σ_A and σ_B be continuous. Let U_A and U_B be vNM utility functions with U_A more risk averse. Suppose that for each $M \in \mathbb{R}$ there exist lotteries $S >_c R$ such that*

$$\frac{U_B(S) - U_B(R)}{U_A(S) - U_A(R)} \geq M.$$

Then there are reversals for (U_A, F_A, σ_A) and (U_B, F_B, σ_B) .

Proof of Lemma 6. Let U_A and U_B be as in the statement of the lemma. Let $Z, Z', \Lambda \in \mathcal{L}$ such that Z and Z' have the same distribution, and Λ is uniformly distributed on $[0, 1]$ and independent of Z and Z' . Let $F^A(t) = F_A\left(\frac{t}{\sigma_A(Z, Z')}\right)$ and $F^B(t) = F_B\left(\frac{t}{\sigma_B(Z, Z')}\right)$ and let $f^A(t) = \frac{d}{dt}F^A(t)$ and $f^B(t) = \frac{d}{dt}F^B(t)$. Let $\varepsilon_A > f^A(0)$ and $0 < \varepsilon_B < f^B(0)$, which is possible since $f^B(0)$ is positive. Since f^A and f^B are continuous at 0, there is $\delta > 0$ such that for all $0 < t < \delta$, $f^A(t) < \varepsilon_A$ and $f^B(t) > \varepsilon_B$.

By assumption, there are lotteries $X >_c Y$ such that

$$\frac{U_B(X) - U_B(Y)}{U_A(X) - U_A(Y)} \geq \frac{\varepsilon_A}{\varepsilon_B}.$$

Since the inequality only depends on the distributions of X and Y , we may choose X and Y to be independent of Λ . We define the random variables S_λ and R_λ by

$$S_\lambda(\omega) = \begin{cases} X(\omega) & \Lambda(\omega) \leq \lambda \\ Z(\omega) & \Lambda(\omega) > \lambda \end{cases}$$

and

$$R_\lambda(\omega) = \begin{cases} Y(\omega) & \Lambda(\omega) \leq \lambda \\ Z'(\omega) & \Lambda(\omega) > \lambda. \end{cases}$$

Note that for $\lambda \in (0, 1)$, $S_\lambda >_c R_\lambda$ and

$$U_A(S_\lambda) - U_A(R_\lambda) = \lambda(U_A(X) - U_A(Y)),$$

since $U_A(Z) = U_A(Z')$. Hence,

$$\frac{U_B(S_\lambda) - U_B(R_\lambda)}{U_A(S_\lambda) - U_A(R_\lambda)} = \frac{U_B(X) - U_B(Y)}{U_A(X) - U_A(Y)} \geq \frac{\varepsilon_A}{\varepsilon_B}, \quad (1.15)$$

and $U_A(S_\lambda) - U_A(R_\lambda), U_B(S_\lambda) - U_B(R_\lambda) > 0$.

Let $F_\lambda^A(t) = F_A\left(\frac{t}{\sigma_A(S_\lambda, R_\lambda)}\right)$, $F_\lambda^B(t) = F_B\left(\frac{t}{\sigma_B(S_\lambda, R_\lambda)}\right)$ and let $f_\lambda^A(t) = \frac{d}{dt}F_\lambda^A(t)$ and $f_\lambda^B(t) = \frac{d}{dt}F_\lambda^B(t)$. Since $S_\lambda \rightarrow Z$ and $R_\lambda \rightarrow Z'$ as $\lambda \rightarrow 0$, continuity of σ_A and σ_B , i.e., $\sigma_A(S_\lambda, R_\lambda) \rightarrow \sigma_A(S, R)$ and $\sigma_B(S_\lambda, R_\lambda) \rightarrow \sigma_B(S, R)$, imply that $f_\lambda^A \rightarrow f^A$ and $f_\lambda^B \rightarrow f^B$ compactly (Lemma 5). Thus, there is $\delta' > 0$ and λ_A such that for all $\lambda \in (0, \lambda_A)$ and all $t \in [0, \delta']$, $f_\lambda^A(t) < \varepsilon_A$ and $F_\lambda^A(t) < \frac{1}{2} + t\varepsilon_A$. Likewise, there is λ_B such that for all $\lambda \in (0, \lambda_B)$ and all $t \in [0, \delta']$, $F_\lambda^B(t) > \frac{1}{2} + t\varepsilon_B$.

For λ small enough,

$$U_A(S_\lambda) - U_A(R_\lambda) < \delta' \quad \text{and} \quad U_B(S_\lambda) - U_B(R_\lambda) < \delta'.$$

Thus, for λ small enough, we have

$$\begin{aligned} F_\lambda^B(U_B(S_\lambda) - U_B(R_\lambda)) &> \frac{1}{2} + (U_B(S_\lambda) - U_B(R_\lambda))\varepsilon_B \\ &\geq \frac{1}{2} + (U_A(S_\lambda) - U_A(R_\lambda))\varepsilon_A \\ &> F_\lambda^A(U_A(S_\lambda) - U_A(R_\lambda)). \end{aligned}$$

The second inequality follows from (1.15). □

Proof of Theorem 1. Let U_A be more risk averse than U_B . In Section 1.4, we showed that if $U_A \geq_{\text{rel}} U_B$ then there are continuous (indeed, constant) σ_A and σ_B and $F \in \mathcal{F}$ such that there are no reversals under (U_A, F, σ_A) and (U_B, F, σ_B) . We now show that when there will always be reversals if U_A is not relatively more concave than U_B . Indeed, suppose that, for all $k > 0$, it is not the case that $kU_A - U_B$ is concave, i.e., kU_A is not absolutely more concave than U_B . Then, by Proposition 4, for each $k > 0$, there are lotteries $S >_c R$ such that

$$k(U_A(S) - U_A(R)) < U_B(S) - U_B(R).$$

Hence, the result follows from Lemma 6. □

1.D Comparison with Standard Comparative Risk Notions

Note that U_A is more risk-averse than U_B in the Arrow-Pratt sense if and only if, for all x ,

$$\frac{u_B''(x)}{u_A''(x)} \leq \frac{u_B'(x)}{u_A'(x)}.$$

While stochastic comparative risk depends on bounding the ratio of second derivatives by a constant, the traditional Arrow-Pratt notion of comparative risk depends on bounding this ratio by the ratio of marginal utilities. Thus, when $u_B'(x)/u_A'(x)$ is bounded, the traditional notion of comparative risk is sufficient for the stochastic notion. On the other hand, Ross' (1981) stronger comparative notion requires

$$\frac{u_B''(x)}{u_A''(x)} \leq k \leq \frac{u_B'(x)}{u_A'(x)}, \quad (1.16)$$

for some $k > 0$ and all x , implying the boundedness of u_B''/u_A'' , i.e., $u_A \geq_{\text{rel}} u_B$. Moreover, Ross (1981) shows (1.16) is equivalent to the existence of k such that $ku_A - u_B$ is concave and decreasing, while $u_A \geq_{\text{rel}} u_B$ does not require $ku_A - u_B$ to be decreasing.

The next proposition states that, in the presence of noise related by a scale factor, Ross' (1981) notion captures stochastic comparative risk aversion stemming from preferences rather than greater choice precision. To formalize this, we introduce the following notion of relative choice precision.

We say that A has *greater choice precision* than B if, whenever X first-order dominates Y , A chooses X more than B .

Proposition 7. *Let $U_A = \mathbb{E}[u_A]$ and $U_B = \mathbb{E}[u_B]$ be vNM utilities with U_A more risk averse than U_B . Define $F_A(t) = F(t/\sigma_A)$ and $F_B(t) = F(t/\sigma_B)$ for some $\sigma_A, \sigma_B > 0$ and $F \in \mathcal{F}$. Then there are no reversals for (U_A, F_A) and (U_B, F_B) , and B has greater choice precision if and only if $u_A/\sigma_A - u_B/\sigma_B$ is decreasing and concave.*

Proof of Proposition 7. For any $X, Y \in \mathcal{L}$, A chooses X over Y more than B if and only if

$$U_A(X)/\sigma_A - U_B(X)/\sigma_B \geq U_A(Y)/\sigma_A - U_B(Y)/\sigma_B. \quad (1.17)$$

Thus, if $u_A/\sigma_A - u_B/\sigma_B$ is decreasing and concave, then A chooses X less often when X first-order dominates Y and more often when X concave-order dominates Y .

Conversely, suppose that there are no reversals and B has greater choice precision. It follows from Proposition 4 that $u_A/\sigma_A - u_B/\sigma_B$ is concave. Moreover, for any $x > y$, a degenerate lottery X paying x first-order dominates Y paying y . Since B has greater choice precision, the right-hand side of (1.17) is larger, and we conclude that $u_A/\sigma_A - u_B/\sigma_B$ is decreasing.

□

1.E Expo-power Utilities

We illustrate the expo-power utility function, a two-parameter family proposed by Saha (1993) to capture increasing/decreasing absolute/relative risk aversion,²⁹ suffers from the same problems as CARA and CRRA utilities. The increasing and concave expo-power utility function is given by

$$u_{a,r}(x) = \frac{1 - e^{-ax^{1-r}}}{a},$$

for $x \geq 0$ and positive a and $0 \leq r < 1$. Note that the absolute risk aversion is given by

$$A_{a,r}(x) = \frac{r}{x} + a(1-r)x^{-r},$$

for $x > 0$, which is decreasing, constant, or increasing depending on the parameters. Thus, unlike CARA or CRRA utility, the family of expo-power utilities, parameterized by a and r , is not totally ordered by absolute/relative risk aversion.

Importantly, for any pairs of coefficients (a_1, r_1) and (a_2, r_2) such that $A_{a_1, r_1}(x) \geq A_{a_2, r_2}(x)$ so that $U_1 = \mathbb{E}[u_{a_1, r_1}(X)]$ is more risk-averse than $U_2 = \mathbb{E}[u_{a_2, r_2}(X)]$, it must hold that $r_1 = r_2$ and $a_1 > a_2$. A simple calculation shows that when $r_1 = r_2$ and $a_1 > a_2$, $u''_{a_2, r_2}(x)/u''_{a_1, r_1}(x)$ tends to infinity as $x \rightarrow \infty$. Thus, we have the following corollary.

Corollary 3. *Let $F_A, F_B \in \mathcal{F}$ and σ_A, σ_B be continuous. For any distinct expo-power utility functions U_A and U_B there are reversals for (U_A, F_A, σ_A) and (U_B, F_B, σ_B) .*

Corollary 3 highlights that the paradoxical properties arising from CARA (CRRA) utilities are pervasive, and that they do not depend on the strong assumption of constant absolute (relative) risk aversion.

²⁹Notably, Holt and Laury (2002) use the expo-power utility function to model increasing relative risk aversion exhibited in their data.

1.F Sum-ex and Sum-power Utility Functions

In Section 1.5, we demonstrated that for $h > \ell \geq 0$ the utility families

$$u_a(x; h, \ell) = a \cdot \text{cara}_h(x) + \text{cara}_\ell(x) \quad a \geq 0$$

and

$$v_a(x; h, \ell) = a \cdot \text{crra}_h(x) + \text{crra}_\ell(x) \quad a \geq 0,$$

which are parametrized by the coefficient a placed on the more risk-averse cara or crra utility do not lead to reversals under homoskedastic Fechnerian noise. A utility function taking the form $u_a(x; h, \ell)$ is a sum-ex utility and one taking the form $v_a(x; h, \ell)$ is a sum-power utility. In this appendix, we briefly discuss the axioms that characterize the sum-ex and sum-power families as well as an interpretation of the risk parameter a .

One-switch Utilities

Bell (1988) and Bell (1996) showed that sum-ex and sum-power utilities are characterized by *one-switch* properties on preferences over gambles and decreasing absolute risk aversion.

Definition 4. We say that a Bernoulli utility u satisfies the additive one-switch rule if, for all X and Y

$$\mathbb{E}[u(X + t) - u(Y + t)]$$

is single-crossing in t . Likewise, we say that u satisfies the multiplicative one-switch rule if, for all X and Y

$$\mathbb{E}[u(kX) - u(kY)]$$

is single-crossing in $k > 0$.

The family of CARA utilities is similarly characterized by a zero-switch condition, which requires that preferences over gambles are invariant to background wealth. However, we often observe individuals who become less risk averse as their wealth levels increase, exhibiting *decreasing absolute risk aversion* (DARA). Formally, we say that a Bernoulli utility function u exhibits DARA if $-u''(x)/u'(x)$ is decreasing in x .

As argued by Bell (1988), the additive one-switch property is normatively appealing under DARA, since one may start to prefer a riskier gamble once their wealth exceeds some point and their aversion to risk diminishes. However, this preference should

not be reverted at even higher levels of background wealth, at which the individual is even more risk tolerant. Similarly, Bell (1996) argues that the multiplicative one-switch rule is a normatively appealing criterion for an individual comparing two investments with different stochastic returns.

The following proposition is due to Bell (1988) and Bell (1996) and Pedersen and Satchell (2001).

Proposition 8. *Suppose a concave infinitely differentiable Bernoulli utility function u satisfies decreasing absolute risk aversion. Then*

1. *u satisfies the additive one-switch rule if and only if $u(x)$ is a sum-ex utility function;*
2. *u satisfies the multiplicative one-switch rule if and only if $u(x)$ is a sum-power utility function.*

Interpreting the Risk Coefficient

In Section 1.5, we showed that for sum-ex and sum-power utilities

$$u_a(x; h, \ell) = a \cdot \text{cara}_h(x) + \text{cara}_\ell(x), \quad v_a(x; h, \ell) = a \cdot \text{crra}_h(x) + \text{crra}_\ell(x),$$

a higher coefficient a corresponds to greater risk aversion in the Arrow-Pratt sense and an increased likelihood of choosing safer lotteries under identical noise. Thus, we advocated for replacing CARA and CRRA utilities with sum-ex and sum-power utilities in estimating risk attitudes.

In this section, we further provide an interpretation of the risk coefficient a that is to be estimated for these families. Specifically, we show that the coefficient a can be interpreted as a background wealth level. That is, the preference of an individual with a lower coefficient a is identical to that of an individual with a larger coefficient and more background wealth. This equivalence is formalized in the next proposition.

Proposition 9. *Let $a > b > 0$ and $h > \ell > 0$. Then there are $\alpha, \beta, t, k > 0$, $\gamma, \delta \in \mathbb{R}$ such that*

$$u_b(x; h, \ell) = \alpha \cdot u_a(x + t; h, \ell) + \gamma \quad \text{for all } x \in \mathbb{R} \quad (1.18)$$

$$v_b(x; h, \ell) = \beta \cdot v_a(k \cdot x; h, \ell) + \delta \quad \text{for all } x > 0. \quad (1.19)$$

Proof of Proposition 9. For sum-ex utility, let

$$\alpha = \left(\frac{a}{b}\right)^{\frac{\ell}{h-\ell}}, \quad \gamma = \frac{b - \alpha \cdot a}{h} + \frac{1 - \alpha}{\ell}, \quad t = \frac{1}{h - \ell} \ln\left(\frac{a}{b}\right).$$

For sum-power utility, if $\ell \neq 1$ and $h \neq 1$ let

$$\beta = \left(\frac{a}{b}\right)^{\frac{\ell-1}{h-\ell}}, \quad \delta = \frac{a \cdot \beta - b}{1-h} + \frac{\beta - 1}{1-\ell}, \quad k = \beta^{\frac{1}{\ell-1}}.$$

Plugging in these parameters, we obtain (1.18) and (1.19). For h or ℓ equal to 1, taking the limits of the expressions for β , δ , and γ works as well, completing the proof. $\square$

Proposition 9 highlights a benefit that sum-ex and sum-power share with CARA and CRRA in the deterministic setting. One important feature of CARA utility is that background wealth does not affect preferences and thus does not need to be separately estimated. By Proposition 9, the estimation of the coefficient on the higher CARA utility can itself be interpreted as an estimation of background wealth which does not, therefore, need to be separately estimated. Likewise, CRRA is often used as a preference specification over investments where it has the benefit that the level of stakes to be invested do not need to be estimated since they do not affect the preference. While sum-power preferences are not invariant to changes in stakes, the coefficient on the higher CRRA utility can be interpreted as revealing the level of stakes to be invested, which do not then need to be separately estimated.

1.G Multinomial Choice Extensions

Homogeneous Costs

Proposition 10. *Let ρ_A have the APU representation (U_A, c) and ρ_B have the representation and (U_B, c) where U_A and U_B are vNM utilities with U_A more risk averse. The following are equivalent:*

1. *There are no compound reversals;*
2. *There are no FOSD reversals;*
3. $U_A \geq_{\text{abs}} U_B$.

Proposition 10 generalizes Proposition 4 to the broader APU framework by establishing the equivalence between $\geq_{\text{abs}}$ and no reversals under a shared cost function c , extending the result beyond binary choices from concave ordered pairs.

Proof of Proposition 10. Since each FOSD reversal is a compound reversal, (1) implies (2). We now show that (3) implies (1).

Fix a menu M . Since c is strictly convex, ρ_A^M is the unique argmax of

$$\sum_{X \in M} U_A(X)p(X) - c(p(X)),$$

and ρ_B^M is the unique argmax of

$$\sum_{X \in M} U_B(X)p(X) - c(p(X)).$$

If $\rho_A^M = \rho_B^M$, then there is no compound reversal, so suppose they are distinct. We thus have the following two inequalities:

$$\begin{aligned} \sum_{X \in M} U_A(X)\rho_A^M(X) - c(\rho_A^M(X)) &> \sum_{X \in M} U_A(X)\rho_B^M(X) - c(\rho_B^M(X)), \\ \sum_{X \in M} U_B(X)\rho_B^M(X) - c(\rho_B^M(X)) &> \sum_{X \in M} U_B(X)\rho_A^M(X) - c(\rho_A^M(X)). \end{aligned}$$

Summing the inequalities gives

$$\sum_{X \in M} (U_A - U_B)(X)\rho_A^M(X) > \sum_{X \in M} (U_A - U_B)(X)\rho_B^M(X).$$

Since $U_A \geq_{\text{abs}} U_B$ by (3), $U_A - U_B = \mathbb{E}[v]$ for some concave function v , so we do not have a compound reversal.

Finally, we show that (2) implies (3). Contrapositively, suppose that $U_A - U_B = \mathbb{E}[v]$ for some function v that is not concave. Then there are lotteries $S >_c R$ such that $\mathbb{E}[v(S)] < \mathbb{E}[v(R)]$. Let $M = \{S, R\}$.

The maximization of APU yields the first-order conditions:

$$\begin{aligned} U_A(S) - U_A(R) &= c'(\rho_A^M(S)) - c'(1 - \rho_A^M(S)) \\ U_B(S) - U_B(R) &= c'(\rho_B^M(S)) - c'(1 - \rho_B^M(S)). \end{aligned}$$

Let $g(t) = c'(t) - c'(1 - t)$. Since c is strictly convex, c' is strictly increasing, so g is strictly increasing.³⁰ Subtracting the first-order condition for B from that of A yields

$$\mathbb{E}[v(S) - v(R)] = g(\rho_A^M(S)) - g(\rho_B^M(S)).$$

Since $\mathbb{E}[v(S)] < \mathbb{E}[v(R)]$, we conclude that $\rho_A^M(S) < \rho_B^M(S)$, which is an FOSD reversal. $\square$

³⁰Since c is strictly convex, $c(y) < c(x) + c'(y)(y - x)$ and $c(x) < c(y) + c'(x)(x - y)$ for $x < y$, which yield $(c'(y) - c'(x))(y - x) > 0$. Thus c' is strictly increasing.

Heterogeneous Costs

In this section, we show that the APU representation (U, c) for ρ reduces to the FEU model of Section 1.3 on binary menus under a smoothness assumption on the cost function. Recall that a cost function is a function $c: [0, 1] \rightarrow \mathbb{R} \cup \{\infty\}$ that is strictly convex and continuously differentiable on $(0, 1)$ and satisfies $\lim_{p \rightarrow 0} c'(p) = -\infty$. Also, recall that $\mathcal{F}$ is the set of all C_+^1 functions $F: \mathbb{R} \rightarrow [0, 1]$ satisfying $F(t) + F(-t) = 1$.

Lemma 7. *Let ρ have the APU representation (U, c) . If $c \in C_+^2$, then for all binary menus,*

$$\rho(X, Y) = F(U(X) - U(Y))$$

for some $F \in \mathcal{F}$.

Proof of Lemma 7. Fix an APU model (U, c) and define $g(p) = c'(p) - c'(1 - p)$. Suppose that the restriction of c to $(0, 1)$ is a C_+^2 function. Then $g \in C_+^2$. Moreover, since g must satisfy $g(1 - p) = -g(p)$ and $\lim_{p \rightarrow 0} c'(p) = -\infty$, we have $\lim_{p \rightarrow 0} g(p) = -\infty$ and $\lim_{p \rightarrow 1} g(p) = \infty$. Let $F: \mathbb{R} \rightarrow (0, 1)$ be the inverse of g . By the inverse function theorem, $F \in C_+^1$. Moreover, we have $1 - F(t) = F(-t)$, so $F \in \mathcal{F}$.

Finally, we show that the Fechnerian model (U, F) and the APU model (U, c) coincide for binary menus. Fix a binary menu $\{X, Y\}$ and let p be the choice probability of X under the APU model (U, c) . The maximization of the APU model yields the first-order condition:

$$U(X) - U(Y) = c'(p) - c'(1 - p) = g(p).$$

Applying F to the first and last terms, we obtain $F(U(X) - U(Y)) = p$.

□

Proof of Theorem 2. First, we show (3) implies (1). Let $k > 0$ such that $kU_A \geq_{\text{abs}} U_B$ and let c be any cost function. Let ρ_A and ρ_B have the APU representations (kU_A, c) and (U_B, c) , respectively. By Proposition 10, ρ_A and ρ_B do not yield compound reversals. Since $(U_A, \frac{c}{k})$ represents ρ_A and (U_B, c) represents ρ_B , we have (3) implies (1) by setting $c_A = c/k$ and $c_B = c$.

Since each FOSD reversal is a compound reversal, (1) implies (2). We conclude by showing that (2) implies (3). Contrapositively, consider U_A that is not relatively

more concave than U_B . Then by Lemma 7, it follows from Theorem 1 that, for any cost functions $c_A, c_B \in C_+^2$, there exists a binary menu $\{S, R\}$ where $S >_c R$ that generates a reversal, which is an FOSD reversal. $\square$

Compound Reversals for CARA- and CRRA- based RPMs

In Section 1.6, we show that CARA- and CRRA- based RPMs do not yield FOSD reversals. To complete the proof of Proposition 6, we now construct an example where these models yield compound reversals.

Proof of Proposition 6. Let $p = 10^{-5}$. Consider the menu $M = \{X, Y, Z\}$ with the following distributions:

$$X = \begin{cases} 20 & \text{w.p. } 1-p \\ 10 & \text{w.p. } p \end{cases}, \quad Y = \begin{cases} 23 & \text{w.p. } \frac{1-p}{2} \\ 20 & \text{w.p. } \frac{p}{2} \\ 19 & \text{w.p. } \frac{1-p}{2} \\ 10 & \text{w.p. } \frac{p}{2} \end{cases}, \quad Z = \begin{cases} 27 & \text{w.p. } \frac{1-p}{2} \\ 20 & \text{w.p. } p \\ 17 & \text{w.p. } \frac{1-p}{2} \end{cases}.$$

For CARA random parameter models, let ε be uniformly distributed on $[0, 0.5]$. Let A and B have absolute risk coefficients 1.16 and 0.1, respectively. Under the random parameter model with ε added to risk coefficients, we have

$$\rho_A^M(X) = \rho_A^M(Z) > 0 \quad \text{and} \quad \rho_B^M(Y) = 1.$$

Since $Y >_c X \frac{1}{2} Z$, $Y >_c W$ for any W that is a mixture of X, Y, Z with positive equal weights on X and Z . Thus, we have a compound reversal. Similarly, we have a compound reversal of CRRA random parameter models from the same menu, when A and B have relative risk coefficients 18.77 and 2.2, respectively, and ε is uniformly distributed on $[0, 12]$. $\square$

BIBLIOGRAPHY

- Apesteguia, Jose and Miguel A Ballester (2018). “Monotone stochastic choice models: The case of risk and time preferences.” In: *Journal of Political Economy* 126.1, pp. 74–106.
- Barseghyan, Levon, Francesca Molinari, Ted O’Donoghue, and Joshua C Teitelbaum (2018). “Estimating risk preferences in the field.” In: *Journal of Economic Literature* 56.2, pp. 501–564.
- Becker, Gordon M, Morris H DeGroot, and Jacob Marschak (1963). “Stochastic models of choice behavior.” In: *Behavioral Science* 8.1, pp. 41–55.
- Bell, David E. (1988). “One-switch utility functions and a measure of risk.” In: *Management Science* 34.12, pp. 1416–1424.
- (1996). “Measuring Risk and Return for Portfolios.” In: *Wise Choices: Decisions, Games, and Negotiations*. Ed. by Richard J. Zeckhauser, Ralph L. Keeney, and James K. Sebenius. Boston, MA: Harvard Business School Press, pp. 102–118. ISBN: 0-87584-677-7.
- Blavatskyy, Pavlo R (2007). “Stochastic expected utility theory.” In: *Journal of Risk and Uncertainty* 34.3, pp. 259–286.
- (2011). “Probabilistic risk aversion with an arbitrary outcome set.” In: *Economics Letters* 112.1, pp. 34–37.
- Bruner, David M (2017). “Does decision error decrease with risk aversion?” In: *Experimental Economics* 20.1, pp. 259–273.
- Buschena, David and David Zilberman (2000). “Generalized expected utility, heteroscedastic error, and path dependence in risky choice.” In: *Journal of Risk and Uncertainty* 20.1, pp. 67–88.
- Debreu, Gerard (1958). “Stochastic choice and cardinal utility.” In: *Econometrica* 26.3, pp. 440–444.
- Farquhar, Peter H and Yutaka Nakamura (1987). “Constant exchange risk properties.” In: *Operations Research* 35.2, pp. 206–214.
- Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki (2015). “Stochastic choice and revealed perturbed utility.” In: *Econometrica* 83.6, pp. 2371–2409.
- Harless, David W and Colin F Camerer (1994). “The predictive utility of generalized expected utility theories.” In: *Econometrica* 62.6, pp. 1251–1289.
- Harrison, Glenn W., John A. List, and Charles Towe (2007). “Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion.” In: *Econometrica* 75.2, pp. 433–458.

