# Discrete Random Expected Utility

Igor Kopylov[*] Kevin Peng[†‡]

November 2, 2025

**JOB MARKET PAPER**
[click here for latest version]

**SELECT SLIDES FOR ASSOCIATED EXPERIMENTAL PROJECT INCLUDED IN APPENDIX B.** (Peng, 2025)

**Abstract**

We model stochastic choice rules via finitely many types $\theta$ that maximize distinct expected utility functions and use endogenous tie-breaking rules. In general, the likelihoods $\mu_A(\theta)$ of such types are allowed to depend on the feasible menu $A$, and we also derive representations where the distribution $\mu(\theta)$ is unaffected by $A$. This invariant case provides a discrete version for the *random expected utility* of Gul and Pesendorfer (2006), but we use distinct axioms and identification methods. More generally, we study representations where the menu-dependent type distribution $\mu_A(\theta)$ accommodates various kinds of *context dependence*. In particular, we show that the standard monotonicity principle imposed on stochastic choice data can be used to characterize *self-selection* in type likelihoods. In other words, type $\theta$ should not become more likely when new alternatives are added to a feasible menu, but do not improve the best choice for $\theta$. Both the discrete type space $\Theta$ and the bivariate function $\mu$ can be identified uniquely in our model. Finally, we discuss applications to heterogeneous risk attitudes and beliefs, and to filtering measurement noise.

[*]Department of Economics, University of California, Irvine. E-mail: ikopylov@uci.edu.
[†]Department of Economics, University of California, Irvine. E-mail: kuangyup@uci.edu
[‡]We are grateful to John Duffy, Stergios Skaperdas, Jay Lu, Yusufcan Masatligolu, and Chris Chambers for their comments.

# 1 Introduction

Observed choices in consumption menus can be naturally *stochastic* when they are produced by distinct rational (i.e., utility maximizing) agents. Data is often presented as an aggregation of choices from such distinct agents. It can be seen in a wide range of empirical work, such as preferred modes of transportation (McFadden, 2001) and choice of fishing sites (Train, 1998). This heterogeneity may also result from more behavioral motivations where a single individual can exhibit distinct random responses in identical choice problems (see Agranov and Ortoleva (2017)).

The classic random utility model (RUM) by Block and Marschak (1960) aims to represent stochastic choice data by a stable probability distribution $\mu$ across *types* that include all possible utility functions. In the general case where types are unrestricted, Falmagne (1978) characterizes RUM via non-negativity of Block and Marschak's polynomials. Gul and Pesendorfer (2006) (henceforth, GP) refine this model to *random expected utility* (REU) where all types should maximize expected utility functions rather than arbitrary ones. Other refinements of RUM have restricted types by single-crossing conditions (Apesteguia, Ballester, and Lu, 2017), quasi-linearity (McFadden, 1973; Williams, 1977; Daly and Zachery, 1979; Yang and Kopylov, 2023), and other assumptions. Behavioral patterns, such as ambiguity aversion (Lu, 2021), have also been discussed within the context of random utility. All of these refinements preserve the assumption that $\mu$ is invariant of the consumption menu.

In this paper, we propose a novel version of random utility that has

- a discrete (i.e., finite) type space $\Theta$ where all types $\theta \in \Theta$ combine expected utility maximization with endogenous tie-breaking rules;
- a distribution of types $\mu_A$ that can depend on the menu $A$ and capture context dependence and self-selection;
- a unique identification for both $\Theta$ and $\mu_A$.

We still consider a representation where the distribution $\mu(\theta)$ is invariant of $A$. This case is a discrete version of the GP's REU (henceforth, DREU), but our axioms and identifications are more parsimonious than theirs and rely more on algebraic rather than analytical techniques. Moreover, the algebraic approach makes the consumption domain $X$ more flexible than in GP, where $X$ is a finite-dimensional simplex of lotteries. In our framework, $X$ is convex, but need not be closed, or bounded, or finite dimensional. For example, $X$ may consist of monetary gambles with payoffs on the real line, or $X$ can be the non-negative orthant of consumption bundles.