- He, Junnan and Paulo Natenzon (2024). “Moderate utility.” In: *American Economic Review: Insights* 6.2, pp. 176–195.
- Hey, John D (1995). “Experimental investigations of errors in decision making under risk.” In: *European Economic Review* 39.3-4, pp. 633–640.
- Hey, John D and Chris Orme (1994). “Investigating generalizations of expected utility theory using experimental data.” In: *Econometrica* 62.6, pp. 1291–1326.
- Holt, Charles A. and Susan K. Laury (2002). “Risk aversion and incentive effects.” In: *American Economic Review* 92.5, pp. 1644–1655.
- Holzmeister, Felix and Matthias Stefan (2021). “The risk elicitation puzzle revisited: Across-methods (in) consistency?” In: *Experimental Economics* 24.2, pp. 593–616.
- Keffert, Henk and Nikolaus Schweizer (2024). “Stochastic monotonicity and random utility models: the good and the ugly.” In: *arXiv preprint arXiv:2409.00704*.
- Kihlstrom, Richard E., David Romer, and Steve Williams (1981). “Risk aversion with random initial wealth.” In: *Econometrica* 49.4, pp. 911–920.
- Loomes, Graham (2005). “Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data.” In: *Experimental Economics* 8.4, pp. 301–323.
- Loomes, Graham, Peter G Moffatt, and Robert Sugden (2002). “A microeconomic test of alternative stochastic theories of risky choice.” In: *Journal of Risk and Uncertainty* 24.2, pp. 103–130.
- Luce, R. D. and P. Suppes (1965). “Preference, utility and subjective probability.” In: *Handbook of Mathematical Psychology*. Ed. by R. D. Luce, R. R. Bush, and E. Galanter. Vol. III. New York: Wiley, pp. 249–410.
- Meissner, Thomas, Xavier Gassmann, Corinne Faure, and Joachim Schleich (2023). “Individual characteristics associated with risk and time preferences: A multi country representative survey.” In: *Journal of Risk and Uncertainty* 66.1, pp. 77–107.
- O’Donoghue, Ted and Jason Somerville (Sept. 2024). “Investigating Risk Preferences Using Experiments.” In: *Handbook of Experimental Methods in the Social Sciences*. to appear.
- Pedersen, Christian S and Stephen E Satchell (2001). “On the characterisation of investor preferences by changes in wealth.” In: *The Geneva Papers on Risk and Insurance Theory* 26.3, pp. 175–193.
- Rabin, Matthew (2000). “Risk aversion and expected-utility theory: a calibration theorem.” In: *Econometrica* 68.5, pp. 1281–1292.
- Ross, Stephen A. (1981). “Some stronger measures of risk aversion in the small and the large with applications.” In: *Econometrica* 49.3, pp. 621–638.

- Rothschild, Michael and Joseph E. Stiglitz (1970). “Increasing risk: I. A definition.” In: *Journal of Economic Theory* 2.3, pp. 225–243.
- Saha, Atanu (1993). “Expo-power utility: a ‘flexible’ form for absolute and relative risk aversion.” In: *American Journal of Agricultural Economics* 75.4, pp. 905–913.
- Shubbatt, Cassidy and Jeffrey Yang (2024). “Tradeoffs and comparison complexity.” In: *arXiv preprint arXiv:2401.17578*.
- Strzalecki, Tomasz (2025). *Stochastic Choice Theory*. Econometric Society Monographs. Cambridge University Press.
- Tversky, Amos and J Edward Russo (1969). “Substitutability and similarity in binary choices.” In: *Journal of Mathematical Psychology* 6.1, pp. 1–12.
- Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström (2011). “Heterogeneity in risky choice behavior in a broad population.” In: *American Economic Review* 101.2, pp. 664–694.
- Wilcox, Nathaniel T (2008). “Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison.” In: *Risk Aversion in Experiments*. Emerald Group Publishing Limited, pp. 197–292.
- (2011). “‘Stochastically more risk averse:’ A contextual theory of stochastic discrete choice under risk.” In: *Journal of Econometrics* 162.1, pp. 89–104.
 - (2021). “Utility measurement: Some contemporary concerns.” In: *A Modern Guide to Philosophy of Economics*, pp. 14–27.
- Yaari, Menahem E. (1969). “Some remarks on measures of risk aversion and on their uses.” In: *Journal of Economic Theory* 1.3, pp. 315–329.

Chapter 2

MEASURING ECONOMIC PREFERENCES IN THE PRESENCE OF NOISE: THE CONNECTIONS BETWEEN CHOICES AND VALUATIONS

2.1 Introduction

Since the original economic experiment of Thurstone (1931), researchers have been interested in using behavioral measurement to assess individual preferences, evaluate adherence to specific underlying axioms, and develop predictions based on models fit to measurement data. To date, there exists no consensus on how such measurement should be conducted, with different researchers advocating for different methods.

Two principal measurement tools are eliciting binary choices between options and eliciting valuations of options using some index of value. Though in an “idealized environment” these two tools would yield equivalent measurement — and thus identical information on individual preferences — the literature suggests that they do not. Lichtenstein and Slovic’s famous “preference reversals,” popularized in economics by Grether and Plott (1979) and extended in numerous ways, showed systematic deviations between the measurements delivered by choices and valuations. The proportion of subjects choosing a safer over a riskier alternative often exceeds the proportion of subjects whose valuations are higher for the safer rather than the riskier alternative. More recently in experimental economics a literature has emerged reinforcing the point that choices and valuations differ (see, e.g., Freeman and Mayraz, 2019; Freeman, Halevy, and Kneeland, 2019). Researchers have interpreted these failures of “procedural invariance” in several ways: either valuations, choices, or both are understood as failing to credibly measure the preferences of interest;¹ or there do not exist stable economic primitives like preferences in the first place, with preferences constructed contextually in each environment (Tversky and Thaler, 1990; Slovic, 1995; Lichtenstein and Slovic, 2006).

This manuscript unpacks the meaning of the term “idealized environment” in the above paragraph to: a) assess whether said ideal is reasonable to expect; b) provide

¹For example, some researchers relate the putative simplicity of design to credibility of measurement and argue in favor of choices as the “gold standard” for measurement (Freeman, Halevy, and Kneeland, 2019).

a basis for analysis of the relationship between choices and valuations when the environment is realistic rather than idealized; and c) conduct new analytic approaches on existing data to understand if choices and valuations truly differ and provide quantitative assessments of any deviations.

Our exercise begins with two basic observations. First, we note that economic measurement of behavior — like all scientific measurement — is inherently noisy.² There is no a priori reason to expect that different measuring devices are subject to identical noise, and so we do not impose this assumption. Second, we note that individuals are fundamentally heterogeneous in their underlying preferences. Even with the standard assumption of stable primitives (Stigler and Becker, 1977) and no noise in measurement, one would not expect homogeneous choices or valuations.

With these preliminaries in hand, we provide three groups of theoretical results focused on testing for preference stability. First, we show that even if individuals have stable preferences which they attempt to maximize in each type of measurement, standard approaches to examining whether choices and valuations are “calibrated” will generally yield evidence of inconsistency. Even with noise symmetrically distributed about zero for both choices and valuations, it is only when noise is identical that calibration obtains. If noise differs between choices and valuations, choices and valuations may diverge widely under stable, heterogeneous preferences, whether said noise is independent of preferences or not, and whether preference heterogeneity is symmetric or not. We characterize the set of attainable joint measurements for choices and valuations under these cases, and find that they cover wide ranges of deviations between choices and valuations.

Second, with further assumptions governing the heterogeneity of preferences, further restrictions can be made on attainable measurements. We show that if preferences are symmetric and unimodal, then choice and valuation measurements coincide if and only if the choice experiment is conducted at the mean preference. This is true both for aggregate assessment of calibration and, under additional regularity assumptions, conditional assessment given a certain valuation — what we term “conditional calibration.” Thus, it is only in the knife-edge case of having conducted a choice experiment at exactly the mean preference that the standard intuitive approaches to testing for preference reversals are credible. Truly credible tests must rely instead on evaluating the distance between experimental results and the regions of attainable

²This observation is most often attributed to Mosteller and Noguee (1951), though even Thurstone (1931) carries some recognition of noise in measurement.

joint measurements under stable preferences.

Third, imposing functional assumptions on preference heterogeneity and noise permits the analyst to go further still: estimating the distribution of valuations and choices accounting for preference and noise components, providing tests of stable preferences, and estimating the magnitude of inconsistencies. We develop an example of one such technique assuming preference heterogeneity and noise are normal and jointly independent, which leads to a simple probit estimator for evaluating preference stability.

We conduct empirical analyses on the body of prior observations from the preference reversal and choices-vs-valuations literatures (375 experiments in total), as well as the data from two recent papers — McGranaghan et al. (2024a) and McGranaghan et al. (2024b) (424 experiments in total). Across these datasets, approximately 75-80% of observations indicate the classic inconsistency between choices and valuations, with individuals being more likely to choose a safer alternative than their valuations would imply. Importantly, however, very few observations lie outside of the attainable regions for stable preferences once noise and heterogeneity are taken into account, and even fewer can be statistically differentiated from the attainable regions. 95% of the prior literature’s experiments and 100% of the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) experiments are consistent with stable preferences under our theoretical framework without restriction on preference heterogeneity. Even restricting to symmetric preferences, we still find that approximately 80-85% of experiments are consistent with stability.

Our theoretical exercise highlights the special importance of choice experiments conducted at the mean preference; it is these experiments that could actually be used as the basis for the standard tests of calibration and conditional calibration. Actually studying these properties in the prior literature is impossible because prior data do not provide an indication of the location of the mean preference. Moreover, much of the recent literature collects choices and valuations from different subjects.³ Thus, richer analyses motivated by our theoretical development, making use of measures of mean preference or linking choices with valuations cannot be conducted with data from the prior literature. Importantly, however, the data from McGranaghan et al. (2024a) and McGranaghan et al. (2024b) are actually fit for purpose.

³The preference reversal literature generally implemented both choices and valuations on the same subjects. However, this literature largely predates the era of online data repositories so only summary statistics can be recovered from the published papers, and complete analysis is impossible.

We conduct these richer analyses using the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets. Interestingly, we find that for choice experiments conducted close to the mean preference and for individuals whose valuations indicate they should be close to indifferent, choice probabilities are indeed remarkably close to 0.5, consistent with stable preferences. Our structured analysis demonstrates the quantitative implications of this fact under the assumption of joint normality: choices and valuations reveal preferences that are extremely highly correlated and deviate by only around 20 cents in expected value on average.

Interestingly, different types of experiments within the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets reveal different directional deviations between choice and valuation preferences that balance out to approximately zero overall. When deciding between a certain amount and a binary lottery, choices are significantly more risk averse than valuations, with a difference in mean preferences on the order of \$3 in expected value. When deciding between a certain amount and a trinary lottery, choices are significantly less risk averse than valuations, with a difference in mean preferences on the order of -\$3 in expected value. And when deciding between two binary lotteries, choices and valuations recruit preferences that cannot be statistically differentiated.

Our theoretical and empirical development has three critical implications for the central question of measurement of economic preferences and the assessment of preference stability. First, and most importantly *prima facie* inconsistencies across measurement techniques do not necessarily imply a failure of stable preferences. We deliver credible ways to assess preference stability when choice measurement and valuation measurement are both noisy if the analyst is willing to make specific assumptions about the noise and about the heterogeneity of preferences.

Second, we provide guidance for experimental design and analysis of preference stability. A credible design comparing choices to valuations must conduct choices and valuations within subject and select choice experiments in such a way as to cover the mean preference. Then, the researcher can examine the relationship between valuations and choices for choice experiments close to their best estimate of the mean preference, investigate adherence to the attainable regions under stable preferences, and conduct more structured empirical tests like those we propose.

Third, our new statistical facts on failures of preference stability under maintained functional assumptions for preference heterogeneity are in contrast with the existing literature. First, instead of choices reliably recruiting greater risk aversion than valu-

ations (which a cursory look at the data and traditional tests would imply), it appears that overall the two recruit quite similar preferences. Second, instead of choices yielding more risk aversion regardless of the context, results depend on the nature of the problem. If certain amounts are contrasted with binary lotteries, choices are more risk averse, but with trinary lotteries, choices are more risk tolerant. And, if binary lotteries are contrasted with each other, choices show no perceptible difference from valuations. This adds nuance to a literature that generally draws broader conclusions about the robustness and directionality of preference reversals.⁴ Further work with purpose-built experiments would be valuable for assessing whether our nuanced findings on preference instability are robust and investigating their source.

This manuscript proceeds as follows: section 2.2 reviews prior work evaluating procedural invariance and testing for preference stability; section 2.3 presents our theoretical framework for evaluating the relationship between choices and valuations in the presence of preference heterogeneity and noisiness in measurement, developing results related to both calibration and conditional calibration and providing guidance for empirical analysis under maintained assumptions for noise and heterogeneity; section 2.4 presents the applications of these findings to the empirical datasets of the prior literature and McGranaghan et al. (2024a) and McGranaghan et al. (2024b); and section 2.5 provides a discussion drawing out implications for the empirical literature on procedural invariance, and concludes.

2.2 Failures of Procedural Invariance and Tests of Preference Stability

Several canonical research designs have been deployed to study failures of procedural invariance by contrasting choices and valuations.⁵ First, in the preference reversal literature initially developed by Lichtenstein and Slovic (1971) and later popularized in economics by Grether and Plott (1979) the analyst considers two lotteries: a high chance of a moderate prize (called a P -bet), and a low chance of a high prize (called

⁴ One exception is Freeman and Mayraz (2019) who follow the standard testing procedure and focus on one choice between a certain amount and a binary lottery and one choice between two binary lotteries. They conclude that “embedding a pair-wise choice in a choice list increases the fraction of subjects choosing the riskier lottery when the safer alternative is certain but does not significantly affect choices when the safer alternative is risky.” While we reproduce these two patterns when the riskier lottery is binary, we find precisely the opposite of the first effect when the riskier lottery is trinary.

⁵There is also a substantial literature contrasting choices made in isolation to those made in the presence of other choices and a random payment mechanism under the recognition that a non-expected-utility decisionmaker may exhibit isolation failures, i.e., different choices across the two designs (Holt, 1986; Starmer and Sugden, 1991; Cox, Sadiraj, and Schmidt, 2015; Harrison and Swarthout, 2014). Our theoretical results naturally extend to investigating such isolation failures, and we briefly discuss this connection in the conclusion, section 2.5.

a \$-bet). Valuations for both lotteries are elicited alongside a choice between the two. The data set consisting of valuations and choices is analyzed by investigating the shares of individuals who: 1) state a higher valuation for the \$-bet and choose the P -bet (what is considered the “standard reversal”); 2) state a higher valuation for the P -bet and choose the \$-bet (the “non-standard reversal”). It has generally been found that standard reversals exceed non-standard reversals, leading people overall to be more likely to choose the safer alternative than their valuations would imply. This apparent failure of procedural invariance has often been interpreted as a sharp critique of the foundation of utility theory, that individuals have stable preferences to begin with.

Second, in the more recent literature on choices versus valuations, the analyst considers a single lottery. In a first condition, the valuation for the lottery is elicited using a multiple price list design offering subjects a series of binary choices between the lottery and an improving alternative. Then, a single choice from the multiple price list is examined in a separate isolated choice condition comparing the lottery to one of the price list alternatives. The researcher compares the choice probability in the isolated condition to the corresponding choice probability associated with the same choice in the multiple price list condition, i.e., the empirical frequency of exhibiting a valuation for the lottery in excess of the value of the alternative in question. For example, Freeman, Halevy, and Kneeland (2019) elicit the probability equivalent valuation $(\$4, p)$ for the lottery $(\$3, 1)$ and compare the choice probability for $(\$3, 1)$ over $(\$4, 0.8)$ in the isolated choice condition to the proportion of subjects exhibiting a probability equivalent of $p = 0.8$ or higher by choosing $(\$3, 1)$ over $(\$4, 0.8)$ in the relevant row of the multiple price list condition. They find that subjects are systematically more likely to choose the certain $(\$3, 1)$ in the isolated choice condition than in the multiple price list condition, which again indicates that subjects are more likely to choose the safer alternative than their valuations would imply.⁶

These two paradigms for identifying failures of procedural invariance differ in several ways. Preference reversal studies generally make use of two valuations and one choice in a within-subject design, while choice-vs-valuation studies generally make use of one valuation and one choice in a between-subject design. The anal-

⁶Freeman, Halevy, and Kneeland (2019) also consider an identical exercise with the lotteries $(\$3, .5)$ and $(\$4, .4)$ and find similar, though statistically insignificant, directional effects of subjects being more likely to choose the safer alternative than their valuations would imply. The statistical differences between one safe alternative and one risky alternative as opposed to two risky alternatives resonates with our own findings discussed in section 2.4.

ysis also differs in that preference reversal studies often contrast the difference in directional reversals, while choices-vs-valuation studies compare choice and valuation proportions. Ultimately, however, the evaluation of stable preferences in both types of paradigms boils down to comparing measurements across choice-valuation experimental pairs; inference is drawn from the deviation between the two under the maintained assumption that stable preferences imply equal measurements on aggregate. In the next section we consider whether this assumption truly follows from stable preferences if choices and valuations both are noisy measures of their underlying theoretical counterpart. We develop attainable joint combinations of choice-valuation measures and show that equality is, indeed, not predicted generically, in contrast with the maintained assumption of these prior exercises. In developing our theoretical analysis, we propose a framework that is sufficiently general to encompass the differing paradigms for preference reversals and choices-vs-valuation studies.

2.3 Theoretical Considerations

Suppose we want person i to compare a lottery X to a sure payment of amount y . Two ways in which we can ask the person to make this comparison are:

- **Choice task:** We ask person i to make a binary choice between X and y .
- **Valuation task:** We ask person i to state the value y_i that is of equal value to them as X , i.e., such that $y_i \sim_i X$.

A common approach used to analyze experimental data on these two tasks is to compare two quantities: $\mathbb{P}_c(y)$, the empirical choice frequency of y over X and $\mathbb{P}_v(y)$, the empirical proportion of individuals who exhibit a valuation for X that is lower than a given y .⁷

Importantly, in experimental data, the quantities $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ will reflect a combination of choice noise at the individual level and heterogeneity in preferences

⁷Note that comparing $\mathbb{P}_c(y)$ to $\mathbb{P}_v(y)$ is equivalent to examining if a greater share of individuals switch from choosing X to valuing y more highly (i.e., $y > y_i$) compared to the share who switch from choosing y to valuing X more highly (i.e., $y < y_i$). To see this, let a be the share choosing y but valuing X more highly, b be the share choosing y and valuing y more highly, and c be the share choosing X but valuing y more highly. Note that $\mathbb{P}_c(y) = a + b$ and $\mathbb{P}_v(y) = b + c$. Thus, the difference $\mathbb{P}_c(y) - \mathbb{P}_v(y) = a - c$, the difference in shares of incongruous observations. The literature on preference reversals has often documented the latter difference in shares, while the more recent literature comparing choices and valuations provides between-subjects comparison of $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$.

at the aggregate level. Hence, in building a stochastic model to make predictions for these quantities, we need to start with a model of stochastic choice at the individual level (which we do in Section 2.3), and then incorporate heterogeneity in preferences to generate stochastic choice at the aggregate level (which we do in Section 2.3).

Before proceeding, we make two observations about the generality of the forthcoming results. First, in the preference reversal literature, choices and valuations are often collected for a pair (X, Z) of non-degenerate lotteries: Subjects choose between X and Z and provide valuations such as certainty equivalents for both. Second, in the choice-vs-valuations literature at times choices and valuations are also collected for a pair (X, Z) of non-degenerate lotteries: Subjects choose between X and Z and the valuation is measured as a specific lottery Z_i , that is indifferent to X . This measurement of Z_i could be conducted through a stated payment or a stated probability. The quantities $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ have natural counterparts in these settings as do all forthcoming assumptions and results. For the theoretical part of this paper, we therefore discuss the simpler setting where y is a certain payoff and y_i is a subject's valuation for X .

Individual Behavior

We begin with a model of individual-level behavior, and in particular how underlying preferences combine with noise to generate stochastic choices at the individual level. Throughout, we assume:

- For a choice task between X and y , individual i 's underlying preferences involve an indifference value y_{ci}^* such that $y_{ci}^* \sim_i X$. For their actual choice, individual i chooses y over X when y is larger than $y_{ci} \equiv y_{ci}^* + \varepsilon_{ci}$, where choice noise ε_{ci} has CDF F_{ci} . Individual i thus chooses y with probability $\mathbb{P}_{ci}(y) = \mathbb{P}(y_{ci}^* + \varepsilon_{ci} \leq y) = F_{ci}(y - y_{ci}^*)$.
- Likewise, for a valuation task, individual i 's underlying preferences involve an indifference value y_{vi}^* such that $y_{vi}^* \sim_i X$. For their actual valuation, individual i reports valuation $y_{vi} = y_{vi}^* + \varepsilon_{vi}$, where valuation noise ε_{vi} has CDF F_{vi} . Thus, the probability that individual i names a valuation for X that is lower than y is $\mathbb{P}_{vi}(y) = \mathbb{P}(y_{vi}^* + \varepsilon_{vi} \leq y) = F_{vi}(y - y_{vi}^*)$.
- We assume that ε_{ci} and ε_{vi} are mean zero.

The key null hypothesis that we assess in this paper is whether each individual has a stable underlying preference that is used for both the choice tasks and the valuation tasks. We state this formally as Assumption 0:

Assumption 0 (stable preferences). *Each individual i has a stable indifference value; in other words, each individual i has $y_{ci}^* = y_{vi}^* \equiv y_i^*$.*

In this section, we are interested in developing predictions for what combinations $(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y))$ are possible under stable preferences combined with various ancillary assumptions. Specifically, we focus on the following notion of attainability:

Definition 1. *Given a set of assumptions on F_{ci} and F_{vi} , the attainable set $\mathcal{A}_i$ is the set of all $(a, b) \in [0, 1]^2$ such that $\mathbb{P}_{ci}(y) = F_{ci}(y - y_i^*) = a$ and $\mathbb{P}_{vi}(y) = F_{vi}(y - y_i^*) = b$ for some y_i^* and some F_{ci} and F_{vi} consistent with the assumptions.*

In this framework, choices and valuations are both noisy measures of underlying preferences. The individual choice probabilities $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$ depend on a combination of (i) how y compares to the individual's underlying preference y_i^* and (ii) how the person's behavior is impacted by choice noise versus valuation noise. To build intuition, we consider a few examples.

First, note that if the choice noise and the valuation noise are identical, then

$$\mathbb{P}_{ci}(y) = F_{ci}(y - y_i^*) = F_{vi}(y - y_i^*) = \mathbb{P}_{vi}(y).$$

Hence:

Remark 2. If noise is identical for choices and valuations, then $\mathcal{A}_i = \{(a, b) | a = b\}$.

Hence, one way to justify the simple null of $\mathbb{P}_{ci}(y) = \mathbb{P}_{vi}(y)$ is to assume that noise is identical for choices and valuations. However, this seems a very strong assumption without supporting evidence, and indeed for our structural analysis in Section 2.3, if we were to assume stable preferences, then we would conclude that choices are noisier than valuations. Once we permit differential noise for choices versus valuations, the attainable set can expand considerably. Indeed, Example 1 highlights that, if the only restriction on choice and valuation noise is that they both must be mean zero, any pattern is possible for $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$.

Example 1: For any y and any $(a, b) \in (0, 1)^2$, one can find a y_i^* and mean-zero F_{ci} and F_{vi} such that $\mathbb{P}_{ci}(y) = a$ and $\mathbb{P}_{vi}(y) = b$. For instance, given a y , choose a

$y_i^* < y$ and a $\delta > y - y_i^*$, and then assume

$$\varepsilon_{ci} = \begin{cases} \delta & \text{with probability } a \\ \frac{-a\delta}{1-a} & \text{with probability } 1-a \end{cases} \quad \varepsilon_{vi} = \begin{cases} \delta & \text{with probability } b \\ \frac{-b\delta}{1-b} & \text{with probability } 1-b \end{cases}.$$

By construction, a person with this y_i^* and these noise distributions would have $\mathbb{P}_{ci}(y) = a$ and $\mathbb{P}_{vi}(y) = b$.

Example 1 highlights that, if we want to have any hope of rejecting the null of stable preferences, we must impose restrictions on the nature of noise. One common restriction in applications is that noise is symmetric about its mean of zero.

Assumption 1 (symmetric noises). *For each i , the distribution of choice noise and valuation noise are both symmetric, i.e., $F_{ci}(x) = 1 - F_{ci}(-x)$ and $F_{vi}(x) = 1 - F_{vi}(-x)$ for all x .⁸*

Figure 2.1 illustrates possible combinations of $(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y))$ under symmetric noise. The curve in panel A depicts a generic curve for fixed F_{ci} and F_{vi} that are symmetric and continuous with the same convex support. In general, once we fix a F_{ci} and F_{vi} , the set of possible $(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y))$ will be a weakly increasing curve, where different points on the curve correspond to different y_i^* (i.e., a larger y_i^* implies a smaller $y - y_i^*$ and thus both $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$ must be weakly smaller). If F_{ci} and F_{vi} are symmetric, then this curve must pass through $(1/2, 1/2)$ and be symmetric around $(1/2, 1/2)$. Finally, if F_{ci} and F_{vi} are continuous with the same convex support, this curve must be strictly increasing from $(0, 0)$ to $(1, 1)$, as depicted by the curve in panel A.

Note that any weakly increasing curve that passes through $(1/2, 1/2)$ must be contained within the two shaded regions in panel A denoted S_y and S_X . Since these will appear in our results, we provide formal definitions:

$$S_y \equiv [1/2, 1]^2 \quad \text{and} \quad S_X \equiv [0, 1/2]^2.$$

The labels reflect that, with symmetric noises, the outcome will be in S_y when $y \succ_i X$, and the outcome will be in S_X when $X \succ_i y$.

Panel B of Figure 2.1 depicts the case where F_{ci} and F_{vi} are both mean-zero normal distributions with variances σ_{ci}^2 and σ_{vi}^2 . The four curves reflect four different values

⁸For clarity, Assumption 1 is written for continuous distributions; for discrete or combined discrete and continuous distributions, the assumption would be $F_{ci}(x) = G_{ci}(-x)$ and $F_{vi}(x) = G_{vi}(-x)$ for all x , where $G_z(x) \equiv \mathbb{P}(\varepsilon_z \geq x)$.

of $k \equiv \sigma_{ci}^2/\sigma_{vi}^2$. When the ratio of the variances is larger than one, which means that choices are noisier than valuations, within S_y the curve must lie above the 45-degree line in the region labeled C_y , whereas within S_x the curve must lie below the 45-degree line in the region labeled C_x . Intuitively, $\mathbb{P}_{ci}(y)$ will be closer to $1/2$ (which reflects random choice), while $\mathbb{P}_{vi}(y)$ will be closer to 0 or 1 (which reflects implementing one's underlying preference). Analogously, when the ratio of the variances is smaller than one, which means that valuations are noisier than choices, within S_y the curve must lie below the 45-degree line in the region labeled V_y , whereas within S_x the curve must lie above the 45-degree line in the region labeled V_x . Finally, the further is the ratio of variances from one, the more the curve deviates from the 45-degree line.