More substantially, we depart from REU by allowing the distribution $\mu_A(\theta)$ to vary with the menu $A$. In general, it is only required that the support $\Theta = \{\theta : \mu_A(\theta) > 0\}$ be finite and invariant of $A$. This $\Theta$ is called a discrete type space. It is used to represent observable likelihoods $\rho(x, A)$ of choosing any element $x$ in any finite consumption menu $A$. Our most general representation takes the form

$$\rho(x, A) = \mu_A\{\theta \in \Theta : x \text{ is the best element for } \theta \text{ in } A\} \tag{1}$$

where $\mu_A(\theta)$ is the probability of type $\theta$ making a choice in the menu $A$.

Each type $\theta$ is assumed to maximize some expected utility function $u_\theta$ in each menu $A$. However, in order for representation (1) to be well-defined, each type $\theta \in \Theta$ must also incorporate some endogenous tie-breaking rule that selects a unique maximizer for $u_\theta$ in $A$. Formally, each $\theta$ is associated with a total, complete, transitive binary relation that satisfies von Neumann-Morgenstern's Independence axiom as well. Such total types have been used in finite-dimensional contexts by Fishburn (1982), Myerson (1986), Blume, Branderburger, and Dekel (1991) and others. We argue that there is a rich family of total types even if $X$ is not finite dimensional. Thus we provide a novel way to accommodate tie-breaking in random utility models. Without totality, Piermont (2022) relaxes GP's model to represent only those choices that are strict maximizers for corresponding types.

We show several characterization results for stochastic choice rules. First, Theorem 2 below characterizes representation where $\mu_A = \mu$ is menu-invariant and relates this result to GP's model. Our main result (Theorem 3 below) characterizes representation (1) and provides a convenient platform for various refinements that impose more structure on the menu-contingent distribution $\mu_A$ and/or types $\theta$. We obtain several such refinements.

Theorem 4 summarizes the implications of the standard monotonicity principle that asserts

$$\rho(x, A) \geq \rho(x, A \cup B)$$

for all menus $A, B \in \mathcal{M}$ and elements $x \in A$. It turns out to be equivalent for the function $\mu_A$ to satisfy

$$\mu_A(\theta) \geq \mu_{A \cup B}(\theta)$$

for all types $\theta \in \Theta$ that have the same maximal element in $A$ and in $A \cup B$. This finding can be interpreted in terms of *self-selection*: types can increase their participation in the menu $A \cup B$, but only if they find better choices in the larger menu. For example, it should be reasonable for gamblers to show up in higher numbers to bookmakers that offer them more favorable odds on some sporting event. Even the expansion of bet types with the rise of online gambling can incentivize new customers to enter the market (Hing et al., 2022). This can also be observed in the dining industry. Garnett et al. (2019) found that doubling the options of vegetarian meals increased sales of such meals by around 15%.

By contrast, self-selection cannot be detected in the general RUM examples because it can keep all Block-Marschak polynomials non-negative. In that example, the presence of self-selection makes the general RUM *misspecified* without rejecting this model outright. Our DREU model does not have this identification issue and disentangles self-selection from the composition of the type space $\Theta$.

More broadly, there are some behavioral patterns, such as reasoned-based heuristics in Shafir, Simonson, and Tversky (1993), where Monotonicity does not hold and hence, self-selection is no longer a plausible explanation. We use *context dependence* as a blanket term to refer to all such patterns (see a literature review by Rabin (1998)). An example of such patterns is extremeness aversion. Sharpe et al. (2008) observed that the removal of the largest drink size (44-oz) from the menu prompted consumers to switch from the 32-oz drink size (which is the largest size in the new menu) to 21 ounces.