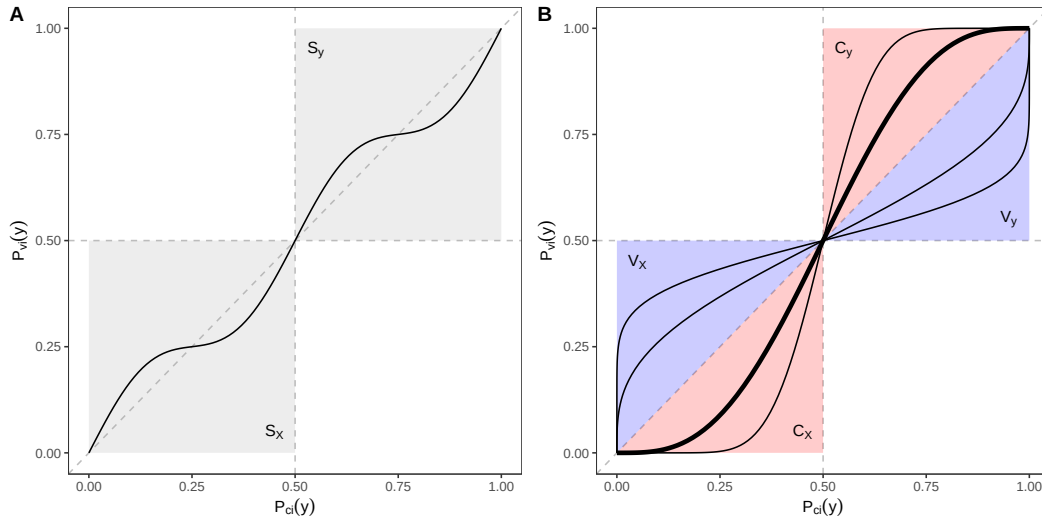


Figure 2.1: Individual Choice and Valuation Probabilities with Varying Noise Assumptions.

Notes: Figure presents relation between individual probabilities, $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$, under varying assumptions for choice and valuation noise distributions, F_{ci} and F_{vi} . Panel A reflects the possible combinations of $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$ when F_{ci} and F_{vi} are symmetric and continuous with convex support yielding Regions S_x or S_y as individual possibilities. Panel B reflects the possible combinations of $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$ when F_{ci} and F_{vi} are both mean-zero normal distributions with $k \equiv \frac{\sigma_{ci}^2}{\sigma_{vi}^2} \in \{4, 2, \frac{1}{2}, \frac{1}{4}\}$. When $k > 1$ (choices noisier than valuations), possible combinations lie in the regions C_x or C_y ; when $k < 1$ (valuations noisier than choices), possible combinations lie in the regions V_x or V_y . Bold line corresponds to $k = 2$.

The curves in Figure 2.1 reflect the attainable set for known noise distributions. When analyzing experimental data, however, one will not know the specific noise distributions, and thus we want to identify the attainable set when we permit all possible noise distributions that satisfy a set of assumptions. From Figure 2.1, one

can see that under symmetric normal-normal noise, if we have the flexibility to choose any σ_{ci}^2 and σ_{vi}^2 , we can produce a curve that goes through any point strictly in the interior of $S_y \cup S_X$. Lemma 1 establishes that under more general symmetric noise, the conclusion is much the same except we can generate outcomes along the edges as well.

Lemma 1. *Under symmetric noises, $\mathcal{A}_i = S_y \cup S_X$.*

The intuition for Lemma 1 is exactly as already expressed above. Whenever $y > y_i^*$ and thus $y \succ_i X$ in terms of underlying preferences, symmetric noise implies $\mathbb{P}_{ci}(y) \geq 1/2$ and $\mathbb{P}_{vi}(y) \geq 1/2$ (because $\mathbb{P}_{ci}(y) = F_{ci}(y - y_i^*) \geq F_{ci}(0) = 1/2$ and $\mathbb{P}_{vi}(y) = F_{vi}(y - y_i^*) \geq F_{vi}(0) = 1/2$). An analogous intuition holds whenever $y < y_i^*$ and thus $y \prec_i X$ in terms of underlying preferences.⁹ In Section 2.B, we construct for each point $(a, b) \in S_y \cup S_X$, a pair of symmetric noises such that $(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y)) = (a, b)$.

The symmetric normal-normal case in Figure 2.1 also illustrates the impact of choices being noisier than valuations versus valuations being noisier than choices. We first formalize the regions in Figure 1:

$$C_y \equiv S_y \cap \{(a, b) \mid b \geq a\} \quad \text{and} \quad C_X \equiv S_X \cap \{(a, b) \mid b \leq a\},$$

and

$$V_y \equiv S_y \cap \{(x, y) \mid y \leq x\} \quad \text{and} \quad V_X \equiv S_X \cap \{(x, y) \mid y \geq x\}.$$

In the normal-normal case, when choices are noisier than valuations (i.e., $\sigma_c^2 > \sigma_v^2$), the outcome must lie in $C_y \cup C_X$, whereas when valuations are noisier than choices (i.e., $\sigma_v^2 > \sigma_c^2$), the outcome must lie in $V_y \cup V_X$. This conclusion generalizes to a more general definition of the noises being ordered (which we introduce later as Assumption 5). However, since one could easily imagine noises not being ordered or being ordered in different ways for different people, and since one cannot easily assess whether the noises are ordered without assuming additional structure, we do not yet make any assumptions on this dimension.

Lemma 1 characterizes the attainable set at the individual level under stable preferences and symmetric noises. If we had many repeated observations of choices and

⁹We note that Assumption 1 is sufficient but not necessary for Lemma 1. All that is required is that F_{ci} and F_{vi} have the same median, in which case $y - y_i^*$ larger than that median must yield an outcome in S_y and $y - y_i^*$ smaller than that median must yield an outcome in S_X .

valuations at the individual level so that we could measure $\mathbb{P}_{ci}(y)$ and $\mathbb{P}_{vi}(y)$, we could assess the null of stable preferences combined with symmetric noises at the individual level. However, in experiments, we virtually always observe aggregate data from a heterogeneous population. Hence, before we can develop tests of the null of stable preferences, we need to build a model of aggregate behavior.

Aggregate Behavior

To specify a model of stochastic choice at the aggregate level, we need to introduce heterogeneity in preferences and heterogeneity in decision noise. To do so, we use the following structure:

- Heterogeneity in preferences: Individual i has $y_i^* = \mu + \eta_i$, where η_i has CDF H and zero mean, so that $\mu \in \mathbb{R}$ is the population mean preference.
- Heterogeneity in decision noise: Individual i has noise distributions $F_{c|\eta_i}$ and $F_{v|\eta_i}$.¹⁰

With this structure, the aggregate choice probability is:

$$\mathbb{P}_c(y) = \mathbb{E}_{\eta_i} [F_{c|\eta_i}(y - \mu - \eta_i)]$$

and

$$\mathbb{P}_v(y) = \mathbb{E}_{\eta_i} [F_{v|\eta_i}(y - \mu - \eta_i)].$$

We are now interested in the attainable set $\mathcal{A}$ for the population proportions $(\mathbb{P}_c(y), \mathbb{P}_v(y))$, which we define analogously to the attainable set $\mathcal{A}_i$ for $(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y))$.

Definition 2. *Given a set of assumptions on H , $F_{c|\eta_i}$, and $F_{v|\eta_i}$, the attainable set $\mathcal{A}$ is the set of all $(a, b) \in [0, 1]^2$ such that $\mathbb{P}_c(y) = a$ and $\mathbb{P}_v(y) = b$ for some H , $F_{c|\eta_i}$, and $F_{v|\eta_i}$ consistent with the assumptions.*

We first note that if we make no assumptions beyond symmetry of noises for all η_i , then the attainable set at the population level is the convex hull of $S_y \cup S_X$.

Proposition 1. *Under stable preferences and symmetric noises, $\mathcal{A} = \text{conv}(S_y \cup S_X)$.*

¹⁰While this specification permits people with different η_i to have different noise distributions, it considers a single noise distribution for everyone with the same η_i . However, by letting $F_{c|\eta_i}(t) = \mathbb{P}(\varepsilon_{ci} < t|\eta_i)$ and $F_{v|\eta_i}(t) = \mathbb{P}(\varepsilon_{vi} < t|\eta_i)$, one can think of these terms as reflecting the average noise distributions among people with the same realized η_i , since in terms of the quantities that one typically observes in an experiment, this would be observationally equivalent.

Panel A in Figure 2.2 depicts the attainable set from Proposition 1. The intuition is straightforward. Lemma 1 establishes that for any point in $S_y \cup S_X$, an individual could exhibit those choice probabilities. If we then put no restrictions on the distribution of preferences and permit people with different y_i^* to have different noise distributions, we can create a population that consists of any distribution we want across points in $S_y \cup S_X$. It immediately follows that we can get average behavior anywhere in $\text{conv}(S_y \cup S_X)$.

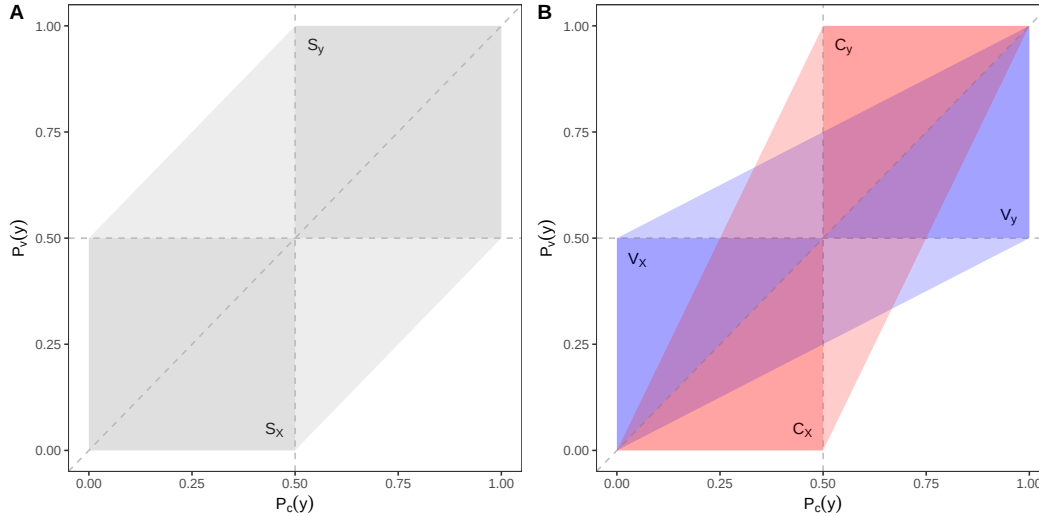


Figure 2.2: Attainable Sets from Propositions 1 and 2.

Notes: Figure presents relation between aggregate probabilities, $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$, under the assumptions of Propositions 1 and 2. Panel A presents the implications of Assumption 1, that choice and valuation noise are both symmetric, yielding an attainable set of $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ combinations $\mathcal{A} = \text{conv}(S_y \cup S_X)$. Panel B presents the implications of Assumption 1 and Assumption 2, that preferences are independent of noise, yielding an attainable set of $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ combinations $\mathcal{A} = \text{conv}(C_y \cup C_X) \cup \text{conv}(V_y \cup V_X)$.

Proposition 1 suggests a possible test of the null of stable preferences: If we want to assume very little—specifically, only that everyone has symmetric noises—then to reject the null of stable preferences we would need to reject that an experiment lies in $\text{conv}(S_y \cup S_X)$. However, that is a pretty large attainable set, and as we will see in Section 2.4, the bulk of existing experiments lie within this set. Hence, one might want to assume more.

A natural next step to restrict the attainable set is to assume that preferences and decision noise are independent from each other:

Assumption 2 (preferences independent of noise). *For all η_i , $F_{c|\eta_i}(s) = F_c(s)$ and $F_{v|\eta_i}(s) = F_v(s)$ for all $s \in \mathbb{R}$.*

An implication of Assumption 2 is that the set of $(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y))$ that could emerge across different y_i^* must be a weakly increasing curve that is symmetric around $(1/2, 1/2)$. Without any restrictions on the distribution of η_i , the attainable set will then be any $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ in the convex hull of this curve. Applying this insight yields Proposition 2.

Proposition 2. *Under stable preferences and symmetric noises, if preferences are independent of noise, then $\mathcal{A} = \text{conv}(C_y \cup C_X) \cup \text{conv}(V_y \cup V_X)$.*

To provide some intuition for Proposition 2, consider the restriction of the attainable set relative to Proposition 1. Specific values of $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ are eliminated; and these values correspond to combinations of correlated types (e.g., combining individuals exhibiting $y_i^* > y$ and valuations noisier than choices with those exhibiting $y_i^* < y$ and choices noisier than valuations). With preferences independent of noise, $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ values attainable by relying on such correlation are no longer permissible.

Proposition 2 suggests another possible test of the null of stable preferences: If we are willing to assume that everyone has symmetric noises and that preferences and noise are independent, then to reject the null of stable preferences we would need to reject that an experiment lies in $\text{conv}(C_y \cup C_X) \cup \text{conv}(V_y \cup V_X)$. This region is illustrated in Figure 2.2. Proposition 2 is proved in Section 2.B.

Propositions 1 and 2 rely on the freedom to use any distribution of preferences that we want. We next consider a natural (and frequently made) restriction on the distribution of preferences:

Assumption 3 (symmetric preferences). *H is symmetric, i.e., $H(x) = 1 - H(-x)$ for all x .*

Under Assumptions 1, 2, and 3, it is natural to rewrite the random variables y_{ci} and y_{vi} as

$$y_{ci} \equiv y_i^* + \varepsilon_{ci} = \mu + \nu_{ci} \quad \text{and} \quad y_{vi} \equiv y_i^* + \varepsilon_{vi} = \mu + \nu_{vi}$$

where $\nu_{ci} = \eta_i + \varepsilon_{ci}$ and $\nu_{vi} = \eta_i + \varepsilon_{vi}$. When preferences are independent of noise (Assumption 2), ν_{ci} and ν_{vi} have the same distributions for all i , and thus we denote their CDFs by F_{ν_c} and F_{ν_v} . We can then write the aggregate choice proportions as:

$$\mathbb{P}_c(y) = F_{\nu_c}(y - \mu) \quad \text{and} \quad \mathbb{P}_v(y) = F_{\nu_v}(y - \mu).$$

Moreover, under symmetric noises (Assumption 1) and symmetric preferences (Assumption 3), v_{ci} and v_{vi} are both symmetric and mean zero. We thus have:

Remark 1: In terms of aggregate proportions $(\mathbb{P}_c(y), \mathbb{P}_v(y))$, a model with symmetric noise and symmetric preferences that are independent of noise is observationally equivalent to a model of an individual with $y_i^* = \mu$ and symmetric mean-zero noise.

Remark 1 immediately leads to Proposition 3:

Proposition 3. *Under stable preferences and symmetric noises, if preferences are symmetric and independent of noise, then:*

1. $\mathcal{A} = S_y \cup S_X$; and
2. If $y > \mu$ then $\mathcal{A} = S_y$, if $y < \mu$ then $\mathcal{A} = S_X$, and if $y = \mu$ then $\mathbb{P}_c(y) = \mathbb{P}_v(y) = \frac{1}{2}$.

Part 1 of Proposition 3 provides another test of the null of stable preferences when comparing aggregate shares in an experiment. If we are willing to make the ancillary assumptions that noise is symmetric and that preferences are symmetric and independent of noise, then we can test the null of stable preferences by assessing whether an experiment lies outside of $S_y \cup S_X$.

Part 2 of Proposition 3 suggests an additional stronger test if the analyst has information on mean preferences—for instance, the sample mean for the reported y_{vi} 's in an experiment. If based on this proxy we are confident that $y > \mu$, then the aggregate shares must be within just S_y , and if instead we are confident that $y < \mu$, then the aggregate shares must be within just S_X . In other words, relative to the simpler test based on part 1, a test based on part 2 could, for instance, reject the null of stable preferences even when the aggregate shares lie in S_y if we also observed that the sample mean for the y_{vi} were clearly larger than y . We illustrate this empirically in Section 2.4.

Part 2 of Proposition 3 provides an initial suggestion that there might be value to move beyond simple comparisons of aggregate shares, and to make use of the connections between choices and valuations. We expand on this idea in the next subsection. But first we provide one final result based on three further assumptions:

Assumption 4 (unimodal preferences). *H has a density h on convex support S , and is strictly unimodal. That is, for any $x_1, x_2 \in S$, $h(x_1) < h(x_2)$ if $x_1 < x_2 \leq 0$ and $h(x_1) > h(x_2)$ if $0 \leq x_1 < x_2$.*

Assumption 5 (ordered noise). *At least one of the following holds:*

1. *Choices are noisier than valuations: $F_v(z) \leq F_c(z)$ for $z \leq 0$ and $F_v(z) \geq F_c(z)$ for $z \geq 0$.*
2. *Valuations are noisier than choices: $F_c(z) \leq F_v(z)$ for $z \leq 0$ and $F_c(z) \geq F_v(z)$ for $z \geq 0$.*

Assumption 6 (distinct noise distributions). *Choice and valuation noise densities satisfy $f_c(x) \neq f_v(x)$ for all x in some open set containing 0.*

The assumption of unimodal preferences is stronger than symmetry, but it is commonly used (e.g., if preferences are normally distributed). The assumption of ordered noise generalizes what we saw for the normal-normal case in Figure 2.1, where σ_c^2/σ_v^2 greater or less than one determined whether choices or valuations are noisier. The third assumption is a technical assumption that the noises are distinct near zero.

Proposition 4. *Under stable preferences and symmetric noises, if preferences are symmetric, unimodal, and independent of noise, and the noise is ordered and distinct, then $\mathbb{P}_c(y) = \mathbb{P}_v(y)$ if and only if $y = \mu$, in which case $\mathbb{P}_c(y) = \mathbb{P}_v(y) = \frac{1}{2}$.*

Proposition 4 highlights a different problem with the typical approach of testing whether the aggregate shares are the same. Up to now, we have emphasized that a simple test of the null $\mathbb{P}_c(y) = \mathbb{P}_v(y)$ is not enough to test the null of stable preferences unless one is willing to make very strong ancillary assumptions (e.g., that the choice noise and the valuation noise are identical). In Proposition 4, we see a second problem: Under the types of regularity assumptions frequently made in economics—symmetric continuous noise and symmetric unimodal preferences—stable preferences would in fact yield $\mathbb{P}_c(y) = \mathbb{P}_v(y)$ only if we happen to be conducting the experiment at $y = \mu$. Since the population mean preference is typically something one would not know when designing an experiment, a strong interpretation is that, in fact, under stable preferences, we should not expect to see $\mathbb{P}_c(y) = \mathbb{P}_v(y)$ in any experiments.

Using Connections Between Valuations and Choices

An advantage of tests based on Propositions 1, 2, and 3(1) is that they do not require linked choices and valuations at the individual level, and thus can be conducted on between-subjects data. If, however, one has linked data at the individual level, additional tests become possible, as suggested in 3(2). In this section, we explore additional ways to use linked data.

Conditional Calibration

A natural conjecture is that one can extend the logic from Proposition 3(2) from the experimental level to the individual level. In particular, while it might be difficult to choose experimental parameters such that y is identical to the population mean μ , we could limit attention to individuals who state a valuation $y_{vi} = y$. If this signals that the experimental y is close to their y_i^* , their choice probability should be close to $\frac{1}{2}$. In other words, because it is true that, under symmetric noise,

$$\mathbb{P}_c[y|y = y_i^*] = \frac{1}{2},$$

and since $E[y_{vi}] = y_i^*$, one might conjecture that

$$\mathbb{P}_c[y|y = y_{vi}] = \frac{1}{2}.$$

We refer to this condition as *conditional calibration*.¹¹

Unfortunately, it turns out that conditional calibration is unlikely to hold even under relatively benign assumptions. In particular, suppose we assume symmetric noises and stable preferences that are symmetric, unimodal, and independent of noise. In addition, we assume two additional nice regularity conditions:

Assumption 7 (independent noise). ε_{ci} and ε_{vi} are independent.

Assumption 8 (positive densities). The densities f_c and f_v are positive in an open interval containing zero.

The following proposition establishes that, even with all these nice regularity conditions, conditional calibration holds only if y is equal to the mean preference.

¹¹Note that while the conditioning event is null, this is formally a regular conditional probability distribution, which is uniquely defined up to a null set.

Proposition 5. *Under stable preferences and symmetric noises, if preferences are symmetric, unimodal, and independent of noise, and the noise is independent with positive densities, then $\mathbb{P}_c[y | y = y_{vi}] = \frac{1}{2}$ if and only if $y = \mu$. Moreover, if $y > \mu$ then $\mathbb{P}_c[y | y = y_{vi}] > \frac{1}{2}$, and if $y < \mu$ then $\mathbb{P}_c[y | y = y_{vi}] < \frac{1}{2}$.*

The intuition for Proposition 5 turns out to be relatively straightforward. Consider an experiment with $y > \mu$. When we observe that a group of people report the valuation $y_{vi} = y$, it need not be the case that $E[y_i^* | y_{vi} = y] = y$; indeed, with symmetric and unimodal preferences, it must be that $E[y_i^* | y_{vi} = y]$ is in between μ and y . Hence, for this group, y will be relatively attractive, and thus it leads to a conditional choice probability greater than $\frac{1}{2}$. An analogous argument holds for an experiment with $y < \mu$, leading to a conditional choice probability less than $\frac{1}{2}$. Proposition 5 is proved in Section 2.B.

Figure 2.3 provides an illustration for the case where H , F_c , and F_v are all normal. Panel A depicts the histogram of y_i^* . In a valuation task, people with different y_i^* will report the same valuation $y_{vi} = y'$ due to noise in valuations. Moreover, since F_v is symmetric, for any $\Delta > 0$, people with $y_i^* = y' + \Delta$ and $y_i^* = y' - \Delta$ are equally likely to report $y_{vi} = y'$. But since H is symmetric and unimodal and since $y' > \mu$, for any Δ the density is larger at $y' - \Delta$ than at $y' + \Delta$, and as a result the conditional mean is dragged from y' toward μ . Panel B illustrates this by depicting the conditional distribution for $y_i^* | y_{vi}=y'$, and $\mathbb{P}(y' | y' = y_{vi})$ is then the shaded area. Finally, panel C depicts $\mathbb{P}(y | y = y_{vi})$ as a function of y , where the outcome from panels A and B is one point in panel C.

We emphasize this result on conditional calibration as a point of caution: it is tempting to think that one way to test for stable preferences is to test whether $\mathbb{P}(y | y_{vi} = y) = 1/2$. Analogous to our discussion after Proposition 4, such a test would be valid only if we were confident that we were conducting an experiment precisely at $y = \mu$. While conditional calibration does not hold generally, Proposition 5 implies that something weaker does hold: congruence between the signs of $y - \mu$ and $\mathbb{P}(y | y_{vi} = y) - 1/2$. In our empirical analysis, we test this implication of Proposition 5.

Structural Analysis

Up to now, our assumptions have involved qualitative features of the distributions of noise and preferences. If one is willing to make structural assumptions about

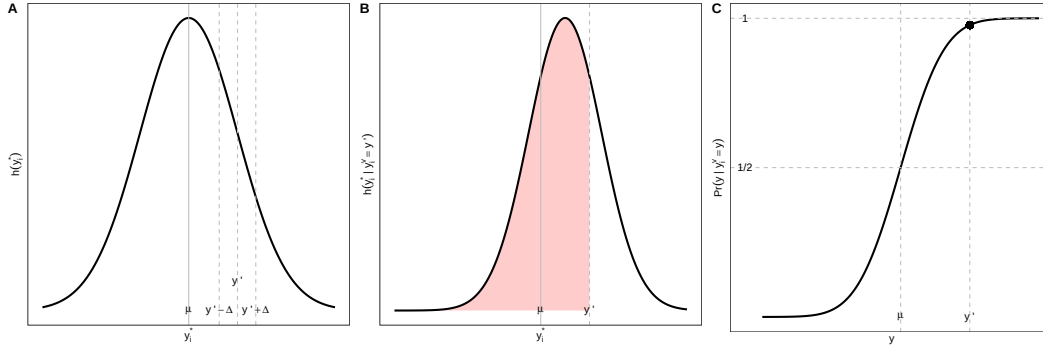


Figure 2.3: Failure of Conditional Calibration from Proposition 5.

Notes: Figure presents illustration of conditional calibration failure under joint normality for H , F_c and F_v . Panel A shows the probability of two individuals with different true preferences stating the same valuation, $y' > \mu$ above the mean preference with the implication that the individuals with true values closer to the mean are more likely under symmetric H . Panel B shows the resulting implication for the conditional distribution $y_{vi}^*|y_{vi}=y'$ with its conditional mean dragged from y' towards μ . Shaded region corresponds to $\mathbb{P}(y'|y' = y_{vi})$. Panel C depicts $\mathbb{P}(y|y = y_{vi})$ as a function of y . Illustrated point corresponds to $\mathbb{P}(y'|y' = y_{vi})$

these distributions, and if one has the appropriate variation in one's data, one can test the null of stable preferences by estimating structural parameters. Moreover, beyond testing the null, a structural approach also permits quantitative estimates of the magnitudes of deviations from the null of stable preferences. In this section, we present an example of such a structural approach in which we assume all distributions are normal.

Suppose that the population distribution of underlying preferences is given by

$$\begin{pmatrix} y_{ci}^* \\ y_{vi}^* \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_c \\ \mu_v \end{pmatrix}, \begin{pmatrix} \gamma_c^2 & \gamma_{cv} \\ \gamma_{cv} & \gamma_v^2 \end{pmatrix} \right).$$

With this structure, the null of stable preferences would mean that $\mu_c = \mu_v$ and $\gamma_c^2 = \gamma_v^2 = \gamma_{cv}$ (which together imply that every individual has $y_{ci}^* = y_{vi}^*$).

We also assume that noise is independent of preferences (Assumption 2), and we further assume that ε_{ci} and ε_{vi} both have mean-zero normal distributions with variances σ_c^2 and σ_v^2 , respectively. Finally, we assume that the two types of noise are independent of each other (Assumption 7).

Recall that a person chooses y over X when y is larger than $y_{ci} \equiv y_{ci}^* + \varepsilon_{ci}$, and the person reports valuation for X of $y_{vi} \equiv y_{vi}^* + \varepsilon_{vi}$. Under the assumptions above, the joint distribution of y_{ci} and y_{vi} is bivariate normal:

$$\begin{pmatrix} y_{ci} \\ y_{vi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_c \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_c^2 + \gamma_c^2 & \gamma_{cv} \\ \gamma_{cv} & \sigma_v^2 + \gamma_v^2 \end{pmatrix} \right). \quad (2.1)$$

If one collects within-subjects responses for both choices and valuations, this framework generates a prediction for $\mathbb{P}(y|y_{vi} = y)$. In particular, because y_{ci} and y_{vi} are joint normal, the conditional distribution of y_{ci} given a specific valuation of y_{vi} is normal, and from that conditional distribution one can derive that

$$\mathbb{P}(y|y_{vi} = y) = \Phi \left(\frac{\mu_v - \mu_c}{V} + \frac{\left(1 - \frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2}\right)}{V} (y - \mu_v) + \frac{\frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2}}{V} (y - y_{vi}) \right)$$

where $V^2 \equiv \sigma_c^2 + \gamma_c^2 - \frac{(\gamma_{cv})^2}{\sigma_v^2 + \gamma_v^2}$. Replacing $y - \mu_v$ with its sample analog from valuations, $y - \bar{y}_{vi}$, yields a simple probit regression:

$$\mathbb{P}(y|y_{vi}) = \Phi(\alpha + \beta_1(y - \bar{y}_{vi}) + \beta_2(y - y_{vi})), \quad (2.2)$$

where $\alpha = \frac{\mu_v - \mu_c}{V}$, $\beta_1 = \frac{\left(1 - \frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2}\right)}{V}$, $\beta_2 = \frac{\frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2}}{V}$, and $\beta_1 + \beta_2 = \frac{1}{V}$. Hence, if one has choice data for multiple experiments (i.e., for multiple values of y) given a fixed X , one can obtain estimates for α , β_1 , and β_2 .¹²

With these estimates in hand, we can test the null of stable preferences, which again means $\mu_v = \mu_c$ and $\gamma_v^2 = \gamma_c^2 = \gamma_{cv}$. Testing $\mu_v = \mu_c$ is equivalent to testing $\alpha = 0$. Within the context of the probit model, testing $\alpha = 0$ asks whether, when projecting to a counterfactual ideal choice experiment conducted precisely with $y - \bar{y}_{vi} = 0$ and $y - y_{vi} = 0$, the choice probability is equal to $\Phi(0) = \frac{1}{2}$. When $\alpha = 0$ is rejected, an estimate for the deviation in mean preferences, $\mu_v - \mu_c$, can be derived from combining regression coefficients,

$$\mu_v - \mu_c = \frac{\alpha}{\beta_1 + \beta_2},$$

with standard error derived from the delta method.