Finally, we discuss applications where the type space $\Theta$ captures heterogeneous *risk attitudes* and *beliefs*. For example, agents make more risk neutral choices when lotteries are presented in a binary format (i.e., only two states of the world have positive probabilities) instead of lotteries that are more complex or compound lotteries (Harrison et al., 2013). Changes in beliefs arise in the form of partition dependence. Ahn and Ergin (2010) observed that agents, when choosing deductible plans, place a higher weight on surgeries when they are individually listed than when they are bundled under the umbrella term "surgery". In such situations, the consumption domain $X$ can be taken consists of lotteries or Anscombe-Aumann (1963) (henceforth, AA) uncertain prospects called acts.

# 2 Preliminaries

Let $X = \{x, y, z, \dots\}$ be a consumption domain. Assume that $X$ is a non-singleton convex subset of a normed linear space $L$. For example, $X$ can be a simplex

$$\Delta^n = \left\{ x \in \mathbb{R}^n_+ : \sum_{i=1}^{n} x_i = 1 \right\}$$

of all *lotteries* on a finite set of deterministic prizes $\{1, ..., n\}$. Then $X$ is a convex subset of $L = \mathbb{R}^n$

Let $\mathcal{U}$ be the set of all non-constant functions $u : X \to \mathbb{R}$ such that for all $x, y \in X$ and

$\gamma \in [0, 1]$,

$$u(\gamma x + (1 - \gamma)y) = \gamma u(x) + (1 - \gamma)u(y). \tag{2}$$

If $X = \Delta^n$ is a simplex of lotteries, then $\mathcal{U}$ consists of all expected utility functions on $X$. In general, $L$ may be infinite dimensional, and functions $u \in \mathcal{U}$ may have interpretations other than expected utility.

A binary relation $\theta$ on $X$ is called a *type* if it is complete, transitive, and satisfies Independence: for all $x, y, z \in X$ and $\gamma \in (0, 1]$,

$$x\theta y \iff [\gamma x + (1 - \gamma)z]\theta[\gamma y + (1 - \gamma)z].$$

In our representations below, types $\theta$ are interpreted as heterogeneous preferences that generate observable choices. A type $\theta$ is called

- *null* if $x\theta y$ for all $x, y \in X$; otherwise, $\theta$ is *non-null*;
- *total* if for all $x, y \in X$, $x\theta y\theta x$ implies $x = y$;
- *mixture continuous* if

  $$\{\gamma \in [0, 1] : z\theta[\gamma x + (1 - \gamma)y]\} \text{ and } \{\gamma \in [0, 1] : [\gamma x + (1 - \gamma)y]\theta z\}$$

  are closed sets for all $x, y, z \in X$. Otherwise, $\theta$ is *discontinuous*.

Let $\mathcal{T}$ be the set of all total types.

Herstein and Milnor's (1953) mixture space theorem asserts that a type $\theta$ is mixture continuous if and only if $\theta$ can be represented by some function $u \in \mathcal{U}$: for all $x, y \in X$,

$$x\theta y \iff u(x) \geq u(y).$$

Moreover, such $u$ is unique up to a positive linear transformation: if $\theta$ is represented by another $u' \in \mathcal{U}$, then $u' = \alpha u + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$.

However, this result cannot be applied directly to total types because they are usually discontinuous. Indeed, let $X$ be a convex subset of a line $\{x + \alpha d : \alpha \in \mathbb{R}\}$ for some $x, d \in L$ such that $d \neq 0$. Call such $X$ *collinear*. Then there are only two possible types $\theta_+$ and $\theta_-$ that are represented by $u(x + \alpha d) = \alpha$ and $u(x + \alpha d) = -\alpha$ respectively. Both are total and mixture continuous.

In the generic *non-collinear* case, $X$ is not a subset of any line, and all total types are discontinuous. To show this, take $x, y, z \in X$ such that $x - y$ and $y - z$ are linearly independent. Let $\theta$ be mixture continuous. Then it is represented by some $u \in \mathcal{U}$. Without loss in generality, $u(x) \geq u(y) \geq u(z)$. Then either $u(x) = u(z)$, or $u(y) = u(\gamma x + (1-\gamma)z)$ where $\gamma = \frac{u(y)-u(z)}{u(x)-u(z)}$. As $y \neq \gamma x + (1-\gamma)z$, then either $x \sim_\theta z$ or $y \sim_\theta \gamma x + (1-\gamma)z$ implies that $\theta$ is not total.