This approach yields only a partial test of $\gamma_v^2 = \gamma_c^2 = \gamma_{cv}$ due to the inability to separate preference heterogeneity from noise. In particular, combining regression coefficients gives us an estimate of

$$\frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2} = \frac{\beta_2}{\beta_1 + \beta_2},$$

where again the standard error can be derived from the delta method. Because σ_v^2 must be positive, stable preferences—specifically, $\gamma_{cv} = \gamma_v^2$ —implies $0 \leq \frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2} \leq$

¹²The key challenge is obtaining enough variation in $y - \bar{y}_{vi}$ to separately identify α and β_1 since $y - \bar{y}_{vi}$ will only vary across experiments.

1. Hence, testing $H_0 : \frac{\beta_2}{\beta_1 + \beta_2} \in [0, 1]$ provides a partial test of stable heterogeneity. With more information, such as repeated valuations like those implemented in Obviously-Related Instrumental Variables (Gillen, Snowberg, and Yariv, 2019) or preference decomposition exercises (McGranaghan et al., 2024a), more progress can be made on separately identifying σ_v^2 and developing a richer test, but this requires more data than one traditionally has in the comparison of choices to valuations.

Whether applied to testing the null hypotheses of stable preferences or to evaluating the magnitude of preference deviations, the probit estimator of equation (2.2) relies on replacing $y - \mu_v$ with its sample analog, $y - \bar{y}_{vi}$. This introduces the possible challenge of measurement error biasing corresponding estimates. Attenuation bias could naturally lead to measures of α closer to zero, working against finding evidence of preference instability. A natural alternative is to estimate mean preferences alongside the other variables of interest by pursuing a full maximum likelihood approach.

Imagine a person provides a valuation in the region (y_L, y_H) (e.g., the interval that corresponds to the switching point in a “multiple price list” eliciting y_{vi}), and that when offered a choice between X and y , they choose y . The likelihood of such an observation is the probability of (y_{ci}, y_{vi}) falling within the square $(-\infty, y) \times (y_L, y_H)$. Under the normal-normal-normal specification here, it is straightforward to calculate this probability given the joint normal distribution in equation (2.1). Maximizing the log likelihood of observations with respect to the parameters $\mu_c, \mu_v, \sigma_c^2 + \gamma_c^2, \sigma_v^2 + \gamma_v^2, \gamma_{cv}$ is a straightforward exercise yielding estimates of the corresponding parameters using conventional techniques. Similar tests of $\mu_v = \mu_c$ and $\frac{\gamma_{cv}}{\sigma_v^2 + \gamma_v^2} \in [0, 1]$ can be implemented using the maximum likelihood estimates.

Again, the overarching theme of the structured empirical approaches motivated by our analysis is that with more assumptions the analyst may be able to make substantial progress on testing preference stability and evaluating the extent of deviations between the primitives recruited for choices and valuations. We illustrate by applying this approach in Section 2.4.

2.4 Empirical Application

Our theoretical analysis develops a suite of propositions and structural methods for the evaluation of stable preferences across choices and valuations. Propositions 1, 2, and 3(1) may be conducted on aggregate values $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ without information

on the location of the mean preference, μ , while Proposition 3(2) requires information on μ , and Proposition 5 requires both information on μ and connected data on conditional choice $\mathbb{P}_c[y|y_{vi}]$.

These differential data requirements mean that Propositions 1, 2, and 3(1) can actually be investigated with the minimal aggregates reported in prior data sets on preference reversals and choices-versus-valuations studies. We identified 375 prior experiments for which we could obtain $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ combining preference reversal studies from Lu (2025) and choices-vs-valuations studies from Freeman, Halevy, and Kneeland (2019) and Freeman and Mayraz (2019) and Brown and Healy (2018). We harmonize the data such that $\mathbb{P}_c(y)$ corresponds to the likelihood of choosing the safer alternative and $\mathbb{P}_v(y)$ corresponds to the likelihood of stating a higher valuation for the safer alternative (in the case of a preference reversal study) or having a valuation that would indicate a preference for the safer alternative (in the case of a choices-vs-valuations study).

Due to data limitations in the prior literature, the richer analyses of Propositions 3(2) and 5, as well as our more structured analyses, cannot be conducted.¹³ In order to be able to study the full suite of empirical exercises motivated by our theoretical development, we turn to two studies: McGranaghan et al. (2024a) and McGranaghan et al. (2024b). These studies collect both valuations and choices within subject, and choice experiments were chosen to bracket what the authors expected to be the mean preference in each case. Thus, these studies can provide measures appropriate for testing all propositions as well as conducting our structural exercises.

The McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets are described in detail in Appendix 2.C. In these experiments subjects make three types of choices: 1) AB , a choice between a safe certain lottery A and a risky binary lottery B ; 2) AB' , a choice between a safe certain lottery A and a risky trinary lottery B' ; and 3) CD , a choice between a safer binary lottery C and a riskier binary lottery D . Each type of choice is paired with a valuation, fixing one of the lotteries, and eliciting the indifferent alternative (e.g., fixing a lottery B and finding the value of the certain moderate payment in lottery A that makes the subject indifferent). Two different methods of valuation are considered, eliciting either an equivalent high prize or an equivalent moderate prize.

¹³ While preference reversals are generally studied within subject, few manuscripts provide information on mean preferences and often report multiple choice and valuation experiments in one quantity. Choices-versus-valuations studies generally conduct between-subjects designs and also do not report information on mean preferences.

The different types of choices and the different methods of valuation generate issues of comparability across experiments. In order to resolve these issues, we organize the data such that higher levels of the choice experimental value y correspond to a more attractive safer alternative, and higher y_{vi} corresponds to a greater preference for the safer alternative. $\mathbb{P}_c(y)$ corresponds to the likelihood of choosing the safer alternative and $\mathbb{P}_v(y)$ corresponds to the likelihood of having a valuation that would indicate a preference for the safer alternative. For each experiment we calculate an average valuation, $\overline{y_{vi}}$, as an estimate of the mean preference, and can thus calculate the distance of the choice experiment to both the mean preference, $y - \overline{y_{vi}}$, and each individual's valuation, $y - y_{vi}$. To make these values comparable across different types of choices and methods of elicitation, we normalize them to be in expected value units; Appendix 2.C also provides complete details on the normalizations of the data.

Across the two papers, McGranaghan et al. (2024a) and McGranaghan et al. (2024b) have 64 different AB experiments, 120 different AB' experiments, and 240 different CD experiments, yielding 424 experiments in total. Each experiment was completed by an average of 161.4 subjects, yielding a total of 68,448 individual observations.

We study Propositions 1, 2, and 3(1) on both the prior studies from the preference reversal and choices-vs-valuations literatures as well as the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets. We study Propositions 3(2) and 5 and conduct our structural analysis using only the latter. Before conducting these analyses, we begin with the traditional inferential approach assessing whether $\mathbb{P}_c(y) = \mathbb{P}_v(y)$.

Traditional Inference Comparing $\mathbb{P}_c(y)$ to $\mathbb{P}_v(y)$

In Figure 2.4 we plot values of $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ with reference to the 45-degree line of equality, where, again, higher values of each indicate a greater preference for the safer alternative. In Panel A we present the data for the prior literature, denoting whether an experiment is sourced from a preference reversal or a choices-vs-valuations study,¹⁴ and in Panel B we present the data from McGranaghan et al. (2024a) and McGranaghan et al. (2024b). Within the prior literature, 278 of 375 experiments (74%) exhibit $\mathbb{P}_c(y) > \mathbb{P}_v(y)$, and within the McGranaghan et al.

¹⁴Some experiments from the preference reversal literature combine results from multiple experimental conditions and so should not, in principle, be studied as if they were isolated choice-valuation experimental pairs. Nonetheless, these aggregated observations may be instructive, and so we present them alongside the rest for completeness.

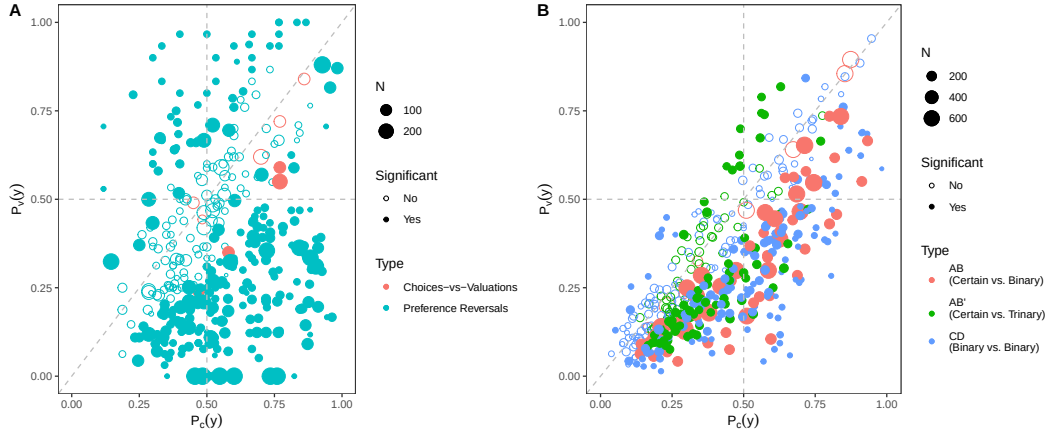


Figure 2.4: Traditional Inferential Approach.

Notes: Figure presents relation between aggregate choice and valuation probabilities, $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$, for 375 experiments from the prior literature (Panel A) and 424 choice-valuation experimental pairs from McGranaghan et al. (2024a) and McGranaghan et al. (2024b) (Panel B). Size corresponds to experimental sample size. Data are harmonized such that greater probability corresponds to greater likelihood of preferring the safer alternative. Solid shading of dots corresponds to observations identified to deviate significantly at the 5% level from $\mathbb{P}_c(y) = \mathbb{P}_v(y)$, the traditional null for stable preferences.

(2024a) and McGranaghan et al. (2024b) data 341 of 424 experiments (80%) exhibit $\mathbb{P}_c(y) > \mathbb{P}_v(y)$. Following conventional wisdom, individuals indeed appear more likely to choose safer alternatives than their valuations would imply.

In order to develop a statistical test for the traditional null of $\mathbb{P}_c(y) = \mathbb{P}_v(y)$, we consider the standard test of proportions under the assumption of independence with z -score

$$z = \frac{\mathbb{P}_v(y) - \mathbb{P}_c(y)}{\sqrt{\frac{\mathbb{P}_v(y)(1-\mathbb{P}_v(y))}{N_v} + \frac{\mathbb{P}_c(y)(1-\mathbb{P}_c(y))}{N_c}}},$$

where N_c and N_v refer to the number of subjects completing choices and valuations, respectively; and we reject the null of $\mathbb{P}_c(y) = \mathbb{P}_v(y)$ if $|z| > 1.645$ (i.e., we impose a one-sided threshold of 0.05). For the prior data, 69% of experiments reject the null hypothesis of $\mathbb{P}_c(y) = \mathbb{P}_v(y)$; and for the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets 61% of experiments reject the null hypothesis of $\mathbb{P}_c(y) = \mathbb{P}_v(y)$.¹⁵

¹⁵Within the paired data of McGranaghan et al. (2024a) and McGranaghan et al. (2024b) we can also evaluate the test of proportions accounting for the correlation between choices and valuations, adjusting the z -score to

$$z_{paired} = \frac{\mathbb{P}_v(y) - \mathbb{P}_c(y)}{\sqrt{\frac{\mathbb{P}_v(y)(1-\mathbb{P}_v(y))}{N} + \frac{\mathbb{P}_c(y)(1-\mathbb{P}_c(y))}{N} - 2\frac{cov_{v,c}}{N}}},$$

This traditional approach of testing $\mathbb{P}_c(y) = \mathbb{P}_v(y)$ reinforces the received wisdom that individuals appear systematically more likely to choose safer alternatives than corresponding valuations would imply in both the prior literature and the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets. An analyst conducting such tests would conclude that choices and valuations often recruit different preferences. Of course, we know that such inference is flawed unless one is willing to assume that choices and valuations are subject to identical noise. Without this strong assumption restricting attainable observations to be those with $\mathbb{P}_c(y) = \mathbb{P}_v(y)$, much more behavior is actually attainable under stable preferences.

Evaluating Propositions 1, 2, and 3(1) Using Only Aggregate Combinations

Propositions 1, 2, and 3(1) elaborate the set of attainable aggregate combinations $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ under stable preferences and varying assumptions for choice and valuation noise and heterogeneity. In Figure 2.5, we present the prior data in Panel A and the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets in Panel B with reference to these various attainable regions.

Notable from Figure 2.5 is the substantial portion of observations that lie within the attainable regions of $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ for all propositions. As Propositions 1, 2, and 3(1) are ordered in their restrictiveness of the attainable regions of $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ under stable preferences, we present the empirical results in reverse order moving from the smallest attainable region in Proposition 3(1) to the largest in Proposition 1.

Proposition 3(1) provides our most restrictive test of stability under the assumption of symmetric preferences and noise with attainable combinations $(\mathbb{P}_c(y), \mathbb{P}_v(y)) \in S_X \cup S_Y$. Within the prior literature, 45% of observations lie outside of $S_X \cup S_Y$; and within the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets 23% of observations lie outside of $S_X \cup S_Y$.

In order to provide a statistical test of Proposition 3(1), we deploy the recently-developed test of Miller, Molinari, and Stoye (2025) for sign congruence between two estimators. Proposition 3(1) requires $\mathbb{P}_c(y) - 0.5$ and $\mathbb{P}_v(y) - 0.5$ to be sign-

where N is the number of paired observations, and

$$cov_{v,c} = \mathbb{P}(y, y) - \sqrt{\mathbb{P}_v(y)(1 - \mathbb{P}_v(y))} * \sqrt{\mathbb{P}_c(y)(1 - \mathbb{P}_c(y))}$$

with $\mathbb{P}(y, y)$ being the empirical proportion of subjects both choosing y over X and exhibiting valuations that imply a preference for y over X . With this adjustment, 67% of McGranaghan et al. (2024a) and McGranaghan et al. (2024b) experiments reject the null hypothesis of $\mathbb{P}_c(y) = \mathbb{P}_v(y)$.

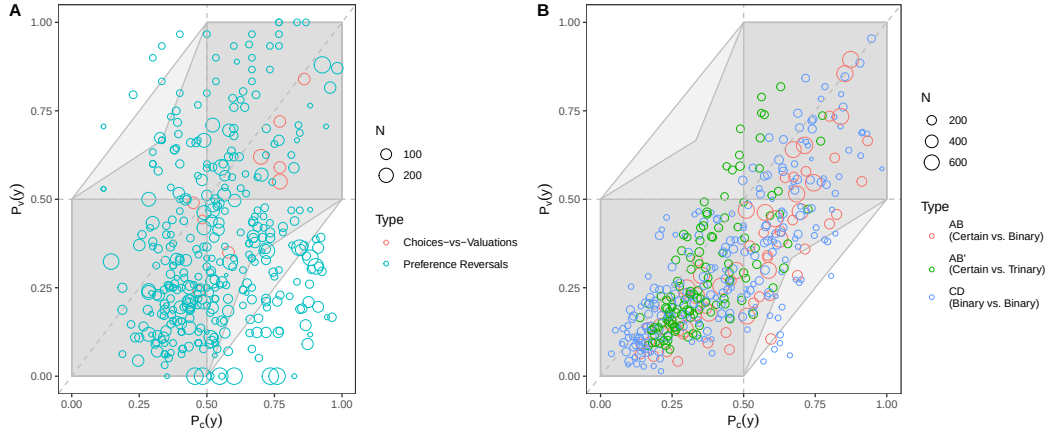


Figure 2.5: Evaluating Propositions 1, 2, and 3(1).

Notes: Figure presents relation between aggregate choice and valuation probabilities, $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$, for 375 experiments from the prior literature (Panel A) and 424 choice-valuation experimental pairs from McGranaghan et al. (2024a) and McGranaghan et al. (2024b) (Panel B). Size corresponds to experimental sample size. Data are harmonized such that greater probability corresponds to greater likelihood of preferring the safer alternative. Gray regions correspond to attainable regions of behavior under stable preferences and assumptions of Propositions 1, 2, and 3(1).

congruent such that $(\mathbb{P}_c(y) - 0.5)(\mathbb{P}_v(y) - 0.5) \geq 0$. The Miller, Molinari, and Stoye (2025) approach considers the two relevant z -scores,

$$z_c = \frac{\mathbb{P}_c(y) - 0.5}{\sqrt{\frac{\mathbb{P}_c(y)(1-\mathbb{P}_c(y))}{N_c}}}, \quad z_v = \frac{\mathbb{P}_v(y) - 0.5}{\sqrt{\frac{\mathbb{P}_v(y)(1-\mathbb{P}_v(y))}{N_v}}},$$

and rejects the null hypothesis of sign congruence if $(\mathbb{P}_c(y) - 0.5)(\mathbb{P}_v(y) - 0.5) < 0$ and $\min(|z_c|, |z_v|) > \bar{z}_\alpha$. The threshold $\bar{z}_\alpha$ is set to a critical value $\Phi^{-1}(1 - \alpha)$, and we select $\alpha = 0.05$ such that $\bar{z}_\alpha = 1.645$.¹⁶

Deploying the Miller, Molinari, and Stoye (2025) sign-congruence test on the prior literature, we find that only 19% of experiments can be statistically differentiated from the attainable region $S_X \cup S_y$, rejecting stability under Proposition 3(1). For the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets, 12% of experiments similarly reject stability under Proposition 3(1). Thus, rather than the majority of observations indicating significant violations of stable preferences, as the traditional analysis would suggest, the majority of observations are actually

¹⁶If $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ are known to be highly negatively correlated (i.e., a correlation smaller than around -0.8) an adjustment to $\bar{z}_\alpha$ is required, and when the correlation is -1, the threshold value for $\alpha = 0.05$ is actually $\bar{z}_\alpha = 1.96$. Russek-Cohen and Simon (1993) Table 2 provides a tabulation of the relevant threshold for $\alpha = 0.05$ depending on the correlation. For testing Proposition 3(1), the issue of negative correlation is largely irrelevant as $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$ should generally be positively correlated; within the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets 414 of 424 experiments have a positive correlation and the smallest correlation is -0.19.

within the attainable region for stable preferences associated with Proposition 3(1) and only few significantly reject stability. Moving to the larger attainable regions associated with Propositions 2 and 1 indicates that most of the experiments that reject stability under Proposition 3(1)'s strict assumptions governing the heterogeneity of preferences and noise may be accommodated if preference heterogeneity is not required to be symmetric or noise is not required to be independent.

In the prior literature 20% of experiments lie outside the attainable region of $\text{conv}(C_y \cup C_X) \cup \text{conv}(V_y \cup V_X)$ for Proposition 2, and 9% can be statistically differentiated from the attainable region, rejecting Proposition 2.¹⁷ In the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets, 4% of observations lie outside $\text{conv}(C_y \cup C_X) \cup \text{conv}(V_y \cup V_X)$, and zero can be statistically differentiated from the attainable region.¹⁸ With the even greater admissibility of Proposition 1, virtually all experiments can be accommodated under stable preferences. In the prior literature, only 12% of experiments lie outside of $\text{conv}(S_y \cup S_X)$ and 5% can be statistically distinguished from the attainable region, rejecting Proposition 1.¹⁹ In the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets, only 1% of experiments lie outside of $\text{conv}(S_y \cup S_X)$, and zero can be statistically distinguished from the attainable region.²⁰

In both the prior literature and the McGranaghan et al. (2024a) and McGranaghan

¹⁷Here, as well, we deploy the technique of sign-congruence testing from Miller, Molinari, and Stoye (2025) using their approach to more general null hypotheses by implementing a base change corresponding to the linear restrictions of the boundaries for $\text{conv}(C_y \cup C_X) \cup \text{conv}(V_y \cup V_X)$. Notably, the base change induces negative correlation that requires adjustment of the threshold value $\bar{z}_\alpha$ in most cases, which we implement in our analysis.

¹⁸Accounting for the paired nature of the data and adjusting the corresponding z -scores (as well as the threshold value for $\bar{z}_\alpha$ given any difference in correlation), we still find that zero experiments can be statistically differentiated from the attainable region.

¹⁹To develop a statistical test, we consider the linear restrictions imposed on by the lower and upper boundaries of $\text{conv}(S_y \cup S_X)$, $L : \mathbb{P}_v = \mathbb{P}_c - 0.5$ and $U : \mathbb{P}_v = \mathbb{P}_c + 0.5$. The associated z -scores for testing these linear restrictions are

$$z_L = \frac{\mathbb{P}_v(y) - \mathbb{P}_c(y) + 0.5}{\sqrt{\frac{\mathbb{P}_v(y)(1-\mathbb{P}_v(y))}{N_v} + \frac{\mathbb{P}_c(y)(1-\mathbb{P}_c(y))}{N_c}}}$$

and

$$z_U = \frac{\mathbb{P}_v(y) - \mathbb{P}_c(y) - 0.5}{\sqrt{\frac{\mathbb{P}_v(y)(1-\mathbb{P}_v(y))}{N_v} + \frac{\mathbb{P}_c(y)(1-\mathbb{P}_c(y))}{N_c}}}.$$

We consider an observation to be significantly outside of the attainable region if $\mathbb{P}_v(y) < \mathbb{P}_c(y)$ and $z_L < -1.645$ or $\mathbb{P}_v(y) > \mathbb{P}_c(y)$ and $z_U > 1.645$.

²⁰Accounting for the paired nature of the data and adjusting the corresponding z -scores, we still find that zero experiments can be statistically differentiated from the attainable region.

et al. (2024b) datasets few observations meaningfully reject the hypotheses of stable preferences generated by our propositions. The analyses of Propositions 1, 2, and 3(1), however, are fundamentally limited as they cannot make use of critical information related to the location of mean preferences or the connections between choices and valuations to provide more stringent tests of preference stability. We turn next to these analyses relying only on the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets

Evaluating Proposition 3(2) Using Information on Mean Preferences

Evaluation of Proposition 3(2) requires the analyst to know the location of a given choice experiment, y , relative to mean preferences, μ . For each of the 424 choice-valuation experimental pairs in McGranaghan et al. (2024a) and McGranaghan et al. (2024b) we have an estimate of the mean preference derived from the average valuation, $\bar{y}_{vi}$, and we construct the distance to mean preference as $y - \bar{y}_{vi}$.

Proposition 3 assumes that preferences are symmetric around their mean. With symmetric noise, this implies that $y_{vi} - \bar{y}_{vi}$ should be symmetric about zero. Before deploying any evaluation of Proposition 3(2) it is useful to evaluate this assumption. Figure 2.9 presents a histogram of $y_{vi} - \bar{y}_{vi}$ across all 68,448 observations with an overlaid symmetric distribution centered at zero for comparison. While the mean value of $y_{vi} - \bar{y}_{vi}$ is 0 by construction, the median value is \$0.48 in expected value indicating a failure of symmetry. This suggests that it would be inappropriate to evaluate Proposition 3(2) strictly because symmetry of preferences may not hold. Nonetheless, the data do not stray far from the symmetric benchmark, such that it may still be instructive to evaluate the proposition.

Proposition 3(2) has three components elaborating attainable choice probabilities when choice experiments are conducted at $y < \mu$, $y = \mu$, and $y > \mu$. In Figure 2.6, we present the combinations $(\mathbb{P}_c(y), \mathbb{P}_v(y))$, focusing on experiments where the distance to mean preference measure $y - \bar{y}_{vi}$ is significantly negative, indistinguishable from zero, and significantly positive,²¹ with reference to corresponding regions of admissibility under Proposition 3(2). Of the 424 experiments, a total of 26% lie outside of their respective attainable regions; and 17% can be statistically differentiated from their respective attainable regions, rejecting stability under Proposition

²¹To assess the significance of the distance to mean preference relative to zero, we conduct a one-sided means test with a threshold z -score of 1.645 and corresponding $\alpha = 0.05$.

3(2).²² These rejections are illustrated as shaded points in Figure 2.6.²³

When compared to our test of Proposition 3(1) on the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets, we find that 5% more experiments reject Proposition 3(2). This highlights the value of having information on the location of the choice experiment relative to mean preference: The analyst can construct stricter tests of stable preferences, as stability requires observations in S_X to be accompanied by $y - \mu < 0$ and observations in S_Y to be accompanied by $y - \mu > 0$. Figure 2.6, Panels A and C illustrate a number of observations within either S_X or S_Y , consistent with Proposition 3(1), but accompanied by the wrong sign of $y - \bar{y}_{vi}$. These experiments are the ones differentially identified by Proposition 3(2) relative to Proposition 3(1) as inconsistent with stable preferences. Importantly, however, they are few, such that the vast majority of experiments remain consistent with stable preferences.

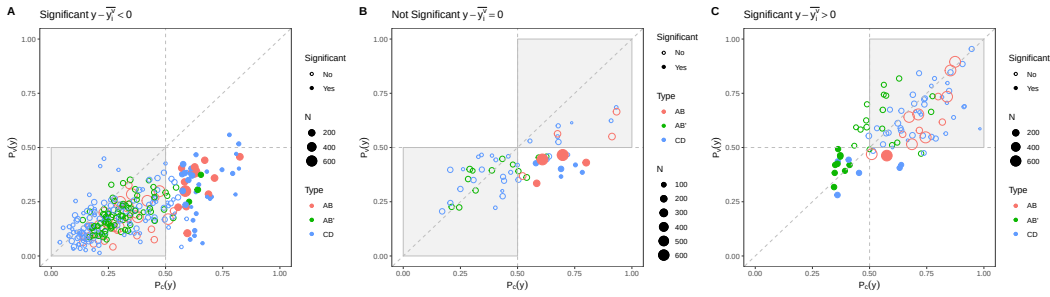


Figure 2.6: Evaluating Proposition 3(2).

Notes: Figure presents relation between aggregate choice and valuation probabilities, $\mathbb{P}_c(y)$ and $\mathbb{P}_v(y)$, for 424 experiments from McGranaghan et al. (2024a) and McGranaghan et al. (2024b). Size corresponds to experimental sample size. Color corresponds to different types of tasks: *AB* certain amount vs. binary lottery (red), *AB'* certain amount vs. trinary lottery (green), and *CD* binary lottery vs. binary lottery (blue). Shaded gray regions correspond to regions of attainable choice-valuation experimental pairs under the assumptions of Proposition 3(2). Solid shading of dots corresponds to observations identified to deviate significantly from the gray shaded region at the 5% level. Panel A presents observations that exhibit significant evidence of $y - \bar{y}_{vi} < 0$ with attainable observations $\mathbb{P}_c(y), \mathbb{P}_v(y) \leq \frac{1}{2}$. Panel C presents observations that exhibit significant evidence of $y - \bar{y}_{vi} > 0$ with attainable observations $\mathbb{P}_c(y), \mathbb{P}_v(y) \geq \frac{1}{2}$. Panel B presents observations where $y - \bar{y}_{vi}$ cannot be statistically differentiated from zero with attainable observations $\mathbb{P}_c(y), \mathbb{P}_v(y) \leq \frac{1}{2}$ or $\mathbb{P}_c(y), \mathbb{P}_v(y) \geq \frac{1}{2}$.

²²For these tests, we again deploy the sign congruence tests of Miller, Molinari, and Stoye (2025) with $\bar{z}_\alpha = 1.645$, but adjust in Figure 2.6 Panels A and C to account for the testing of sign-congruence separately for S_X and S_Y .

²³Panel A identifies 45 rejections out of 279 experiments, Panel B identifies 9 rejections out of 58 experiments, and Panel C identifies 16 rejections out of 87 experiments.

Evaluating Proposition 5 Using Connection Between Choices and Valuations

Proposition 5 states that under suitable regularity assumptions, if a choice experiment is conducted at the mean preference, the conditional choice probability of someone whose valuation is also equal to the mean preference should be $P_c(y|y_{vi} = y = \mu) = \frac{1}{2}$. Similarly, $P_c(y|y_{vi} = y > \mu) > \frac{1}{2}$ and $P_c(y|y_{vi} = y < \mu) < \frac{1}{2}$.

Figure 2.7, Panel A provides an initial illustration of conditional choice probabilities by focusing on individuals where $y_{vi} \approx y$, specifically $|y - y_{vi}| \leq 1.25$ in expected value.²⁴ For these observations we construct 40 bins of $y - \bar{y}_{vi}$ aggregating across different experiments, and within each bin calculate the conditional choice probability $P_c(y|y_{vi} \approx y)$. Under Proposition 5, when $y - \bar{y}_{vi} = 0$, the conditional choice probability should be $\frac{1}{2}$, and when $y - \bar{y}_{vi} > 0$ (< 0) the conditional choice probability should be greater than (less than) $\frac{1}{2}$. The data generally adhere to these broad predictions for sign congruence of $y - \bar{y}_{vi}$ and $P_c(y|y = y_{vi}) - 0.5$, and indeed a linear fit to the full underlying dataset indicates that when $y - \bar{y}_{vi}$ is zero, the conditional choice probability is remarkably close to $\frac{1}{2}$ with an estimate of 0.514 (clustered s.e. = 0.005).

This initial illustration carries with it an indictment of the intuitive conjecture for analyzing preference stability through conditional calibration. Naive analysis would investigate whether subjects who state $y_{vi} \approx y$ exhibit a choice probability of $\frac{1}{2}$ when asked to choose between X and y , irrespective of the distance to mean preferences. Panel A reveals that this choice probability indeed varies substantially with our measure of the distance to mean preferences, $y - \bar{y}_{vi}$; it ranges from approximately 25% to 75% within the data. These extreme choice probabilities provide no concrete rejection of stable preferences unless they are accompanied by the incongruent values for the sign of $y - \bar{y}_{vi}$. We now turn to that more precise analysis of Proposition 5.

Figure 2.7, Panel B evaluates Proposition 5 in the data by again focusing attention on observations with $|y - y_{vi}| \leq 1.25$. For these observations, we relate $y - \bar{y}_{vi}$ to the conditional choice probability $P_c(y|y_{vi} \approx y)$ separately for each experiment. Of the 424 experiments, 423 have individual observations with $|y - y_{vi}| \leq 1.25$, though only 220 have 30 or more observations such that credible estimates for the conditional choice probability $P_c(y|y_{vi} \approx y)$ can be calculated. Of the 423 experiments, 38% exhibit incongruent signs for $y - \bar{y}_{vi}$ and $P_c(y|y_{vi} \approx y) - 0.5$. Again deploying the Miller, Molinari, and Stoye (2025) testing approach, we find that 17% of experiments can be statistically differentiated from the predictions of stable preferences for the

²⁴Of 68,448 individual observations, 19,647 satisfy this criterion.

congruent relationship between $y - \bar{y}_{vi}$ and $P_c(y|y_{vi} \approx y) - 0.5$.²⁵ These significant observations are illustrated in Figure 2.7, Panel B.

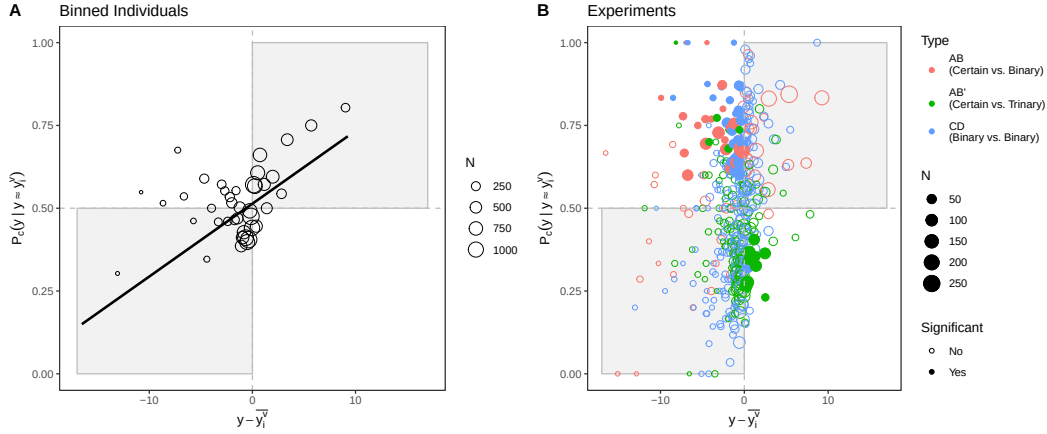


Figure 2.7: Evaluating Proposition 5.

Notes: Figure presents relation between conditional choice probability, $P_c(y|y \approx y_{vi})$ and distance to mean valuation, $y - \bar{y}_{vi}$ for 423 experiments and individual observations of $|y - y_{vi}| \leq 1.25$ from McGranaghan et al. (2024a) and McGranaghan et al. (2024b). Gray regions correspond to regions of attainable behavior under the assumptions of Proposition 5. Panel A presents binned individual data across all experiments along with the size corresponding to the number of observations in a given bin. Panel B presents experiment-level data with the size corresponding to the number of individual observations with $|y - y_{vi}| \leq 1.25$ in each experiment. Panel B color corresponds to different types of tasks: *AB* certain amount vs. binary lottery (red), *AB'* certain amount vs. trinary lottery (green), and *CD* binary lottery vs. binary lottery (blue). Panel B solid shading of dots corresponds to observations identified to deviate significantly from the gray shaded region at the 5% level.

Structured Empirical Analysis

The results from our tests of Propositions 1 to 5 do not suggest substantial deviations from preference stability. While $P_c(y)$ and $P_v(y)$ may appear to differ quantitatively and $P_c(y|y_{vi} \approx y)$ may differ substantially from $\frac{1}{2}$, when analyzed with the inferential lens guided by our theoretical development, few statistical deviations from stable preferences are actually present. However, these analyses make no functional assumptions on the nature of heterogeneity and noise. With further assumptions, stricter stability tests can be generated and quantitative estimates for the extent of preference instability can be constructed. We turn now to these more structured empirical analyses.

²⁵Focusing on those experiments with 30 or more observations, 39% exhibit non-congruent signs and 20% can be statistically differentiated from the attainable region for stable preferences under Proposition 5.

Probit Analysis

In Table 2.1, columns (1) through (4) we conduct the probit regression analysis of equation (2.2) with our normalized distance measures, $y - \bar{y}_{vi}$ and $y - y_{vi}$ as dependent variables. Choice probabilities are systematically related to both distance to mean preferences and distance to individual valuations. When both distances are close to zero the choice probability is projected to be quite close to $\frac{1}{2}$, $\Phi(\alpha) = 0.508$ (s.e. = 0.004). Though we reject the stable preference null of $\alpha = 0$, the quantitative estimate for $\Phi(\alpha)$ suggests a failure of less than 1%-age point overall. Thus, the probit estimator provides consistent evidence with our less-structured evaluation: the data are remarkably close to the benchmark of stable preferences overall. The estimate of $\mu_v - \mu_c$ (measured in expected value units) derived from the probit estimation on the full data set is \$0.205 (0.103). Stable preferences also requires that $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2} \in [0, 1]$, and for the full dataset we obtain an estimate of 0.602 (0.020). Thus, we fail to reject stability by this test, and the point estimate suggests a substantial covariance between the preferences recruited by choices and valuations relative to the total variability of y_{vi} .

In Table 2.1, columns (2)-(4), we repeat the analysis of column (1) separately for the different types of choices in McGranaghan et al. (2024a) and McGranaghan et al. (2024b): Certain vs. Binary AB problems, Certain vs. Trinary AB' problems, and Binary vs. Binary CD problems. For both AB and AB' problems the null of stable preferences is rejected, with estimates suggesting failures on the order of 10%-age points away from $\Phi(0) = \frac{1}{2}$. These correspond to estimated preference deviations $\mu_v - \mu_c$ on the order \$3 in expected value in each case. Interestingly, the deviations are in opposite directions: subjects are more likely to choose a safe certain option than their valuations would imply when the alternative is a binary lottery, while they are less likely to choose a safe certain option when the alternative is a trinary lottery. Correspondingly $\mu_v - \mu_c$ is positive in AB problems and negative in AB' problems. These directional differences counter the general theme of the literature that risk aversion is systematically greater in choices than in valuations (i.e., that $\mu_v - \mu_c > 0$ generally). For CD problems, we cannot reject stable preferences and the point estimate for $\mu_v - \mu_c$ is small, in the neighborhood of 10 cents. In no case do we reject the implication of stability that $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2} \in [0, 1]$, with point estimates suggesting substantial connections between choice and valuation preferences. Interestingly, for AB' problems the point estimates suggests a relative covariance so high that it would be inconsistent with $\gamma_v^2 = \gamma_c^2$, though the estimate of $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$ cannot be statistically

Table 2.1: Probit Estimation

Dependent Variable: Sample:	<i>Chose Safer Alternative = 1</i>			
	All	<i>AB</i>	<i>AB'</i>	<i>CD</i>
		Certain vs. Binary	Certain vs. Trinary	Binary vs. Binary
	(1)	(2)	(3)	(4)
Constant (α)	0.020 (0.010)	0.293 (0.014)	-0.272 (0.018)	0.023 (0.013)
$y - \bar{y}_{vi}$ (β_1)	0.039 (0.002)	0.046 (0.002)	-0.006 (0.006)	0.097 (0.006)
$y - y_{vi}$ (β_2)	0.059 (0.002)	0.051 (0.002)	0.090 (0.005)	0.079 (0.005)
# Observations	68448	25816	16816	25816
# Clusters	3002	3002	2102	3002
$\Phi(\alpha)$	0.508 (0.004)	0.615 (0.005)	0.393 (0.007)	0.509 (0.005)
$\mu_v - \mu_c$	0.205 (0.103)	3.000 (0.137)	-3.277 (0.318)	0.130 (0.074)
$\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$	0.602 (0.020)	0.528 (0.021)	1.077 (0.076)	0.450 (0.026)

Notes: Table presents probit regressions for choice of the safer alternative in a choice experiment on a constant, harmonized distance to mean preference, and harmonized distance to valuation. Standard errors clustered at individual level in parentheses. $\Phi(\text{Constant})$ provides an estimate of the standard normal CDF evaluated at the constant where distance to mean preference and distance to valuation is zero, corresponding to test of stable preferences. Standard error for $\Phi(\alpha)$ calculated using the delta method. Preference difference, $\mu_v - \mu_c$, calculated as $\frac{\alpha}{\beta_1 + \beta_2}$ with standard error calculated via the delta method. Relative covariance, $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$, calculated as $\frac{\beta_2}{\beta_1 + \beta_2}$ with standard error calculated via the delta method.

differentiated from 1.

Full Maximum Likelihood Analysis

To complement the probit analysis of section 2.4 and provide estimates that do not rely on the probit's replacement of μ_v with its sample analog, we also conduct the full maximum likelihood exercise described in section 2.3.²⁶ In the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) experiments there are a total of 84 different problem types upon which we could conduct our maximum likelihood approach. These problem types each fix a choice type, a valuation elicitation method, and a

²⁶An additional difference between the methods is that the full maximum likelihood approach recognizes that price list switch points are actually interval coded, while the probit estimates assume that y_{vi} is the midpoint of the valuations implied by an individual's switch point.

set of probability parameters and vary only the experimental value of y . Thus, each problem type combines valuations with 4-6 different choice experiments. Appendix 2.C provides a description of each problem type.

For each problem type, we conduct the maximum likelihood estimation making use of all valuation data and all choice experiments to provide estimates of μ_c and μ_v alongside the parameters of the variance covariance matrix.²⁷ Figure 2.8, Panel A presents densities for $\mu_v - \mu_c$ derived from the 84 maximum likelihood estimates. The message is essentially the same as that provided by Table 2.1: overall $\mu_v - \mu_c$ is small, estimated to have a median of -23 cents; and this median masks substantial heterogeneity with $\mu_v - \mu_c$ being positive for *AB* problems with a median estimate of \$2.73, negative for *AB'* problems with a median estimate of -\$2.57, and approximately zero for *CD* problems with a median estimate of \$0.10.

Figure 2.8, Panel C presents densities for the relative covariance, $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$. Here as well, the message is consistent with our probit estimates. The relative covariance is generally consistent with stable preferences, with a median estimate of 0.77. This median also masks substantial heterogeneity with $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$ being largest for *AB'* problems and smallest for *CD* problems. Interestingly, among our maximum likelihood estimates, a number suggest deviations from stable preferences with values of $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$ substantially in excess of one.

In order to further compare the probit and full maximum likelihood estimates, we conduct the probit estimation disaggregating into the 84 problem types noted above. Panels A and B of Figure 2.8 compare the resulting estimates for $\mu_v - \mu_c$, while Panels C and D compare the estimates for $\frac{\gamma_{cv}}{\gamma_v^2 + \sigma_v^2}$. For mean preference deviations, the estimates largely agree and are correlated 0.89. For relative covariance, the estimates are similarly in line and are correlated 0.91.

The probit and maximum likelihood estimates of Figure 2.8 reveal that different broad conditions lead to different estimates of preference instability. Importantly, within conditions there is also substantial heterogeneity. Understanding the drivers of this heterogeneity may lead to further insights on when preference stability is more or less likely to hold. In Appendix 2.C, we relate our estimates of $\mu_v - \mu_c$ to specific problem parameters, indicating some important predictive power of the probability of the high prize in the riskier alternative. When this probability is high, choices tend to show less risk aversion relative to valuations, and when this

²⁷ As in our other analyses, these values are then harmonized to make them comparable in expected value units.

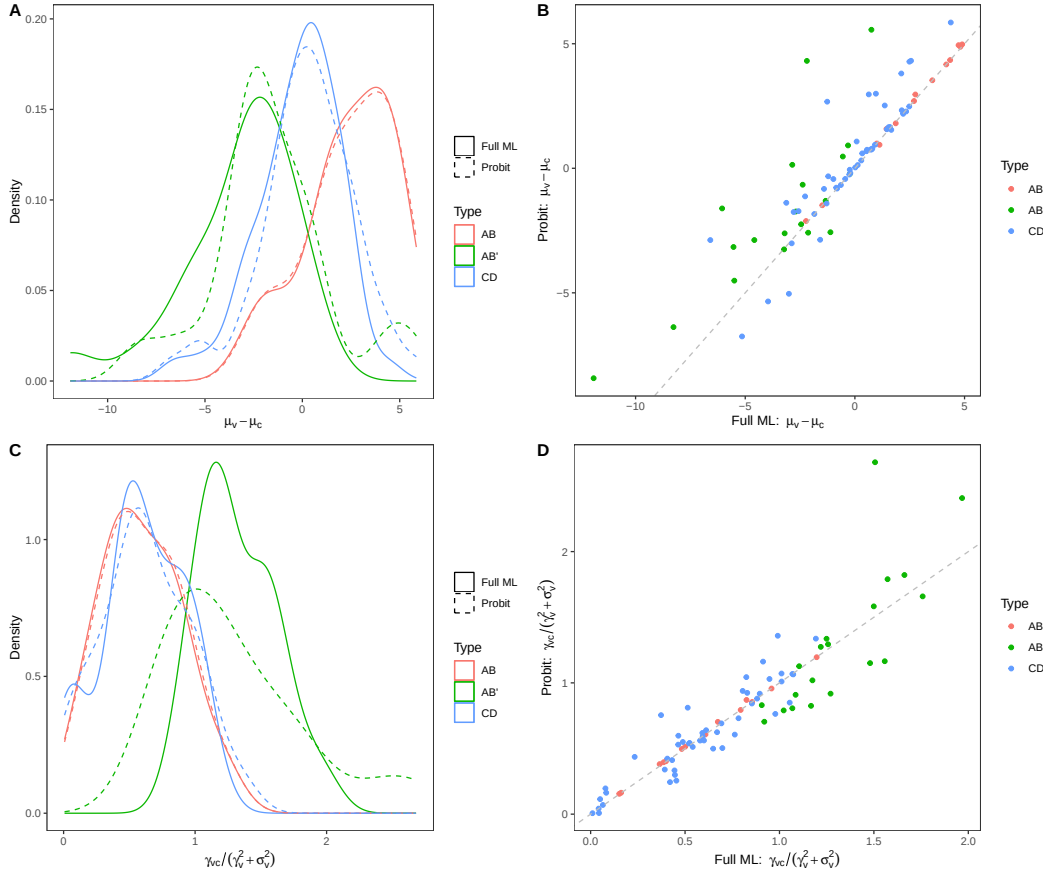


Figure 2.8: Estimated Preferences from Probit and Full ML Analysis.

Notes: Figure presents 84 separate full maximum likelihood and probit estimates for $\mu_v - \mu_c$ and $\gamma_{cv}/(\gamma_v^2 + \sigma_v^2)$ for 424 choice-valuation experimental pairs from McGranaghan et al. (2024a) and McGranaghan et al. (2024b). Color corresponds to different types of tasks: *AB* certain amount vs. binary lottery (red), *AB'* certain amount vs. trinary lottery (green), and *CD* binary lottery vs. binary lottery (blue). Panels A and C present densities for estimated values of $\mu_v - \mu_c$ and $\gamma_{cv}/(\gamma_v^2 + \sigma_v^2)$ separately for full maximum likelihood (solid) and probit (dashed) estimation approaches. Panels B and D present correlation between full maximum likelihood estimates and probit estimates.

probability is low the opposite is true. Understanding why higher probabilities of this form lead to less risk aversion in choices compared to valuations is an interesting puzzle worthy of future study.

Comparing Traditional and New Approaches

Having developed our structured empirical approach for evaluating preference stability and estimating the degree of instability, a natural question is how the lessons derived from this new approach relate to the traditional inferential approach of comparing $\mathbb{P}_c(y)$ to $\mathbb{P}_v(y)$.

In Appendix 2.D we explore one simple approach for comparison which contrasts the traditional deviation in probabilities, $\mathbb{P}_c(y) - \mathbb{P}_v(y)$, to our probit estimate of the same deviation for the ideal experiment where $y - \mu_v = 0$ for the ideal subject with $y - y_{vi} = 0$, $\Phi(\alpha) - \frac{1}{2}$. The two approaches often lead to qualitative distinctions in conclusions: Though highly correlated, 34% of experiments exhibit opposite signs for $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ and $\Phi(\alpha) - \frac{1}{2}$. Quantitatively, distinctions emerge as well: the traditional $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ exceeds $\Phi(\alpha) - \frac{1}{2}$ by a median of 11%-age points.

The fundamental source of this disagreement across approaches is that $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ is measured from all subjects' behavior for any implemented value of y . In contrast $\Phi(\alpha) - \frac{1}{2}$ is calculated as a projection to an idealized experimental value of y with $y - \mu_v = 0$ (and an idealized subject with $y - y_{vi} = 0$). In most instances the analyst has no information with regards to how far the choice experiment y is from either μ_v or y_{vi} and has no ability to restrict their sample in a way to make both distances zero. If they could, and the probit assumptions were correct, the two approaches would coincide.

2.5 Conclusion

Measurement of economic preferences faces important conceptual challenges if said preferences exhibit substantial instability across measurement techniques. In this manuscript we develop tools to assess preference stability across techniques, principally choices and valuations, under maintained assumptions for noisiness in measurement and heterogeneity of preferences. Though a wide body of prior exercises points to failures of procedural invariance across measurements in the form of preference reversals or choice-valuation inconsistencies, we show that the inference drawn from such exercises is substantively flawed. Instead, virtually all prior choice-valuation experimental pairs along with 424 new choice-valuation experimental pairs from McGranaghan et al. (2024a) and McGranaghan et al. (2024b) are consistent with stable preferences if heterogeneity of preferences is not required to be symmetric about mean preferences and noise is not required to be independent of preferences. Even under our most restrictive assumptions, only 15-20% of choice-valuation experimental pairs lie outside of the predictions of stable preferences. More structured analysis reveals the magnitudes of any preference instability is small in expected value overall, but that there are important nuances across tasks.

The conclusions we draw on preference stability derive from a convenience sample of available prior data. Importantly, however, we assemble the largest collection

of experiments to date comparing choices to valuations, so it is plausible that some conclusions drawn from our empirical application will be useful for future experimental exercises.

One critical question for future exercises is “How should preferences be measured?” Some researchers argue for specific methodology under impressionistic judgments about the superiority of choice tasks for revealing preferences or the superiority of valuation tasks for supporting introspection about preferences (see, e.g., Freeman, Halevy, and Kneeland, 2019; Freeman and Mayraz, 2019). These arguments carry two flaws: first, they are made under the assumption that choices and valuations recruit different preferences without our necessary tests for instability; and second, they presume that one measurement must be a more accurate reflection of some “true” preference of interest. Permitting the assumption that choices and valuations recruit different preferences, who is to say that one is more “true” than another.

Even if one preference is “true” and the other is not, if they are highly correlated the distinction may not be highly relevant. If choices and valuations recruit similar mean preferences with a high degree of covariance, as our exercises reveal overall, then the question of “How should preferences be measured?” becomes a question of which technique is noisier. In Appendix 2.E, we explore this question and provide bounds on the correlation in preferences recruited by choices and valuations. Interestingly, we find that choices are substantially more variable than valuations, $\gamma_c^2 + \sigma_c^2 > \gamma_v^2 + \sigma_v^2$, for virtually all problem types. If choices and valuations were to recruit a common preference, then this would indicate that choices are substantially noisier. Of course, our data cannot disentangle variability derived from preference heterogeneity from variability derived from noise, and so in a limiting case (i.e., $\sigma_c^2 = \sigma_v^2 = 0$) this would imply only that the preferences for choices are more heterogeneous. In that same limiting case, one obtains a lower bound for the correlation in preferences recruited by choices and valuations. Interestingly, even these lower bounds are high, similar to those discussed in the psychology literature on stability of personality traits or the economics literature on stability of preferences using identical measurements over time (for discussion on the relationship between stability in economic preference measurements and psychological personality traits, see Meier and Sprenger, 2015). Thus, our data indicate that choices and valuations recruit similar preferences that are substantially correlated, but that valuations do so with plausibly greater precision. It will be critical to know whether these conclusions are robust in other empirical environments.

Future work should take the central insights of this manuscript to deploy empirical tests of stable preferences on purpose-built experimental samples. Critical to the development of such work is the selection of choice-valuation experimental pairs. We show that for the traditional approach of examining equality of choice and valuation measures, the analyst must have conducted the choice experiment at the mean preference. While this is not possible to do in practice, an alternative may be to conduct choice experiments that bracket the mean preference and use local information to consider a counterfactual choice experiment at the mean.²⁸ Thus, future experimental designs for evaluating preference stability should move away from their standard designs to ones where a variety of choice-valuation experimental pairs bracketing mean preferences are considered. For the classic preference reversal studies, this means considering a range of P -bets and $\$$ -bets bracketing where the average true valuations for both are equal. For the choices-vs-valuations studies, this means considering a range of choice experiments, y , bracketing the expected location of the average valuation, $\overline{y_{vi}}$.

Our results also have important implications for assessments of procedural invariance using other experimental techniques. For example, a substantial debate continues on appropriate payment mechanisms in experimental protocols depending on whether subjects “isolate” each experimental choice (see, e.g., Holt, 1986; Starmer and Sugden, 1991; Harrison and Swarthout, 2014; Cox, Sadiraj, and Schmidt, 2015). This debate revolves around evidence of potential inconsistencies between choice proportions depending on whether only a single choice is made or the same choice is embedded in a larger experiment. Our propositions extend naturally to such comparisons and highlight that inferring violations of isolation based on deviations from equality in choice proportions, though intuitive, is substantively flawed if choices are noisy. Thus, this work on procedural invariance and the credibility of specific payment protocols may also require reconsideration.

²⁸This is, in essence, what our probit analysis does, though less structured approaches are also possible.

2.A Additional Figures

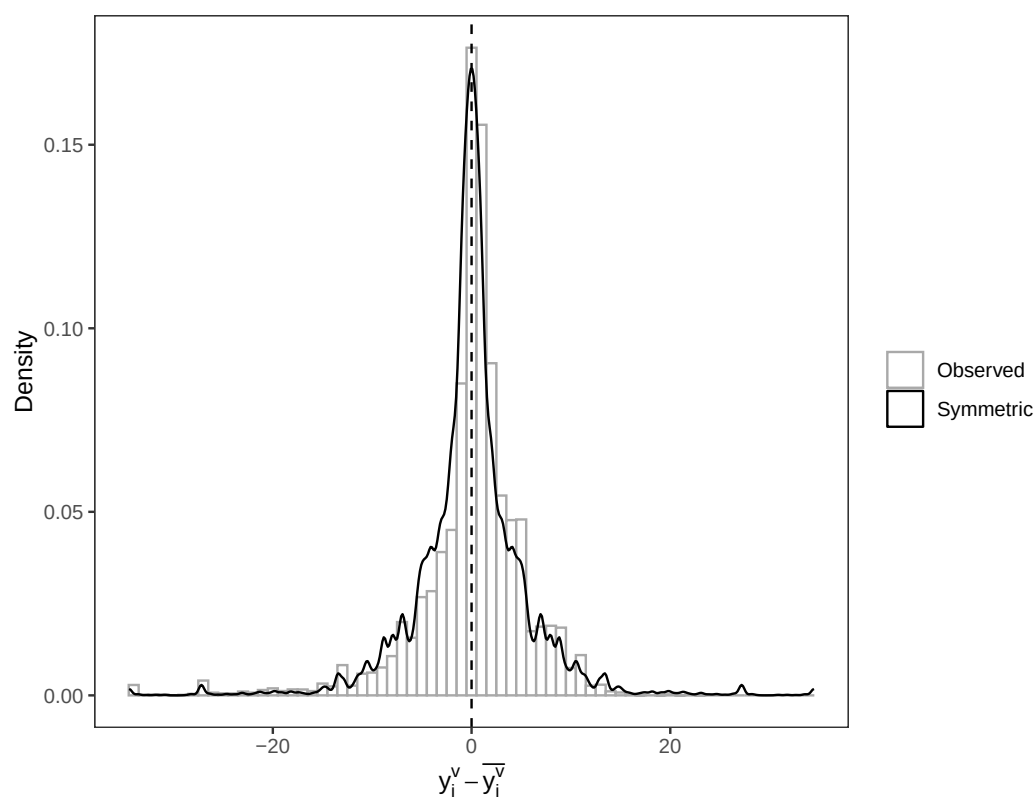


Figure 2.9: Evaluating Symmetry of $y_{vi} - \overline{y_{vi}}$.