For any discontinuous type $\theta$, the mixture space theorem can be still applied to its *closure* $\theta^*$. Define $\theta^*$ as a binary relation on $X$ such that for any $x, y \in X$, $x\theta^* y$ if and only if there are $z_x, x_y$ such that for all $\gamma \in [0,1]$,

$$[\gamma x + (1-\gamma)z_x]\theta[\gamma y + (1-\gamma)z_y]. \tag{3}$$

Obviously, if $x\theta y$, then $x\theta^* y$ as well because (3) holds for $z_x = x$ and $z_y = y$.

**Theorem 1.** *For any type $\theta$, its closure $\theta^*$ is a mixture continuous type, and $\theta^*$ is represented by $u \in \mathcal{U}$ if and only if for all $x, y \in Z$,*

$$x\theta y \implies u(x) \geq u(y). \tag{4}$$

*Moreover, if $\theta^*$ is null, and $X$ is finite dimensional, then $\theta$ is null.*

The proof is in the appendix. Together with the mixture space theorem, this result delivers two possible cases for any non-null type $\theta$. If its closure $\theta^*$ is non-null, then the one-way representation (4) holds for some $u \in \mathcal{U}$ that is unique up to a positive linear transformation. If $\theta^*$ is null, then (4) does not hold for any $u \in \mathcal{U}$. The latter case never happens when $X$ is finite dimensional. Note that (4) is trivial if $u$ is constant, but such $u$ does not belong $\mathcal{U}$.

Representation (4) can be refined further by combining the maximization of $u \in \mathcal{U}$ with a tie-breaking rule that is used when $u(x) = u(y)$. This refinement can be also used to construct total types explicitly. We postpone this discussion until Section 4.

## 2.1 Menus and Discrete Distributions

Let $\mathcal{M} = \{A, B, C \dots\}$ be the set of all *menus*—non-empty finite subsets of $X$. Singletons $\{x\}$ are written as $x$.

For any menu $A \in \mathcal{M}$ and total type $\theta \in \mathcal{T}$, there is a unique element $x \in A$ such that $x\theta y$ for all $y \in A$. Say that $x$ is *best* for $\theta$ in $A$. If $\theta$ is represented by (4), then its best element $x$ must maximize $u$ in $A$, albeit it need not be the unique such maximizer.

For any $A \in \mathcal{M}$ and $x \in A$,

$$\mathcal{T}(x, A) = \{\theta \in \mathcal{T} : x \text{ is best for } \theta \text{ in } A\}.$$

Then the entire $\mathcal{T}$ can be partitioned into the disjoint union

$$\mathcal{T} = \bigcup_{x \in A} \mathcal{T}(x, A). \tag{5}$$

For any $A, B \in \mathcal{M}$ and $\gamma \in [0,1]$, define a mixture

$$\gamma A + (1-\gamma)B = \{\gamma x + (1-\gamma)y : x \in A, y \in B\}.$$

If $\gamma > 0$, then for any $x \in A$, the set $\mathcal{T}(x, A)$ is partitioned into the disjoint union

$$\mathcal{T}(x, A) = \bigcup_{y \in B} \mathcal{T}(\gamma x + (1-\gamma)y, \gamma A + (1-\gamma)B). \tag{6}$$

Indeed, take any total type $t \in \mathcal{T}$. If $x$ is best for $\theta$ in $A$ and $y$ is best for $\theta$ in $B$, then by Independence, $\gamma x + (1-\gamma)y$ is best for $\theta$ in $\gamma A + (1-\gamma)B$. However, if $x$ is not best for $\theta$ in $A$, then by Independence, for any $y \in B$, $\gamma x + (1-\gamma)y$ is not best for $\theta$ in $\gamma A + (1-\gamma)B$. The union in (6) is disjoint because the best element for $\theta$ in $\gamma A + (1-\gamma)B$ is unique.