Notes: Figure presents histogram of “ $y_{vi} - \overline{y_{vi}}$ ” for 68,448 observations in McGranaghan et al. (2024a) and McGranaghan et al. (2024b) (gray bars). Values harmonized to be in expected value units. Solid black line corresponds to symmetrized distribution constructed from a combined dataset of “ $y_{vi} - \overline{y_{vi}}$ ” and “ $\overline{y_{vi}} - y_{vi}$.” Median of “ $y_{vi} - \overline{y_{vi}}$ ” is 0.48.

2.B Proofs

Proof of Lemma 1

Recall that $S_y = [1/2, 1]^2$ and $S_X = [0, 1/2]^2$, and

$$C_y \equiv S_y \cap \{(a, b) \mid b \geq a\} \quad \text{and} \quad C_X \equiv S_X \cap \{(a, b) \mid b \leq a\},$$

and

$$V_y \equiv S_y \cap \{(x, y) \mid y \leq x\} \quad \text{and} \quad V_X \equiv S_X \cap \{(x, y) \mid y \geq x\}.$$

We show that any point in $S_y \cup S_X$ is attainable under symmetric noises. Note that $S_y \cup S_X = C_y \cup C_X \cup V_y \cup V_X$.

Proof of Lemma 1. We prove $\mathcal{A}_i \subset S_y \cup S_X$ in Section 2.3. We now show the converse, i.e., any point in $S_y \cup S_X$ is attainable under some noise and valuation satisfying the assumptions. Indeed, fix any $(a, b) \in C_y$, so that $\frac{1}{2} \leq a \leq b \leq 1$ and let $y_i^* < y$. We seek a pair of continuous and symmetric noise distributions $\sim (F_c, F_v)$ such that $F_c(y - y_i^*) = a$, $F_v(y - y_i^*) = b$. For instance, we can set

$$F_c(x) = \begin{cases} \frac{a - \frac{1}{2}}{y - y_i^*}x + \frac{1}{2} & 0 \leq x < y - y_i^* \\ x + a - (y - y_i^*) & y - y_i^* \leq x < 1 - a + (y - y_i^*) \\ 1 & 1 - a + (y - y_i^*) \leq x \end{cases}$$

and

$$F_v(x) = \begin{cases} \frac{b - \frac{1}{2}}{y - y_i^*}x + \frac{1}{2} & 0 \leq x < y - y_i^* \\ x + b - (y - y_i^*) & y - y_i^* \leq x < 1 - b + (y - y_i^*) \\ 1 & 1 - b + (y - y_i^*) \leq x \end{cases},$$

where the negative parts are then determined by symmetry.

An analogous construction shows that we can obtain any point in C_X when $y_i^* > y$. Moreover, a symmetric argument applies for obtaining any point in V_y and V_X for F_c and F_v that satisfy Assumption 1.

□

Proof of Proposition 2

Proof of Proposition 2. Recall that $S_X = [0, \frac{1}{2}]^2$ and $S_y = [\frac{1}{2}, 1]^2$. It follows from Lemma 1 that the attainable region under symmetric noise is a subset of

$\text{conv}(S_y \cup S_x)$. First, we prove that under symmetric noise that is independent of preferences, then $(\mathbb{P}_c(y), \mathbb{P}_v(y)) \in \text{conv}(C_X \cup C_y) \cup \text{conv}(V_X \cup V_y)$.

Let

$$T_1 = \left\{ (x, y) \in \mathbb{R}^2 \mid y < 2x - 1, y < \frac{1}{2}x \right\}, \quad T_2 = \left\{ (x, y) \in \mathbb{R}^2 \mid y > 2x, y > \frac{1}{2}x + \frac{1}{2} \right\}.$$

Note that $\text{conv}(C_X \cup C_y) \cup \text{conv}(V_X \cup V_y) = \text{conv}(S_X \cup S_y) \setminus (T_1 \cup T_2)$. Let F_c and F_v be symmetric CDFs. We show that for all $x, y \in \mathbb{R}$ and for all $\alpha \in [0, 1]$,

$$\alpha F_c(x) + (1 - \alpha)F_c(y) = \frac{2}{3} \quad (2.3)$$

implies

$$\alpha F_v(x) + (1 - \alpha)F_v(y) \geq \frac{1}{3}. \quad (2.4)$$

Without loss of generality, assume $x \geq y$. First, we observe that $x > 0$ since (2.3) implies $F_c(x) \geq \frac{2}{3}$. If $y \geq 0$, then (2.4) always holds since $F_v(x), F_v(y) \geq \frac{1}{2}$. Thus, we only consider the case when $x > 0 > y$.

If $x \leq -y$, then $F_c(-x) = 1 - F_c(x) \geq F_c(y)$. Thus, we have

$$\alpha F_c(x) + (1 - \alpha)(1 - F_c(x)) \geq \alpha F_c(x) + (1 - \alpha)F_c(y) = \frac{2}{3}. \quad (2.5)$$

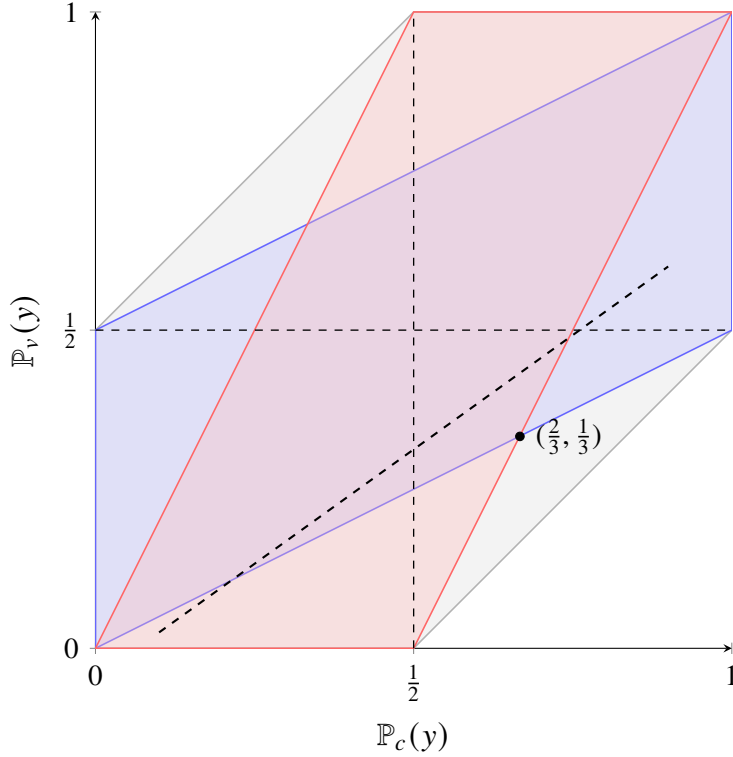
Rearranging (2.5), we obtain $\alpha \geq \frac{F_c(x) - \frac{1}{3}}{2F_c(x) - 1}$. Since $F_c(x) \geq \frac{2}{3}$, we have $\alpha \geq \frac{2}{3}$ as the right hand side term is minimized when $F_c(x) = 1$. Moreover, since $x > 0$, we have $F_v(x) \geq \frac{1}{2}$. Thus, (2.4) holds.

If $x > -y$, then $F_v(y) \geq F_v(-x) = 1 - F_v(x)$. Moreover,

$$\begin{aligned} \alpha F_v(x) + (1 - \alpha)F_v(y) &\geq \alpha F_v(x) + (1 - \alpha)F_v(-x) = (2\alpha - 1)F_v(x) + (1 - \alpha) \\ &\geq \min \left\{ \alpha, \frac{2\alpha - 1}{2} + 1 - \alpha \right\} \\ &= \min \left\{ \alpha, \frac{1}{2} \right\}. \end{aligned}$$

Note that $\alpha = \frac{\frac{2}{3} - F_c(y)}{F_c(x) - F_c(y)}$ by (2.3). Since $F_c(x) \geq \frac{2}{3}$ and $F_c(y) \leq 0.5$, we have $\alpha \geq \frac{1}{3}$, which shows that (2.3) implies (2.4). Thus, any point that is a convex combination of two points in $(F_c(x), F_v(x)), (F_c(x), F_v(y))$ and whose first coordinate is $\frac{2}{3}$ must have a higher second coordinate than $\frac{1}{3}$ and therefore cannot

lie in T_1 . In other words, the convex hull of the set $\{(F_c(x), F_v(x)) \mid x \in \mathbb{R}\}$ does not intersect T_1 . By symmetry, we conclude the same for T_2 . Thus, the attainable region under symmetric noise that is independent of preferences is $\text{conv}(C_X \cup C_Y) \cup \text{conv}(V_X \cup V_Y)$.



Next, we show that every point of $\text{conv}(C_X \cup C_Y) \cup \text{conv}(V_X \cup V_Y)$ can be attained. We show that there is a single pair of noise distributions satisfying the assumptions such that any point $(a, b) \in \text{conv}(C_Y \cup C_X)$ is attainable under some population distribution H . The proof for $\text{conv}(V_Y \cup V_X)$ follows a symmetric argument.

Let the valuation noise be uniformly distributed on $[-1, 1]$ and the choice noise be uniformly distributed on $[-3, -2] \cup [2, 3]$, so that they are both symmetric and choice is noisier than valuation noise. Note that for any y ,

$$(\mathbb{P}_{ci}(y), \mathbb{P}_{vi}(y)) = \begin{cases} (1, 1) & y_i^* \leq y - 3 \\ (\frac{1}{2}, 1) & y_i^* \in [y - 2, y - 1] \\ (\frac{1}{2}, 0) & y_i^* \in [y + 1, y + 2] \\ (0, 0) & y_i^* \geq y + 3. \end{cases}$$

Since $(1, 1), (\frac{1}{2}, 1), (\frac{1}{2}, 0), (0, 0)$ form the extreme points of $C_Y \cup C_X$, $(\mathbb{P}_c(y), \mathbb{P}_v(y))$ can be any point in $\text{conv}(C_Y \cup C_X)$ under some population distribution of y_i^* . $\square$

Proof of Proposition 4

Recall that the aggregate choice probability of choosing y over X and the aggregate probability of reporting a valuation for X lower than y are given by

$$\mathbb{P}_c(y) = G_c(y - \mu), \mathbb{P}_v(y) = G_v(y - \mu),$$

where G_c and G_v are CDFs of $\eta_i + \varepsilon_i^c$ and $\eta_i + \varepsilon_i^v$, respectively. An important implication of homogeneous and symmetric noise is that when η_i is symmetric, G_c and G_v are symmetric.

Moreover, if η_i is unimodal, then if choices are noisier, $\eta_i + \varepsilon_i^c$ is noisier than $\eta_i + \varepsilon_i^v$ (and vice versa if valuations are noisier).

Proof of Proposition 4. Let η_i with a symmetric and unimodal CDF H and $\varepsilon_i^c, \varepsilon_i^v$ be symmetric and independent of η_i , where ε_i^c is noisier than ε_i^v . The proof for when ε_i^v is noisier than ε_i^c follows a symmetric argument. Let G_c and G_v be the CDFs of $\eta_i + \varepsilon_i^c$ and $\eta_i + \varepsilon_i^v$, respectively. Since ε_i^c and ε_i^v are symmetric and independent of symmetric η_i , G_c and G_v are symmetric, i.e., $G_c(x) = 1 - G_c(-x)$ and $G_v(x) = 1 - G_v(-x)$ for all x .

We first show that $G_c(t) \neq G_v(t)$ when $t > 0$. The case when $t < 0$ follows a symmetric argument. Let F_c and F_v be the respective CDFs of ε_i^c and ε_i^v . Let $D(t) = F_v(t) - F_c(t)$. Note that since F_v and F_c are symmetric, D is odd, i.e., $D(t) = -D(-t)$ for all t . Since η_i is independent of ε_i^c and ε_i^v , we have

$$G_v(t) - G_c(t) = \int_{-\infty}^{\infty} (F_v(t-x) - F_c(t-x))h(x)dx = \int_{-\infty}^{\infty} D(t-x)h(x)dx,$$

and by setting $s = t - x$, we have

$$G_v(t) - G_c(t) = \int_{-\infty}^0 D(s)h(t-s)ds + \int_0^{\infty} D(s)h(t-s)ds.$$

Substituting $s = -x$ in the first integral, we have

$$\int_0^{\infty} D(-x)h(t+x)dx,$$

which equals

$$\int_0^{\infty} -D(x)h(t+x)dx,$$

since D is odd. Thus, we have

$$G_v(t) - G_c(t) = \int_0^{\infty} D(x)(h(t-x) - h(t+x))dx.$$

Since ε_i^c is noisier than ε_i^v , we have $D(x) \geq 0$ for all $x \geq 0$ and since h is unimodal, $h(t-x) - h(t+x) \geq 0$ for all $t, x \geq 0$. Moreover, since ε_i^c and ε_i^v are distinct near 0 and since h is strictly unimodal, we have

$$D(x) > 0, \quad h(t-x) - h(t+x) > 0 \text{ for all } t \geq 0$$

for some small positive x . Thus, for all $t > 0$, $G_v(t) - G_c(t) > 0$.

□

Proof of Proposition 5

Proof. We first show that conditional calibration holds at $y = \mu$ under Assumptions 1, 3, and 7. Recall that under Assumption 7, we have

$$\mathbb{P}_c[\mu \mid y_{vi} = \mu] = \frac{\int F_c(\mu - y_i^*) f_v(\mu - y_i^*) dH(y_i^*)}{\int f_v(\mu - y_i^*) dH(y_i^*)}.$$

By Assumption 1, we can write $F_c(x) = \frac{1}{2} + O(x)$ for some odd function O . Thus, we have

$$\mathbb{P}_c[\mu \mid y_{vi} = \mu] = \frac{1}{2} + \frac{\int O(\mu - y_i^*) f_v(\mu - y_i^*) h(y_i^*) dy_i^*}{\int f_v(\mu - y_i^*) h(y_i^*) dy_i^*}.$$

Note that the left side of the above equation cannot depend on μ , so it is without loss of generality to set $\mu = 0$. It follows that the numerator of the second term is zero since O is odd, f_v is even by Assumption 1, and h is symmetric about $\mu = 0$ under Assumption 3.

We next show that under additional Assumptions 4, 7 and 8, conditional calibration only holds at μ . Consider $\mathbb{P}_c[y \mid y_{vi} = y]$ where $y < \mu = 0$. By Assumptions 1, 3 and 4, $f_v(z)h(y-z) \leq f_v(-z)h(y+z)$, for all $z \geq 0$. Note that this inequality is strict on some interval around zero by Assumptions 4 and 8. Hence, since $f(z)h(y-z) + f(-z)h(y+z)$ is even and $f(z)h(y-z) - f(-z)h(y+z)$ is non-positive for $z \geq 0$, $f(z)h(y-z)$ can be written as $E(z) + N(z)$ where E is even and N is non-positive on $\mathbb{R}_+$ and negative for small enough positive arguments.

Thus, letting $z = y - y_i^*$, we have

$$\begin{aligned} \mathbb{P}_c[y \mid y_{vi} = y] &= \frac{1}{2} + \frac{\int O(y - y_i^*) f_v(y - y_i^*) h(y_i^*) dy_i^*}{\int f_v(y - y_i^*) h(y_i^*) dy_i^*} \\ &= \frac{1}{2} + \frac{\int O(z) E(z) dz}{\int (E(z) + N(z)) dz} + \frac{\int O(z) N(z) dz}{\int (E(z) + N(z)) dz}. \end{aligned}$$

Since $O(z)E(z)$ is odd, we have

$$\mathbb{P}_c[y \mid y_{vi} = y] = \frac{1}{2} + \frac{\int O(z)N(z) dz}{\int (E(z) + N(z)) dz}.$$

By Assumption 8, O is positive near zero. Since N is negative for small enough positive arguments, $\mathbb{P}_c[y \mid y_{vi} = y] < \frac{1}{2}$. A similar argument shows that when $y > \mu = 0$, $\mathbb{P}_c[y \mid y_{vi} = y] > \frac{1}{2}$. $\square$

2.C The McGranaghan et al. (2024a) and McGranaghan et al. (2024b) Datasets

The primary focus of McGranaghan et al. (2024a) and McGranaghan et al. (2024b) is testing Expected Utility in both choices and valuations using versions of Allais' (1953) Common Ratio (CR) and Common Consequence (CC) situations. For their choice paradigm, these studies ask the individual to make choices of the following forms:

$$AB : A \equiv (M, 1) \text{ vs. } B \equiv (H, p; 0, 1 - p),$$

$$AB' : A \equiv (M, 1) \text{ vs. } B' \equiv (H, pr; M, 1 - r; 0, r(1 - p)),$$

$$CD : C \equiv (M, r; 0, 1 - r) \text{ vs. } D \equiv (H, pr; 0, 1 - pr),$$

for fixed prizes $M, H > M$, $p \in (0, 1)$, and $r \in (0, 1)$. In different choice experiments, different combinations of (p, r, H, M) are considered. AB problems ask the individual to choose between a safe certain option A and a risky binary lottery B ; AB' problems ask the individual to choose between a safe certain option A and a risky trinary lottery B' ; and CD problems ask the individual to choose between two binary lotteries, a safer option C and a riskier option D .

Across the two papers McGranaghan et al. (2024a) and McGranaghan et al. (2024b) have 64 different AB choice experiments, 120 different AB' choice experiments, and 240 different CD choice experiments conducted at different values of the parameter vector (p, r, H, M) , yielding a total of 424 choice experiments. Each choice experiment was completed by an average of 161.4 subjects yielding a total of 68,448 individual choices.

For each choice experiment, McGranaghan et al. (2024a) and McGranaghan et al. (2024b) also collect corresponding valuations. They consider two different types of valuation tasks: choose- m and choose- h . In choose- m valuation tasks, the authors fix $(p, r, H = \$30)$ and ask subjects to state the value of $M = m_i^v$ that make the two options indifferent. In the related choice tasks, M was varied in a way to attempt to bracket the average m_i^v valuation. In choose- h valuation tasks, the authors fix

$(p, r, M = \$p \times 30)$ and ask subjects to state the value $H = h_i^v$ that make the two options indifferent. In the related choice tasks, H was varied in a way to attempt to bracket the average h_i^v valuation.

More risk-averse individuals should state lower values of m_i^v in choose- m tasks and higher values of h_i^v in choose- h tasks. Thus, the proportion of subjects whose valuations imply that they would prefer the safer option at a given M (with fixed (p, r, H)) is $\mathbb{P}_v(M) = \mathbb{P}(M > m_i^v)$ in choose- m tasks; and the proportion of subjects whose valuations suggest that they would prefer the safer option at a given H (with fixed (p, r, M)) is $1 - \mathbb{P}_v(H) = \mathbb{P}(h_i^v > H)$ in choose- h tasks. Similarly, for related choice tasks the choice probabilities for the *safer* option would be $\mathbb{P}_c(M) = \mathbb{P}(m_i^c < M)$ and $1 - \mathbb{P}_c(H) = \mathbb{P}(h_i^c > H)$, where $m_i^c = m_i^* + \epsilon^c$ and $h_i^c = h_i^* + \epsilon^c$ reflect effective indifference points applied in choice following the theoretical development of section 2.3.

To harmonize across choose- m and choose- h environments we consider the probability of preferring the safe option at a given value of y :

$$\mathbb{P}_j(y) = \begin{cases} \mathbb{P}_j(M), & \text{if choose-}m; M = y \\ 1 - \mathbb{P}_j(H) & \text{if choose-}h; H = y \end{cases}$$

for $j = \{c, v\}$.

For each of the 424 choice experiments it is straightforward to take the corresponding valuation tasks and assess the mean valuation, $\overline{m}_i^v$ or $\overline{h}_i^v$, to provide a proxy for mean preferences, μ . We can then calculate the distance to the mean preference estimate for a given choice experiment as $M - \overline{m}_i^v$ or $\overline{h}_i^v - H$.²⁹ Clearly, this distance measure carries a different meaning in different choice experiments. In order to normalize our notion of distance to the mean preference, we adjust all distances to expected value units. We use as our measure of distance to mean preference,

$$y - \overline{y}_{vi} = \begin{cases} r^{1-1_{AB}}(M - \overline{m}_i^v), & \text{if choose-}m; M = y \\ p * r^{1-1_{AB}}(\overline{h}_i^v - H) & \text{if choose-}h; H = y. \end{cases}$$

We similarly harmonize the distance to an individual's valuation

$$y - y_{vi} = \begin{cases} r^{1-1_{AB}}(M - m_i^v), & \text{if choose-}m; M = y \\ p * r^{1-1_{AB}}(h_i^v - H) & \text{if choose-}h; H = y. \end{cases}$$

²⁹The directional difference of considering the choice experimental value minus the average valuation for choice experiments connected to choose- m tasks and the opposite for choose- h tasks is implemented so that larger values of distance correspond intuitively to greater likelihood of choosing the safe option.

These normalize quantities for $\mathbb{P}_c(y)$, $\mathbb{P}_v(y)$, $y - \overline{y_{vi}}$, and $y - y_{vi}$ are the ones we use in our analysis of the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets.

Problem Types and Full Maximum Likelihood Estimation

In the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) dataset problems are differentiated by their type $\{AB, AB', CD\}$, whether the valuations are choose- m or choose- h , and the problem parameters (p, r) . There are a total of 84 combinations of choose- m /choose- $h \times (p, r) \times \{AB, AB', CD\}$.³⁰ Within each problem type, there are a number of choice experiments consisting of different values of M for a fixed (p, r, H) or different values of H for a fixed (p, r, M) (i.e., different values of the choice experimental variable y). Within each problem type we have an average of around 5 choice experiments, leading to our total of 424 experiments.

Maximum Likelihood Estimation

The 84 problem types are appropriate for deploying our full maximum likelihood approach as across the five or so values of y we can articulate the likelihood of both choosing the safer option and stating a valuation within a given range. Let y^e be the experimental value of y assigned to a given individual. Under the normal-normal-normal formulation, the likelihood of the individual choosing the safe alternative y^e over X and exhibiting a valuation in $(y_{i,L}^v, y_{i,H}^v)$ is

$$L_{i,safe}(e) = N \left(\begin{pmatrix} -\infty, y^e \\ y_{i,L}^v, y_{i,H}^v \end{pmatrix}; \begin{pmatrix} \mu_c \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_c^2 + \gamma_c^2 & \gamma_{cv} \\ \gamma_{cv} & \sigma_v^2 + \gamma_v^2 \end{pmatrix} \right).$$

And the likelihood of choosing the riskier alternative X over y^e and exhibiting a valuation in $(y_{i,L}^v, y_{i,H}^v)$ is

$$L_{i,risky}(e) = N \left(\begin{pmatrix} y^e, \infty \\ y_{i,L}^v, y_{i,H}^v \end{pmatrix}; \begin{pmatrix} \mu_c \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_c^2 + \gamma_c^2 & \gamma_{cv} \\ \gamma_{cv} & \sigma_v^2 + \gamma_v^2 \end{pmatrix} \right).$$

Let $\mathbf{1}_{safe}(e)$ be an indicator for choosing the safe alternative y^e over X . Then, the individual's likelihood is

$$L_i(e) \mathbf{1}_{safe}(e) L_{i,safe}(e) + (1 - \mathbf{1}_{safe}) L_{i,risky}(e).$$

³⁰Parameter r is irrelevant for AB comparisons and so all values of r are collapsed for a given p for these problems.

For a collection of individuals, i , and choice experiments, e , the total log likelihood would thus be

$$LL = \sum_i \sum_e \log(L_i(e)).$$

Within the McGranaghan et al. (2024a) and McGranaghan et al. (2024b) datasets each individual valuation is related to only one choice experiment so LL is comprised of only $\sum_i$ terms. Maximizing the log likelihood, LL , of observations with respect to the parameters $\mu_c, \mu_v, \sigma_c^2 + \gamma_c^2, \sigma_v^2 + \gamma_v^2, \gamma_{cv}$ is a straightforward exercise yielding estimates of the corresponding parameters using conventional techniques.

In our implementation we implement 84 ML estimators, one for each problem type, using the conjoint gradient algorithm with a maximum of 1000 iterations. If convergence is not achieved within 1000 iterations, we take the final values as our estimates.

Problem Parameters and Estimates of Preference Deviations

The 84 full maximum likelihood estimates (as well as the corresponding probit estimates) of Figure 2.8 reveal that different broad conditions lead to different quantitative estimates of preference instability, $\mu_v - \mu_c$. Median estimates indicate that for AB problems $\mu_v - \mu_c > 0$, for AB' problems $\mu_v - \mu_c < 0$, and for CD problems $\mu_v - \mu_c \approx 0$.

Importantly, within conditions there is also substantial heterogeneity, with each of AB , AB' , and CD indicating some instances of $\mu_v - \mu_c > 0$ and some instances of $\mu_v - \mu_c < 0$. Understanding the drivers of this heterogeneity may lead to further insights on when preference stability is more or less likely to hold.

To analyze the sources of this within-problem heterogeneity we regress the full maximum likelihood estimate of $\mu_v - \mu_c$ on the probability parameter p , controlling for the full suite of interactions choose- m /choose- $h \times \{AB, AB', CD\}$. We find that the value of p is significantly negatively correlated with the value of “ $\mu_v - \mu_c$ ” recovered from the full maximum likelihood exercise (regression coefficient = -5.73, s.e. = 0.67); when p is high, choices tend to show less risk aversion relative to valuations and the opposite when p is low. For tasks where probability parameter r is relevant (AB' and CD problems), we conduct the same analysis with r and find that r is statistically unrelated to the value of “ $\mu_v - \mu_c$ ” (regression coefficient = -1.16, s.e. = 1.34). Figure 2.10 illustrates these results.

Findings of preference instability may be quite sensitive not only to the type of lotteries used (e.g., certain vs. binary compared to binary vs. binary), but also to the probabilities implemented. Understanding the deeper reasons for these predictable effects on instability will be a critical path for future studies.

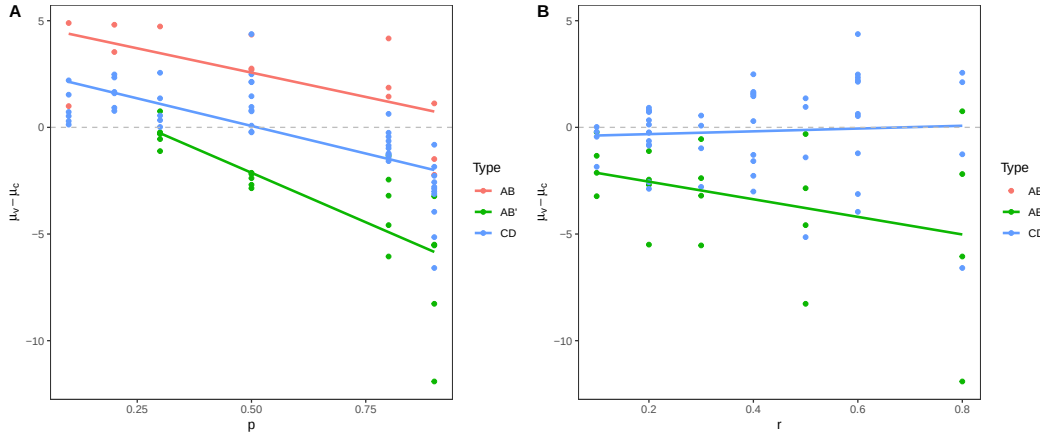


Figure 2.10: Comparative Statics With Respect to Problem Parameters.