For any finite set $\Theta \subset \mathcal{T}$, let $\mathcal{D}_\Theta$ be the set of all probability distributions $\mu : 2^{\mathcal{T}} \to \mathbb{R}_+$ that have $\Theta$ as their *support*. In other words,

- for all $\Psi \subset \mathcal{T}$, $\mu(\Psi) = \sum_{\theta \in \Psi \cap \Theta} \mu(\theta)$,
- for all $\theta \in \Theta$, $\mu(\theta) > 0$,
- $\mu(\mathcal{T}) = \sum_{\theta \in \Theta} \mu(\theta) = 1$.

Such distributions are called *discrete*. When $\mu : \mathcal{M} \to \mathcal{D}_\Theta$ specifies discrete distributions contingent on each menu $A$, then these distributions are written as $\mu_A$ rather than $\mu(A)$.

# 3 Main Representation Results

Let $\Omega$ be the set of all pairs $(x, A)$ such that $A \in \mathcal{M}$ and $x \in A$, that is, $x$ is a feasible element in a menu $A$. Such pairs are called *trials*.

A function $\rho : \Omega \to [0,1]$ is called a *stochastic choice rule* (scr) if

$$\sum_{x \in A} \rho(x, A) = 1 \text{ for all } A \in \mathcal{M}. \tag{7}$$

Here, the probability $\rho(x, A)$ of any trial $(x, A) \in \Omega$ can be interpreted as the likelihood of $x$ being chosen when the menu $A$ is feasible.

A stochastic choice rule $\rho$ has a *discrete type space* $\Theta \subset \mathcal{T}$ if $\Theta$ is finite, and there is a function $\mu : \mathcal{M} \to \mathcal{D}_\Theta$ such that

$$\rho(x, A) = \mu_A(\mathcal{T}(x, A)) \text{ for all } (x, A) \in \Omega. \tag{8}$$

is the likelihood that the menu-dependent discrete distribution $\mu_A$ assigns to all total types $\theta$ for which $x$ is best in $A$. Partition (5) implies that (8) satisfies normalization (7). By Theorem 1, total types $\theta \in \mathcal{T}$ can be interpreted in terms of expected utility maximization. Thus we refer to representation (8) as *discrete random expected utility* (DREU).

A stochastic choice rule $\rho$ is represented by a *discrete distribution* $\mu$ if $\mu \in \mathcal{D}_\Theta$ for some finite $\Theta \subset \mathcal{T}$, and

$$\rho(x, A) = \mu(\mathcal{T}(x, A)) \text{ for all } (x, A) \in \Omega. \tag{9}$$

Obviously, (9) is a special case of (8) where the discrete distribution $\mu_A$ is invariant of the menu $A$.

First, we characterize representation (9) and relate it to GP's REU. For any $(x, A) \in \Omega$, $B \in \mathcal{M}$, and $\gamma \in [0, 1]$, let

$$\rho(x, A, \gamma, B) = \sum_{y \in B} \rho(\gamma x + (1-\gamma)y, \gamma A + (1-\gamma)B)$$

be the aggregate likelihood of an element in $\gamma x + (1-\gamma)B$ being chosen when $\gamma A + (1-\gamma)B$ is feasible.

**Axiom 1** (Additivity)**.** *For all $(x, A) \in \Omega$, $B \in \mathcal{M}$, and $\gamma \in (0, 1)$,*

$$\rho(x, A) = \rho(x, A, \gamma, B).$$

This condition is implied by (9) and the partition (6). Note that Additivity is equivalent to the apparently weaker subadditivity condition:

$$\rho(x, A) \geq \rho(x, A, \gamma, B) \tag{10}$$

for all $(x, A) \in \Omega$, $B \in \mathcal{M}$, and $\gamma \in (0, 1)$. Indeed, by definition of stochastic choice rules, subadditivity (10) implies

$$1 = \sum_{x \in A} \rho(x, A) \geq \sum_{x \in A} \sum_{y \in B} \rho(\gamma x + (1-\gamma)y, \gamma A + (1-\gamma)B) \geq 1,$$

and hence, Additivity.