Notes: Figure presents 84 full maximum likelihood estimates for $\mu_v - \mu_c$ from McGranaghan et al. (2024a) and McGranaghan et al. (2024b) against the problem parameters p and r . Color corresponds to different types of tasks: AB certain amount vs. binary lottery (red), AB' certain amount vs. trinary lottery (green), and CD binary lottery vs. binary lottery (blue).

2.D Connecting Traditional and Structured Analyses

In Figure 2.11 we relate the deviation in probabilities, $\mathbb{P}_c(y) - \mathbb{P}_v(y)$, to our probit estimate of the corresponding difference for the ideal choice experiment, $\Phi(\alpha) - \frac{1}{2}$.³¹ While 80% of choice-valuation experimental pairs exhibit $\mathbb{P}_c(y) - \mathbb{P}_v(y) > 0$, only 50% of pairs exhibit $\Phi(\alpha) - \frac{1}{2} > 0$. Overall, the disagreement in direction is substantial, with 34% of experimental pairs having opposing signs for $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ and $\Phi(\alpha) - \frac{1}{2}$. The traditional approach finds that 44% of pairs exhibit significantly positive $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ compared to 2% significantly negative. The probit approach finds that 47% of pairs exhibit significantly positive estimates of $\Phi(\alpha) - \frac{1}{2}$ compared to 44% significantly negative.³² The source of this essential difference in message is illustrated along the horizontal line at zero in Figure 2.11, Panel A where a collection of six AB' experimental pairs shows five with positive values of $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ (three

³¹To construct $\Phi(\alpha) - \frac{1}{2}$ at the choice-valuation experimental pair level, we apply our probit estimate from the relevant collection of choice-valuation experimental pairs to each element. That is, if 5 pairs were used for a given probit, the same value of $\Phi(\alpha) - \frac{1}{2}$ applies to all. Thus, in Figure 2.11 values of $\Phi(\alpha) - \frac{1}{2}$ in horizontal groupings of approximately 5 observations.

³²We conduct the test of $H_0 : \alpha = 0$ to determine whether $\Phi(\alpha) - \frac{1}{2}$ is significantly positive or negative by the new approach, avoiding any delta method numerical approximations.

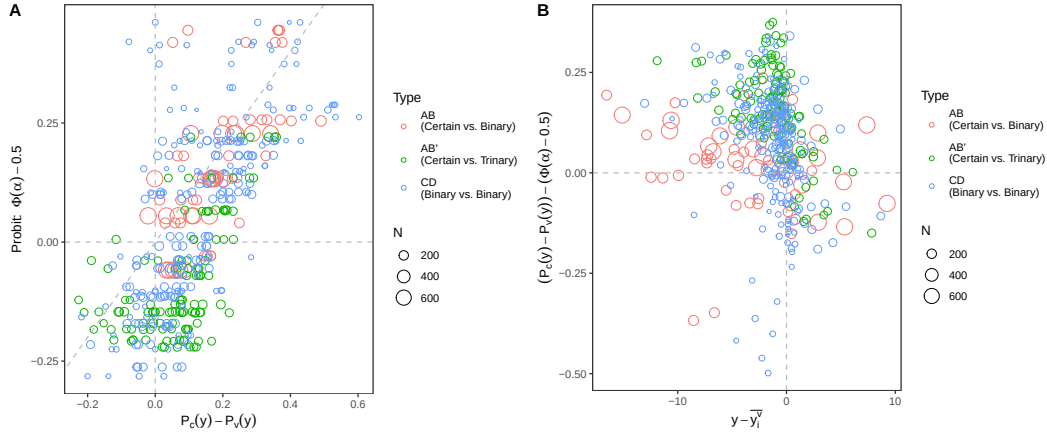


Figure 2.11: Connecting Traditional and Structured Analyses.

Notes: Figure presents comparison between traditional and structured inferential approaches to testing stable preferences for 424 choice-valuation experimental pairs from McGranaghan et al. (2024a) and McGranaghan et al. (2024b). Color corresponds to different types of tasks: *AB* certain amount vs. binary lottery (red), *AB'* certain amount vs. trinary lottery (green), and *CD* binary lottery vs. binary lottery (blue). Data are harmonized such that greater probability corresponds to greater likelihood of preferring the safer alternative. Panel A relates the difference $P_c(y) - P_v(y)$, the traditional measure of instability, to the probit estimate $\Phi(\alpha) - 0.5$, the structured measure of instability. 34% of choice-valuation experimental pairs exhibit disagreement in sign. Panel B relates the extent of disagreement to the location of the choice experiment relative to our estimate of mean preferences, $y - \bar{y}_{vi}$; correlation $\rho = -0.22$, ($p < 0.01$).

significantly so) and one without. When this collection of pairs is considered by the probit technique we estimate $\alpha = 0.01$ (0.06) with $\Phi(\alpha) = 0.51$. This means that when distance to both the mean valuation and a given individual's valuation are both zero, we project a choice probability of approximately 50% because in this instance choices and valuations recruit similar primitives. However, none of the six experimental pairs exhibits $P_c(y) - P_v(y) \approx 0$; the average absolute deviation is 16%-age points.

Our analysis suggests that one driver of the disagreement between the traditional approach and our more structured analysis will be the value of the choice experiment, y , and how far it is from the average preference for valuation, μ_v . Figure 2.11, Panel B relates our measure of $y - \bar{y}_{vi}$ to the extent of disagreement between $P_c(y) - P_v(y)$ and $\Phi(\alpha) - \frac{1}{2}$. The extent of disagreement is significantly correlated with $y - \bar{y}_{vi}$, $\rho = -0.22$ ($p < 0.01$). Further, $y - \bar{y}_{vi}$ is also highly correlated with whether there will be disagreements in the signs of $P_c(y) - P_v(y)$ and $\Phi(\alpha) - \frac{1}{2}$, $\rho = -0.16$ ($p < 0.01$). The disagreement between the traditional approach and ours emerges when choice experiments are selected far from the primitives recruited by valuations. In these instances, when choices and valuations recruit similar

primitives $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ can suggest otherwise; and when choices and valuations recruit different primitives $\mathbb{P}_c(y) - \mathbb{P}_v(y)$ can suggest an opposing deviation.

2.E Variability of Measurement and Bounding the Correlation in Preferences

Our full maximum-likelihood estimation provides estimates of $\sigma_c^2 + \gamma_c^2$ and $\sigma_v^2 + \gamma_v^2$. Under the assumption of stable preferences, $\gamma_c^2 = \gamma_v^2$, and the difference between the estimates reveals which measurement is noisier. In Figure 2.12, Panel A we relate $\sqrt{\sigma_c^2 + \gamma_c^2}$ to $\sqrt{\sigma_v^2 + \gamma_v^2}$ for our 84 maximum likelihood estimates. In 89% of cases choices are estimated to have more variability than valuations, consistent with them being noisier.

In addition to the variability of choices and valuations, our maximum likelihood estimates capture the covariance in preferences, γ_{cv} . If choices and valuations were measured with no noise, $\sigma_c^2 = \sigma_v^2 = 0$, then $\frac{\gamma_{cv}}{\sqrt{\sigma_c^2 + \gamma_c^2} \sqrt{\sigma_v^2 + \gamma_v^2}} = \frac{\gamma_{cv}}{\gamma_c \gamma_v}$, the correlation between the preferences recruited in choice and valuation elicitation. Thus the measure $\frac{\gamma_{cv}}{\sqrt{\sigma_c^2 + \gamma_c^2} \sqrt{\sigma_v^2 + \gamma_v^2}}$ provides a lower bound for the correlation in preferences. Figure 2.12, Panel B presents densities for this value. The median value of $\frac{\gamma_{cv}}{\sqrt{\sigma_c^2 + \gamma_c^2} \sqrt{\sigma_v^2 + \gamma_v^2}}$ is 0.32. Understood as a measure of correlation in preferences, this quantity is within the general range considered for stability of preferences and personality traits (for discussion see Meier and Sprenger (2015)). Importantly, there are sizable differences in correlations across problems with *CD* problems showing the lowest stability, *AB* problems being intermediate, and *AB'* problems showing the highest stability.

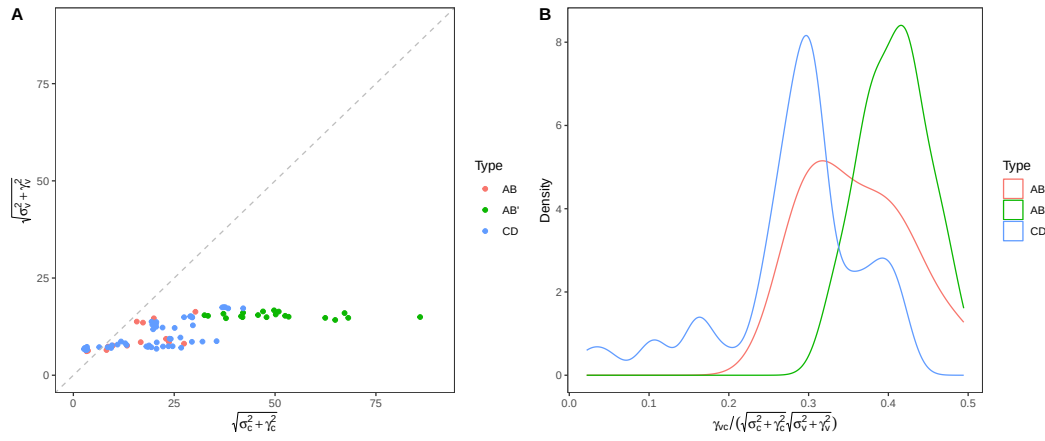


Figure 2.12: Variability of Elicitation and Correlation in Preferences.

Notes: Figure presents 84 maximum likelihood estimates of variability in choices and valuations along with their covariance based on the experiments of McGranaghan et al. (2024a) and McGranaghan et al. (2024b). Color corresponds to different types of tasks: *AB* certain amount vs. binary lottery (red), *AB'* certain amount vs. trinary lottery (green), and *CD* binary lottery vs. binary lottery (blue). . Colors correspond to different problem types. Panel A relates $\sqrt{\sigma_c^2 + \gamma_c^2}$ to $\sqrt{\sigma_v^2 + \gamma_v^2}$. Panel B provides densities of $\frac{\gamma_{cv}}{\sqrt{\sigma_c^2 + \gamma_c^2} \sqrt{\sigma_v^2 + \gamma_v^2}}$.

BIBLIOGRAPHY

- Allais, Maurice (1953). “Le comportement de l’homme rationnel devant le risque: critique des postulats et axiomes de l’école américaine.” In: *Econometrica*, pp. 503–546.
- Brown, Alexander L. and Paul J. Healy (2018). “Separated decisions.” In: *European Economic Review* 101, pp. 20–34.
- Cox, James C., Vjollca Sadiraj, and Ulrich Schmidt (2015). “Paradoxes and mechanisms for choice under risk.” In: *Experimental Economics* 18.2, pp. 215–250.
- Freeman, David J., Yoram Halevy, and Terri Kneeland (2019). “Eliciting risk preferences using choice lists.” In: *Quantitative Economics* 10.1, pp. 217–237.
- Freeman, David J. and Guy Mayraz (2019). “Why choice lists increase risk taking.” In: *Experimental Economics* 22, pp. 131–154.
- Gillen, Ben, Erik Snowberg, and Leeat Yariv (2019). “Experimenting with measurement error: Techniques with applications to the caltech cohort study.” In: *Journal of Political Economy* 127.4, pp. 1826–1863.
- Grether, David M. and Charles R. Plott (1979). “Economic theory of choice and the preference reversal phenomenon.” In: *American Economic Review* 69.4, pp. 623–638.
- Harrison, Glenn W. and J. Todd Swarthout (2014). “Experimental payment protocols and the bipolar behaviorist.” In: *Theory and Decision* 77.3, pp. 423–438.
- Holt, Charles A. (1986). “Preference reversals and the independence axiom.” In: *American Economic Review* 76.3, pp. 508–515.
- Lichtenstein, Sarah and Paul Slovic (1971). “Reversals of preference between bids and choices in gambling decisions.” In: *Journal of Experimental Psychology* 89.1, p. 46.
- eds. (2006). *The Construction of Preference*. New York: Cambridge University Press.
- Lu, Yong (2025). “Preference reversals: a meta-analysis.” In: *Available at SSRN* 5464034.
- McGranaghan, Christina, Kirby Nielsen, Ted O’Donoghue, Jason Somerville, and Charles D. Sprenger (2024a). *Connecting common ratio and common consequence preferences*. Tech. rep. Working paper.
- (2024b). “Distinguishing common ratio preferences from common ratio effects using paired valuation tasks.” In: *American Economic Review* 114.2, pp. 307–347.

- Meier, Stephan and Charles D. Sprenger (2015). “Temporal stability of time preferences.” In: *Review of Economics and Statistics* 97.2, pp. 273–286.
- Miller, Douglas L., Francesca Molinari, and Jörg Stoye (2025). “Testing sign congruence between two parameters.” In: *Journal of Applied Econometrics*.
- Mosteller, Frederick and Philip Noguee (1951). “An experimental measurement of utility.” In: *Journal of Political Economy* 59.5, pp. 371–404.
- Russek-Cohen, Estelle and Richard M. Simon (1993). “Qualitative interactions in multifactor studies.” In: *Biometrics*, pp. 467–477.
- Slovic, Paul (1995). “The construction of preference.” In: *American Psychologist* 50.5, p. 364.
- Starmer, Chris and Robert Sugden (1991). “Does the random-lottery incentive system elicit true preferences? An experimental investigation.” In: *American Economic Review* 81.4, pp. 971–978.
- Stigler, George J. and Gary S. Becker (1977). “De gustibus non est disputandum.” In: *American Economic Review* 67.2, pp. 76–90.
- Thurstone, Louis L. (1931). “The indifference function.” In: *The Journal of Social Psychology* 2.2, pp. 139–167.
- Tversky, Amos and Richard H. Thaler (1990). “Anomalies: preference reversals.” In: *Journal of Economic Perspectives* 4.2, pp. 201–211.

Chapter 3

A SEPARABILITY FOUNDATION FOR RANDOM COEFFICIENTS LOGIT

3.1 Introduction

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

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

¹Random coefficients logit is a workhorse of empirical IO (Nevo, 2000; Moon, Shum, and Weidner, 2018; Gandhi and Nevo, 2021), popularized by the seminal paper of Berry, Levinsohn, and Pakes, 1995. It belongs to the broader class of mixed logit models, although some empirical work uses the two terms interchangeably (see, e.g., the first paragraph in Nevo, 2000). Following Strzalecki, 2025, we reserve the term mixed logit for mixtures over unrestricted utility functions and use random coefficients logit for mixtures of multinomial logit rules with linear utilities. See the literature review and §3.3 for further discussion.

through the variance of revenue or other uncertain quantities.

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

Our monotonicity axiom requires that an action that yields a dominating outcome (i.e., higher in every dimension) is taken with higher probability. This axiom places an ordinal restriction on the choice probabilities within a menu but places no restrictions across menus. Continuity is the technical assumption that a small change in outcomes leads to a small change in choice probabilities.

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

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

From the point of view of an analyst (say, employed by the rideshare company) who

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

Our main result is that random coefficients logit rules are the only ones that satisfy monotonicity, continuity and IID (Theorem 1). Moreover, the coefficient distribution is identical across all menus. Thus, even though the IID axiom only restricts predictions for product menus, its conjunction with monotonicity and continuity implies that all choices—including in non-product menus—are made according to the same rule.

This result provides a simple foundation for this widely used choice rule. It also shows that one of the axioms is violated by any other stochastic choice rule, including one-shot probit and separable probit. Indeed, one-shot probit violates IID, while separable probit violates monotonicity (see §3.5). The theorem thus highlights that modeling even one decision instance with a rule that is inconsistent with random coefficients logit carries hidden global assumptions. Regardless of how behavior is modeled on other menus, such a decision maker must either be influenced by the presence of irrelevant decisions (violating IID), fail to choose better actions with higher probability (violating monotonicity), or be extremely sensitive to outcomes (violating continuity).

In our main result we model separability by the addition of the outcomes in $\mathbb{R}^n$. In §3.6 we carry out a similar exercise, in a setting with an abstract outcome space and an abstract operation that captures separability. This highlights that it is separability—rather than the particular additive structure of $\mathbb{R}^n$ —that is the main driver of our results.

Related Literature

In this paper we provide an axiomatization of random coefficients logit. This term is often conflated with mixed logit, and we follow Strzalecki, 2025 to distinguish the two. Mixed logit allows unrestricted utility heterogeneity and can approximate any random utility model (McFadden and Train, 2000), so axiomatizing mixed logit is close in spirit to axiomatizing stochastic rationality more generally (Manski, 1977;

Falmagne, 1978; McFadden and Richter, 1990; Clark, 1996). A characterization of mixed logit along these lines was obtained by Saito, 2018.

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

The restriction to linear utilities is substantive: linear-utility mixtures do not span all random utility behaviors unless the number of alternatives is small relative to the outcome dimension (Saito, 2018; Lu and Saito, 2025). Such mixtures give rise to linear pure characteristic models (LPCM), which we discuss further in §3.3 alongside a characterization of a related class by Gul and Pesendorfer, 2006.

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

A separate literature microfound multinomial logit and related rules through information or cognitive frictions. Matějka and McKay, 2015 show that multinomial logit captures the behavior of a utility-maximizing individual with entropy-based at-

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

Our analysis of separability in general outcome spaces (§3.6) connects to a classical literature showing that choice independence across dimensions yields additively separable utility (Debreu, 1959; Fishburn, 1965; Segal, 1992; Wakker, 2013; Cho, 2022). Similarly, we demonstrate that choice independence across product menus implies that choices are driven by a utility function that is additively separable across the product components.

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

3.2 Model

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

label representing the compound action of taking a_1 and a_2 together. Note that this condition implies that $\mathcal{A}$ is infinite.² We further assume that $\mathcal{A}$ is countable.

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

The key setting we study is $\mathcal{O} = \mathbb{R}^n$; in §3.6 we generalize beyond $\mathbb{R}^n$. To build intuition it is useful to keep in mind the example of $\mathcal{O} = \mathbb{R}$. This benchmark outcome space can be used to model decision-makers who compare actions by a single number, such as their monetary reward—and we accordingly refer to outcomes as payoffs. We use $\mathcal{O} = \mathbb{R}$ in the examples used to illustrate the definitions of this section.

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

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

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

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

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

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

This framework sits between two standard stochastic-choice domains. In the classical random-utility tradition, a decision problem is a set of primitive alternatives, and all choice-relevant information is carried by the identities of those alternatives (see, e.g., Block and Marschak, 1959; Luce and Suppes, 1965; Falmagne, 1978). In pure-characteristics models, alternatives are represented directly as points in $\mathbb{R}^n$, so alternatives with the same vector are identified at the level of the model (see, e.g., Saito, 2018; Lu and Saito, 2025). Our framework keeps these two roles distinct. As in characteristics-based models, the relevant information about action a in menu

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

(A, o) is contained in $o(a)$. However, we do not reduce the menu to the set of outcomes $o(A) \subset \mathbb{R}^n$. Retaining action labels lets us form compound actions, such as (a_1, a_2) , and relate actions across different menus. The roles played by actions and outcomes in our framework fit the game-theoretic tradition in which the actions available to a player are mere labels and only the payoffs matter; indeed, for $O = \mathbb{R}$, a menu (A, o) can be viewed as a one-player normal-form game with action set A and payoff function o .

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

One family of widely used stochastic choice rules consists of the *independent additive random utility models* (IARU), which are given by

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

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

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

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

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

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

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

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

It is parameterized by a probability measure μ over the logit parameter $\beta \in \mathbb{R}^n$. More generally than random coefficients logit, one can consider mixtures of (non-linear)

multinomial logit rules, which is the class of mixed logit rules. In the empirical literature these terms are often conflated (Nevo, 2000; Gandhi and Nevo, 2021), but in our setting the difference is important, as we discuss in §3.3.

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

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

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

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

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

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

Alternatively put, continuity stipulates that very small changes in the outcomes result in very small changes in choice probabilities. This axiom excludes stochastic choice rules that describe individuals who pay excessive attention to even negligible differences in outcomes. For instance, it is violated by rules for $\mathcal{O} = \mathbb{R}$ that always select one of the highest-outcome actions, regardless of how small the advantage is. Nevertheless, continuity is a common modeling choice made for good reason:

³Our results hold more generally for any partial order, including the trivial one in which each outcome is only comparable to itself. In §3.6 we formally consider general partial orders.

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

people do not always choose the dominant action, especially when the difference between outcomes is minuscule.

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

$$A = A_1 \times A_2 \quad \text{and} \quad o(a_1, a_2) = o_1(a_1) + o_2(a_2). \quad (3.1)$$

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

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

The first test would be well-modeled by the menu (A_1, o_1) , with $A_1 = \{a, b\}$ and $o_1(a) = 0, o_1(b) = 1$. Likewise, the second test would be well-modeled by (A_2, o_2) , with $A_2 = \{s, t\}$ and $o_2(s) = 0, o_2(t) = 1$. The joint decision the subject faces is the product menu

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

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

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

This is not a product menu, even though the action set is a product set (indeed, the same product set), because the outcomes are not additively separable. This example highlights that we do not assume that arbitrary combinations of decisions give rise to product menus. While any such combination involves taking the Cartesian product of the underlying action sets, the associated outcomes need not take the separable form (3.1).

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

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

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

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

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

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

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

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

Decomposability requires that for the product menu,

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

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

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

The IID axiom is a weakening of decomposability, allowing for such correlations while maintaining consistency with predictions for the components. Given a product menu $(A, o) = (A_1, o_1) \otimes (A_2, o_2)$, we denote the marginal choice probability of $a_1 \in A_1$ by

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

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

$$\Phi(a_1 | A, o) = \Phi(a_1 | A_1, o_1) \tag{3.2}$$

for all $a_1 \in A_1$.

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

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

then

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

Thus, IID allows for predictions such as

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

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

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

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

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

IID is satisfied by random coefficients logit, and in particular is satisfied by linear multinomial logit, which moreover satisfies decomposability. The class of rules that satisfy IID is very large. For instance, we can construct such a rule Φ by specifying any choice distribution on each non-product menu, and then defining Φ on product menus to be the product measure given by Φ on each component menu. This construction will in fact satisfy decomposability, and is very flexible.

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

$$\text{SIARU}^u((a_1, a_2) \mid A, o) = \text{SIARU}^u(a_1 \mid A_1, o_1) \cdot \text{SIARU}^u(a_2 \mid A_2, o_2). \quad (3.3)$$

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

That is, choices are made independently in each dimension. For example, if a consumer makes two unrelated choices—e.g., purchasing a milk and a soap—they choose the best milk subject to some noise and the best soap subject to some additional noise, with the noises independent for the two choices. Such a model could also be applied in a dynamic setting: a consumer chooses milk both today and tomorrow, and faces independent shocks for the two periods.⁶ See §3.5 for a formal definition and discussion.

Implications of the Axioms

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

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

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

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

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

Surprisingly, this intuition is wrong. It turns out that if we assume that Φ is monotone and satisfies IID, then $\Phi(a_1 \mid A, o) = 5\%$ implies that $\Phi(a_3 \mid A, o) \geq 83.8\%$. This bound is tight: if we make the stronger assumption that Φ is monotone and decomposable, then $\Phi(a_1 \mid A, o) = 5\%$ implies that $\Phi(a_3 \mid A, o) = 83.8\%$. As we explain in detail below, this is a consequence of our main results. Furthermore, there is a unique Φ that is monotone, decomposable, continuous, and satisfies $\Phi(a_1 \mid A, o) = 5\%$; see Appendix §3.F for details.

⁶We thank an anonymous referee for suggesting this example.

3.3 IID and Random Coefficients Logit

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

$$\text{RCL}^\mu = \int \text{MNL}^\beta d\mu(\beta) \quad (3.4)$$

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

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

The theorem is proved in §3.7 for $\mathcal{O} = \mathbb{R}$ and the multidimensional case is covered in Appendix 3.C. Some of the core ideas are explained in §3.4.

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

In the one-dimensional case $\mathcal{O} = \mathbb{R}$ where outcomes are interpreted as payoffs, $\beta \in \mathbb{R}_{\geq 0}$ determines the (random) noise level of a decision maker who is optimizing, but may stochastically choose a dominated outcome. In the example of the rideshare driver with outcomes $(r, -t, -g, -m)$ where the revenue r is measured in dollars, we can write $\beta = \gamma \cdot p$, where γ is in $\mathbb{R}_{\geq 0}$ and $p = (p_1, p_2, p_3, p_4)$ is a non-negative vector normalized to $p_1 = 1$. We can now interpret the components p_2, p_3 and p_4 as the prices that the driver assigns to their time, gas and mileage. The driver is stochastically optimizing their expected profit, with noise γ . In the choice under ambiguity example, where outcomes are vectors of payoffs indexed by state θ , one can instead normalize p to sum to one, recasting it as a probability measure over the state space Θ . Under this normalization, the decision maker becomes a Bayesian stochastic expected payoff maximizer.

When some coordinates of the outcome vector represent costs rather than payoffs, it is natural to adjust the notion of monotonicity so that the decision maker favors actions with lower values in those coordinates.⁷ Theorem 1 extends straightforwardly to arbitrary partial orders on $\mathbb{R}^n$, since its proof invokes the monotonicity axiom only to ensure that two actions with identical outcomes are chosen with equal probability. Accordingly, if monotonicity is imposed with respect to a given partial order $\geq'$ on $\mathbb{R}^n$, the corresponding version of Theorem 1 states that the rule coincides with RCL^μ for a measure μ supported on the set of $\beta \in \mathbb{R}^n$ for which the linear function $x \rightarrow \beta \cdot x$ is $\geq'$ -monotone. At the extreme, taking $\geq'$ to be the trivial order ($x \geq' y$ if and only if $x = y$) reduces monotonicity to the requirement that equal-outcome actions are selected with equal probability, and μ can be supported on all of $\mathbb{R}^n$. We discuss monotonicity with respect to general partial orders further in §3.6.

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

Mixed Logit, Random Coefficients Logit, and Universal Approximation

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

This property of mixed logit highlights that it is an extremely flexible model, making it challenging to identify and estimate. In contrast, random coefficients logit—

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

in which the utilities u are restricted to be linear—is widely used because it is parsimonious and hence identifiable, while retaining sufficient flexibility to capture many phenomena.⁸

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

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

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

Since $\text{MNL}^{t\beta}$ with $t \rightarrow +\infty$ approximates the choice of a rational individual with $u(x) = \beta \cdot x$, one can show that $\text{LPCM}^\mu(a \mid A, o)$ can be approximated by mixing over such linear logits with β distributed according to μ . As limits of rules satisfying IID and monotonicity, LPCM^μ with μ supported on $\mathbb{R}_{\geq 0}^n$ also satisfies these axioms (this can also be verified directly). However, LPCM^μ violates the continuity assumption and thus is ruled out in Theorem 1.⁹

3.4 Correlations, Decomposability, and Linear Multinomial Logit

The IID axiom does not preclude correlations between separable choices. In this section we tackle two questions. First, we want to better understand what types of

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

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

correlations IID allows, once it is taken in conjunction with continuity and monotonicity. Second, we would like to understand the implications of requiring that there are no correlations, i.e., replacing IID with the stronger axiom of decomposability.

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

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

$$\Phi(o_1(a_1) \cdot o_2(a_2) \mid A, o) \geq \Phi(o_1(a_1) \mid A, o) \cdot \Phi(o_2(a_2) \mid A, o). \quad (3.5)$$

We say that Φ exhibits *zero correlation* if (3.5) holds with equality. A menu (A, o) is said to be non-trivial if o is non-constant.

In the example of the lab experiment in which every correct answer yields a payoff of one, non-negative correlation means that when a subject answers the first question correctly they are (weakly) more likely to answer the second one correctly.

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

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

The proof of Proposition 1 appears in §3.B, and is straightforward. Together with Theorem 1, we conclude that every rule that satisfies our axioms has non-negative correlation, which is not immediately evident from the axioms. In §3.5, we provide an example demonstrating that probit models do not always exhibit non-negative correlation, highlighting the connection between this property and our axioms.

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

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

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

The proof of Proposition 2 appears in Appendix 3.D, and uses Theorem 1 and Proposition 1. We now sketch a direct proof for Proposition 2 for the case $O = \mathbb{R}$. This also provides a simple explanation of the core ideas of the proof of our main result, Theorem 1. The formal proof of Theorem 1 follows a similar path and makes use of the De Finetti Theorem to show that Φ is a convex combination of decomposable rules.

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

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

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

For clarity, consider the particular example

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

¹⁰An extreme point of a convex set is one that cannot be written as a non-trivial convex combination of others. In this light, representation (3.4) mirrors the Choquet theorem: any choice rule satisfying the axioms can be expressed as a mixture of extreme points. Fox et al., 2012 show that this representation is unique, that is $\text{RCL}^\mu = \text{RCL}^{\mu'}$ implies $\mu = \mu'$, so the space of such rules forms an infinite-dimensional simplex. This fact also follows from Proposition 5 in §3.7, which is a part of the proof of our main theorem. In fact, this proposition shows that μ is already identified by the choice probabilities on menus with integer outcomes.

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

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

$$(A, o) \otimes \underbrace{(B, r) \otimes (B, r) \otimes \cdots \otimes (B, r)}_{n \text{ times}}, \quad (3.6)$$

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

$$\Phi(a_3 | A, o) \cdot \Phi(b_0 | B, r)^{59} = \Phi(a_2 | A, o) \cdot \Phi(b_1 | B, r)^{59}. \quad (3.7)$$

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

3.5 Probit Rules

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

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

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

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

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

$$(B, r) = \left\{ \begin{array}{cc} b_0 & b_1 \\ 0 & 1 \end{array} \right\} \quad (3.8)$$

and its “square”

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

The probit predictions are

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

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

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

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

Separable linear probit (SProbit) is a separable IARU model, where the utility is linear and shocks are Gaussian. As in (3.3), it coincides with one-shot linear probit for non-product menus, and for product menus it is given by

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

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

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

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

and let

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

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

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

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

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

3.6 Utility Representations for Abstract Outcome Spaces $\mathcal{O}$

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

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

To define product menus, we endow the outcome space $\mathcal{O}$ with a binary operation $*$ corresponding to combining outcomes in a separable way. To build intuition, consider $\mathcal{O} = \mathbb{R} \times \mathbb{R}_{\geq 0}$, where the first component is interpreted as the mean and the second component as the standard deviation of a stochastic Gaussian monetary

reward. In this setting, product menus correspond to combined choices in which the stochastic payoffs are independent of each other, so that the operation $*$ is given by

$$\begin{pmatrix} m_1 \\ \sigma_1 \end{pmatrix} * \begin{pmatrix} m_2 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} m_1 + m_2 \\ \sqrt{\sigma_1^2 + \sigma_2^2} \end{pmatrix}. \quad (3.9)$$

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

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

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

A rule Φ is decomposable if

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

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

The only requirement that we impose on $*$ is the existence of an irrelevant outcome: there exists $e \in \mathcal{O}$ such that $e * x = x * e = x$ for any $x \in \mathcal{O}$. The decision-maker does not care about an action a having outcome e in the sense that combining a with any other action b with outcome x does not change the decision-maker's perception of b . The operation $*$ defined in (3.9) admits the irrelevant outcome $(0, 0)$. It is also commutative and associative, but in general we assume neither commutativity nor associativity.

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

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

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

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

$$u(x * y) = u(x) + u(y) \quad \text{for all } x, y \in \mathcal{O}. \quad (3.10)$$

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

For example, when $(O, *) = (\mathbb{R}^n, +)$, any linear function $u(x) = \beta \cdot x$ is a utility representation. Moreover, all continuous utility representations take this form, which follows from a fact about the Cauchy functional equation $u(x) + u(y) = u(x + y)$.¹¹ Monotonicity then pins down the signs of the coefficients β_i : under the standard partial order, it requires $\beta_i \geq 0$. Similarly, when $(O, *)$ is the Gaussian outcome space (3.9), the function $u(m, \sigma) = \beta m + \gamma \sigma^2$ is a utility representation. This functional form again exhausts all continuous representations, and the natural monotonicity requirement, which reflects that lower-variance outcomes are chosen more frequently, yields $\gamma \leq 0$.

Our usage of the term utility representation for functions $u: O \rightarrow \mathbb{R}$ satisfying (3.10) can be motivated by the well-established connection between cardinal utility values and separability: since we think of $x * y$ as a separable combination of outcomes x and y , the identity (3.10) is the respective separability condition on utility u . This interpretation relates our analysis to the characterization of separable utility by Debreu, 1959, who shows that choice independence across dimensions pins down a separable utility (uniquely up to affine transformations).

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

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

$$\Phi(a \mid A, o) = \frac{\exp(u(o(a)))}{\sum_{b \in A} \exp(u(o(b)))} \quad (3.11)$$

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

Informally, Theorem 2 says that there is a canonical way to assign utilities to elements of the outcome space so that choices are governed by a multinomial logit

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

rule with respect to these utilities; moreover, summing these utilities corresponds to combining outcomes in a separable way. The logit scale parameter is normalized to one, as any positive scale factor can always be absorbed by u .

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

3.7 Proof of Theorem 1 for $O = \mathbb{R}$

The remainder of this paper is a proof of our main theorem, which we hope is of technical interest in its own right. We prove the case of $O = \mathbb{R}$ here and extend to $\mathbb{R}^n$ in Appendix 3.C. We first describe a classical result of De Finetti regarding partially exchangeable processes. Then, using IID, we extend our stochastic choice rules to infinite products of menus, to which we apply De Finetti's Theorem. Monotonicity ensures the partial exchangeability of a process defined by the extended choice rule. The mixture of i.i.d. conclusion of De Finetti's theorem then implies that the choice rule is a mixture of multinomial logit rules.

De Finetti's Theorem for Partially Exchangeable Processes

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

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

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

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

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

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

Extending Φ to Infinite Products

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

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

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

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

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

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

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

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

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

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

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

Finally, suppose that $\mathbb{N}$ is partitioned as $\mathbb{N} = N_0 \cup N_1 \cup N_2 \cup \dots$, and that for each block N_k , $(A_{i_1}, o_{i_1}) = (A_{i_2}, o_{i_2}) = \dots$ for indices $i_1, i_2, \dots$ from N_k . We want to show that the law of $(X_1, X_2, \dots)$ is invariant under every finite permutation that maps each block N_k into itself. To this end, it suffices to show, without loss of generality, that if we permute X_{i_1} and X_{i_2} then the joint distribution of any long enough prefix $(X_1, X_2, \dots, X_n)$ of the entire sequence remains unchanged. This follows from the monotonicity of Φ , since the joint distribution of $(X_1, X_2, \dots, X_n)$ is given by $\Phi(\otimes_{i=1}^n (A_i, o_i))$, and by monotonicity, this distribution assigns equal probabilities to sequences yielding the same total payoff, and payoffs are preserved by permuting copies of the same menu, such as X_{i_1} and X_{i_2} , as long as n is larger than both i_1 and i_2 . $\square$

Random Coefficients Logit from De Finetti

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

conditioned on the tail we have that $(X_1, X_2, \dots)$ are independent, with Y_i chosen from F , and Z_i chosen from G , so that

$$F(b_k) = \mathbb{P}[Y_i = b_k \mid \mathcal{T}] \quad \text{and} \quad G(b_0) = \mathbb{P}[Z_i = b_0 \mid \mathcal{T}], \quad \text{for any } i. \quad (3.12)$$

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

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

Denote

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

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

Choose any $a, a' \in B_k$ such that $u_k(a) \geq u_k(a')$ and let $d = u_k(a) - u_k(a')$. By (3.12),

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

Since Y_i and Z_i are independent conditioned on the tail,

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

and so by the law of total expectation

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

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

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

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

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

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

$$\mathbb{E}[F(a) \cdot G(\ell)^d \cdot P] = \mathbb{E}[F(a') \cdot G(h)^d \cdot P]. \quad (3.13)$$

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

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

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

By the linearity of expectation, we have that (3.13) holds for any polynomial P . Thus, taking $P = F(a)G(\ell)^d - F(a')G(h)^d$, we have that $\mathbb{E}[P^2] = 0$, so that

$$F(a)G(\ell)^d = F(a')G(h)^d \text{ almost surely.} \quad (3.14)$$

Note that if $d = 0$ then this proof yields that $F(a) = F(a')$ almost surely. Otherwise, $d = u_k(a) - u_k(a') > 0$. Let E_h be the event that $G(h) = 0$, which is the event that $\beta = -\infty$. It follows from (3.14) that $F(a) = 0$ on E_h . Since a and a' are an arbitrary pair such that $u_k(a) > u_k(a')$, we have $F(b) = 0$ for any b that does not yield lowest payoff. Likewise, we have $F(c) = 0$ for any c that does not yield the highest payoff conditioned on the event E_ℓ where $G(\ell) = 0$ and $\beta = +\infty$. We thus have that conditioned on $\beta = +\infty$, $F = \text{MNL}^{+\infty}(B_k, u_k)$, and likewise conditioned on $\beta = -\infty$, $F = \text{MNL}^{-\infty}(B_k, u_k)$.

Outside the union of E_h and E_ℓ , β is finite, and it follows from (3.14) that

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

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

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

By Lemma 1 from Appendix 3.A, monotonicity implies that μ is supported on $[0, \infty]$.

□

Final Steps

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

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

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

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

Hence, by dominated convergence,

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

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

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

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

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

$$\Phi^k(A, o) = \Phi\left(A, \frac{1}{k} \cdot o\right). \quad (3.15)$$

Note that such rules satisfy monotonicity and IID, since Φ does. Hence, by Proposition 4 again, each $\Phi^k|_{\mathcal{M}_{\mathbb{Z}}} = RCL^{\mu_k}|_{\mathcal{M}_{\mathbb{Z}}}$ for some μ_k . By (3.15) and Proposition 5, it follows that μ is equal to the push-forward of μ_k under the map $\beta \rightarrow k\beta$.

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

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

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

$\lim_n \Phi(a \mid A, o_n) = \frac{1}{2}$. If, however, $\mu(\{+\infty\}) = \varepsilon > 0$, then $\lim_n \Phi(a \mid A, o_n) \leq \frac{1}{2}(1 - \varepsilon)$, violating continuity.

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

3.8 Conclusion

This paper explores a novel approach to stochastic choice. We model the behavior of an individual across a rich variety of situations, under the key assumption that choices remain consistent when decisions are combined in a separable way. Though often implicit, this assumption underpins the validity of experimental analyses that focus on isolated decision problems. In this paper, the corresponding IID axiom is taken as an assumption. But this assumption can be easily tested in a lab, by providing subjects with different menus yielding carefully selected monetary payoffs, and seeing whether their choice probabilities satisfy (3.2), the defining property of IID.

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

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

of our axiomatic approach can deliver a (possibly approximate) characterization on a smaller domain remains a natural question for future work.

3.A Monotonicity of Random Coefficients Logit

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

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

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

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

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

□

3.B Proof of Proposition 1

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

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

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

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

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

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

We can rewrite $f(\beta)$ as

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

and

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

By the Cauchy-Schwarz inequality, we have

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

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

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

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

and

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

It follows that

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

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

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

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

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

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

□

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

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

Since Λ_t satisfies monotonicity, continuity and IID, the one-dimensional case proved in §3.7 gives $\Lambda_t = \text{RCL}^{\nu_t}$ for some ν_t supported on $\mathbb{R}_{\geq 0}$.

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

$$(B_t, u_t) = \left\{ \begin{array}{cc} \ell_t & h_t \\ 0 & e_t \end{array} \right\}. \quad (3.16)$$

Since $(B_{n+1}, u_{n+1}), (B_{n+2}, u_{n+2}), \dots$ is an enumeration, the menus (3.16) will appear twice in $(B_1, u_1), (B_2, u_2), \dots$

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

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

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

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

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

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

Hence Y_i corresponds to the choice in the i th copy of B_k and Z_i^t to the i th copy of B_t , as defined in (3.16).

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

$$F(b_k) = \mathbb{P}[Y_i = b_k \mid \mathcal{T}] \text{ and } G_t(b_t) = \mathbb{P}[Z_i^t = b_t \mid \mathcal{T}], \text{ for any } i. \quad (3.17)$$

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

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

Denote by β the n -dimensional random variable with coordinates

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

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

We claim that $\Phi(B_k, u_k)$ is equal to $\text{RCL}^\mu(B_k, u_k)$. To this end, choose any $a, a' \in B_k$, let d_t denote the t -th coordinate of $u_k(a) - u_k(a')$, and let $d_t^+ = \max\{d_t, 0\}$ and $d_t^- = \max\{-d_t, 0\}$. By (3.17), Y_i and $Z_i^1, \dots, Z_i^n$ are independent conditioned on the tail. By the same argument used in the proof of Proposition 4 we conclude that

$$\begin{aligned} & F(a) \cdot G_1(\ell_1)^{d_1^+} \cdots G_n(\ell_n)^{d_n^+} \cdot G_1(h_1)^{d_1^-} \cdots G_n(h_n)^{d_n^-} \\ &= F(a') \cdot G_1(\ell_1)^{d_1^-} \cdots G_n(\ell_n)^{d_n^-} \cdot G_1(h_1)^{d_1^+} \cdots G_n(h_n)^{d_n^+} \text{ almost surely.} \end{aligned} \quad (3.18)$$

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

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

and we may conclude that $\Lambda_t|_{\mathcal{M}_{\mathbb{Z}}} = \text{RCL}^{\mu_t}|_{\mathcal{M}_{\mathbb{Z}}}$. Thus, by Proposition 5, $\mu_t = \nu_t$, so the support of μ_t is contained in $\mathbb{R}_{\geq 0}$. Thus, β is finite (and F is non-degenerate) almost surely. For any (B_k, u_k) it follows from (3.18) that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.D Proof of Proposition 2

Proof. Since decomposability implies IID, by Theorem 1, $\Phi = \text{RCL}^\mu$ for some μ supported on $\mathbb{R}_{\geq 0}^n$. For $k = 1, \dots, n$, define the stochastic choice rule Φ^k for the outcome space $\mathcal{O} = \mathbb{R}$ by $\Phi^k(a \mid A, o) = \Phi(a \mid A, o_k)$, where $o_k(a) = o(a) \cdot e_k$ where e_k is the standard basis vector whose k th component is 1 and whose other components are 0. It follows that

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

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

3.E Proof of Theorem 2

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

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

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

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

$$u(x) = \ln \frac{p_x}{1 - p_x}. \quad (3.19)$$

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

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

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

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

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

Computing their outcomes, we get

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

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

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

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

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

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

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

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

We now consider an arbitrary menu (A, o) and demonstrate that $\Phi(A, o)$ is given by multinomial logit with the constructed utility function (3.11). Let $a, b \in A$ be two distinct actions. We express the ratio $\Phi(a \mid A, o)/\Phi(b \mid A, o)$ in terms of the constructed utility representation u . Denote the outcomes by $x = o(a)$ and $y = o(b)$ and consider a new menu

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

The two actions

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

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

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

This equality can be rewritten as

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

and thus

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

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

$$\Phi(a \mid A, o) = \frac{\exp(u(o(a)))}{\sum_{b \in A} \exp(u(o(b)))}, \quad (3.20)$$

i.e., Φ is multinomial logit.

To verify that u is monotone, take $x \preceq y$ and consider the binary menu with outcomes x and y . By monotonicity of Φ , the action with outcome y is chosen with probability at least as large as the action with outcome x . Using the representation (3.20), this implies $\exp(u(y)) \geq \exp(u(x))$, and hence $u(y) \geq u(x)$.

It remains to check the equivalence between the continuity of Φ and that of u . If $u: \mathcal{O} \rightarrow \mathbb{R}$ is a continuous utility representation, then Φ given by (3.20) is continuous since the right-hand side of (3.20) is a continuous function of the profile of outcomes. In the opposite direction, we suppose that Φ is continuous and show that $u(x_n) \rightarrow u(x)$ for any sequence $x_n \rightarrow x$ in $\mathcal{O}$. Fix a binary set of actions $A = \{a, b\}$ and consider outcome functions o_n and o given by $o_n(a) = o(a) = x$, $o_n(b) = x_n$, and $o(b) = x$. Thus (A, o_n) converges to the menu with identical outcomes (A, o) . By continuity, $\Phi(A, o_n)$ converges to the uniform distribution $\Phi(A, o)$. By (3.20), we get

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

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

□

3.F Restrictions of Monotonicity and IID

The family of random coefficients logit rules is parameterized by a probability measure, which is an infinite dimensional object. This gives this family considerable flexibility to model a wide range of behavior. Nevertheless, knowing that a rule belongs to this family imposes significant restrictions on the choice probabilities, even within simple menus. As we discussed in §3.2, for the menu

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

it holds for every random coefficients logit rule Φ that $\Phi(a_1 \mid A, o) = 5\%$ implies that $\Phi(a_3 \mid A, o) \geq 83.8\%$. Indeed, this lower bound of the choice probability of a_3 turns out to be $\text{MNL}^\beta(a_3 \mid A, o)$ for the β that satisfies $\text{MNL}^\beta(a_1 \mid A, o) = 5\%$. The

following proposition shows that this is a general property of random coefficients logit rules for $\mathcal{O} = \mathbb{R}$.

Proposition 6. *Let $A = \{a_1, \dots, a_n\}$ and $o(a_1) \leq o(a_2) \leq \dots \leq o(a_n)$ with at least one strict inequality. Let $\beta \in \mathbb{R}$ be the unique parameter satisfying $\text{MNL}^\beta(a_1 \mid A, o) = \Phi(a_1 \mid A, o)$. If Φ is a random coefficients logit rule, then $\Phi(a_n \mid A, o) \geq \text{MNL}^\beta(a_n \mid A, o)$.*

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

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

Before proving this proposition we state and prove a lemma.

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

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

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

The derivatives of L are given by:

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

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

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

Thus, we have:

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

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

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

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

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

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

By Lemma 4,

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

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

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

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

BIBLIOGRAPHY

- Ahn, David S., Federico Echenique, and Kota Saito (2018). “On path independent stochastic choice.” In: *Theoretical Economics* 13.1, pp. 61–85.
- Berry, Steven, James Levinsohn, and Ariel Pakes (1995). “Automobile Prices in Market Equilibrium.” In: *Econometrica* 63.4, pp. 841–890.
- Block, Henry David and Jacob Marschak (1959). “Random orderings and stochastic theories of response.” In.
- Breitmoser, Yves (2021). “An axiomatic foundation of conditional logit.” In: *Economic Theory* 72.1, pp. 245–261.
- Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini (2021). “Axiomatic tests for the Boltzmann distribution.” In: *Journal of Statistical Mechanics: Theory and Experiment* 2021.1, p. 013406.
- (Aug. 2022). “Multinomial Logit Processes and Preference Discovery: Inside and Outside the Black Box.” In: *The Review of Economic Studies*. ISSN: 0034-6527.
- Chambers, Christopher P., Yusufcan Masatlioglu, and Christopher Turansick (2021). “Correlated choice.” In: *arXiv preprint arXiv:2103.05084*.
- Cho, Wonki Jo (2022). “How to add apples and oranges: Aggregating performances of different nature.” In: *Games and Economic Behavior* 131, pp. 222–244.
- Clark, Stephen A. (1996). “The random utility model with an infinite choice space.” In: *Economic Theory* 7, pp. 179–189.
- De Finetti, Bruno (1980). “On the condition of partial exchangeability.” In: *Studies in inductive logic and probability* 2, pp. 193–205.
- Debreu, Gerard (1959). “Topological methods in cardinal utility theory.” In: *Cowles Foundation Discussion Papers* 299.
- (1960). “Review of individual choice behavior: a theoretical analysis, by RD Luce.” In: *American Economic Review* 50.1, pp. 186–188.
- Echenique, Federico, Kota Saito, and Gerelt Tserenjigmid (2018). “The perception-adjusted Luce model.” In: *Mathematical Social Sciences* 93, pp. 67–76.
- Falmagne, Jean-Claude (1978). “A representation theorem for finite random scale systems.” In: *Journal of Mathematical Psychology* 18.1, pp. 52–72.
- Fishburn, Peter C. (1965). “Independence in utility theory with whole product sets.” In: *Operations research* 13.1, pp. 28–45.
- Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari (2012). “The random coefficients logit model is identified.” In: *Journal of Econometrics* 166.2, pp. 204–212.

- Frick, Mira, Ryota Iijima, and Tomasz Strzalecki (2019). “Dynamic random utility.” In: *Econometrica* 87.6, pp. 1941–2002.
- Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack (2025). *Limited Memory, Learning, and Stochastic Choice*. Tech. rep. Working paper.
- Fudenberg, Drew and Tomasz Strzalecki (2015). “Dynamic logit with choice aversion.” In: *Econometrica* 83.2, pp. 651–691.
- Gandhi, Amit and Aviv Nevo (2021). “Empirical models of demand and supply in differentiated products industries.” In: *Handbook of industrial organization*. Vol. 4. 1. Elsevier, pp. 63–139.
- Gul, Faruk, Paulo Natenzon, and Wolfgang Pesendorfer (2014). “Random choice as behavioral optimization.” In: *Econometrica* 82.5, pp. 1873–1912.
- Gul, Faruk and Wolfgang Pesendorfer (2006). “Random expected utility.” In: *Econometrica* 74.1, pp. 121–146.
- Hardy, G. H., J. E. Littlewood, and G. Pólya (1952). *Inequalities*. 2nd. Cambridge University Press.
- Kashaev, Nail, Charles Gauthier, and Victor H. Aguiar (2023). “Dynamic and Stochastic Rational Behavior.” In: *arXiv preprint arXiv:2302.04417*.
- Kashaev, Nail, Martin Plávala, and Victor H. Aguiar (2024). “Entangled vs. Separable Choice.” In: *arXiv preprint arXiv:2403.09045*.
- Kovach, Matthew and Gerelt Tserenjigmid (2022). “Behavioral foundations of nested stochastic choice and nested logit.” In: *Journal of Political Economy* 130.9, pp. 2411–2461.
- Kuczma, Marek (2009). “An introduction to the theory of functional equations and inequalities.” In: Birkhäuser Basel.
- Li, Ricky (2021). “Dynamic Random Choice.” In: *arXiv preprint arXiv:2102.00143*.
- Lu, Jay and Kota Saito (2025). “Mixed Logit Versus Pure Characteristics Models.” In: .
- Luce, R. Duncan (1959). *Individual choice behavior*. Vol. 4. Wiley New York.
- Luce, R. Duncan and Patrick Suppes (1965). “Preferences, utility and subjective probability.” In: *Handbook of mathematical psychology*, pp. 249–410.
- Manski, Charles F. (1977). “The structure of random utility models.” In: *Theory and decision* 8.3, p. 229.
- Matějka, Filip and Alisdair McKay (2015). “Rational inattention to discrete choices: A new foundation for the multinomial logit model.” In: *American Economic Review* 105.1, pp. 272–298.
- Mattsson, Lars-Göran and Jörgen W. Weibull (2002). “Probabilistic choice and procedurally bounded rationality.” In: *Games and Economic Behavior* 41.1, pp. 61–78.

- McFadden, Daniel (1974). “Conditional logit analysis of qualitative choice behavior.” In: *Frontiers in Econometrics*, pp. 105–142.
- (1981). “Econometric Models of Probabilistic Choice.” In: *Structural Analysis of Discrete Data with Econometric Applications*. Ed. by Charles F. Manski and Daniel McFadden. Cambridge, MA: The MIT Press, pp. 198–272.
- McFadden, Daniel and Marcel K. Richter (1990). “Stochastic rationality and revealed stochastic preference.” In: *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, Westview Press: Boulder, CO, pp. 161–186.
- McFadden, Daniel and Kenneth E. Train (2000). “Mixed MNL models for discrete response.” In: *Journal of applied Econometrics* 15.5, pp. 447–470.
- Moon, Hyungsik Roger, Matthew Shum, and Martin Weidner (2018). “Estimation of random coefficients logit demand models with interactive fixed effects.” In: *Journal of Econometrics* 206.2, pp. 613–644.
- Nevo, Aviv (2000). “A practitioner’s guide to estimation of random-coefficients logit models of demand.” In: *Journal of economics & management strategy* 9.4, pp. 513–548.
- Pratt, John W. (1964). “Risk aversion in the small and in the large.” In: *Econometrica* 32.1/2, pp. 122–136.
- Saito, Kota (2018). “Axiomatizations of the mixed logit model.” In.
- Sandomirskiy, Fedor, Po Hyun Sung, Omer Tamuz, and Ben Wincelberg (2025). “Monotonicity and bracketing in games.” In: *arXiv preprint arXiv:2502.11243*.
- Segal, Uzi (1992). “Additively separable representations on non-convex sets.” In: *Journal of Economic Theory* 56.1, pp. 89–99.
- Sims, Christopher A. (2003). “Implications of rational inattention.” In: *Journal of monetary Economics* 50.3, pp. 665–690.
- Steiner, Jakub, Colin Stewart, and Filip Matějka (2017). “Rational inattention dynamics: Inertia and delay in decision-making.” In: *Econometrica* 85.2, pp. 521–553.
- Strzalecki, Tomasz (2025). “Stochastic choice theory.” In: *Cambridge Books*.
- Train, Kenneth E. (2003). “Mixed Logit.” In: *Discrete Choice Methods with Simulation*. Cambridge University Press, pp. 138–154.
- Wakker, Peter P. (2013). *Additive representations of preferences: A new foundation of decision analysis*. Springer Science & Business Media.
- Woodford, Michael (2014). “Stochastic choice: An optimizing neuroeconomic model.” In: *American Economic Review* 104.5, pp. 495–500.
- Yellott, John I. (1977). “The relationship between Luce’s choice axiom, Thurstone’s theory of comparative judgment, and the double exponential distribution.” In: *Journal of Mathematical Psychology* 15.2, pp. 109–144.