For any menu $A$, let $k_\rho(A) = |\{x \in A : \rho(x, A) > 0\}|$ be the number of elements in $A$ that are chosen with a positive probability.

**Axiom 2** (Discreteness)**.** *There is $A \in \mathcal{M}$ such that for all $B \in \mathcal{M}$,*

$$k_\rho(A) \geq k_\rho(B).$$

Thus the maximal number of elements that are chosen with a positive probability in any menu $B$ must be bounded. As types $t \in \mathcal{T}$ are total, then representation (8) and a fortiori, (9) imply Discreteness because for all $B \in \mathcal{M}$, $k_\rho(B) \leq |\Theta|$. Like continuity, Discreteness cannot be rejected by finite datasets. However, in theoretical frameworks, some natural tie-breaking rules can violate Discreteness when types are not required to be total. For example, if $X$ is non-collinear, then for any $u \in \mathcal{U}$, one can construct a menu $A = \{x_1, \ldots, x_n\}$ of an arbitrary size $n$ so that $u(x_i) = u(x_j)$ for all $i, j$. If such ties are broken by the uniform rule, then $\rho(x_i, A) = \frac{1}{n}$ for each $i$, and Discreteness fails.

**Theorem 2.** *A stochastic choice rule $\rho$ satisfies Additivity and Discreteness if and only if $\rho$ is represented by a discrete type distribution $\mu$. Such $\mu$ is unique.*

The proof of this result is constructive. By Discreteness, there is a maximal finite menu $B = \{b_1, \ldots, b_k\}$ such that $\rho(b_i, B) > 0$ for all $i$. For any $x, y \in X$ and any $i = 1, \ldots, k$, let

$$x\theta_i y \iff \rho(\alpha x + (1-\alpha)b_i, \alpha\{x, y\} + (1-\alpha)B) > 0, \tag{11}$$

for sufficiently small $\alpha > 0$. Additivity delivers (9) where

$$\Theta = \{\theta_1, \ldots, \theta_k\} \subset \mathcal{T}$$

and $\mu(\theta_i) = \rho(b_i, B)$. This argument also implies that the size $|\Theta|$ of the support $\Theta$ can be achieved as the maximal number of elements chosen in some menu. All details are in the appendix.

Representation (9) is similar to GP's REU and satisfies their axioms, except for continuity). However, representation (9) has some distinct features. First, the distribution $\mu$ is taken to be discrete rather than continuous. This change in mathematical structure allows the identification of $\mu$ via algebraic rather than analytical techniques. The algebraic approach makes the consumption domain $X$ more flexible than the simplex $\Delta^n$ required by GP. In our framework, $X$ is convex, but need not be closed, or bounded, or finite dimensional. For example, $X$ may consist of monetary gambles with payoffs on the real line, or

$X$ can be the non-negative orthant of consumption bundles. Moreover, countable additivity for $\mu$ is equivalent to finite additivity in DREU.

Second, we assume all types to be total rather than mixture continuous as in GP. Recall that total types are discontinuous in the generic case when $X$ is non-collinear. The restriction to total types makes tie-breaking endogenous. In REU, tie-breaking is avoided: it is required to have probability zero in each menu because of a suitable regularity property on the distribution $\mu$ over all mixture continuous types.

To reinterpret these differences, for any total type $\theta \in \mathcal{T}$ and trial $(x, A) \in \Omega$, let

$$\rho_\theta(x, A) = \begin{cases} 1 & \text{if } x = \theta(A) \\ 0 & \text{if } x \neq \theta(A). \end{cases}$$

Obviously, the function $\rho_\theta : \Omega \to [0, 1]$ is a stochastic choice rule, and representation (9) asserts that

$$\rho = \sum_{\theta \in \Theta} \mu(\theta)\rho_\theta$$

is a convex combination of $\rho_\theta$. By contrast, there is no counterpart choice function $\rho_\theta$ when the type $\theta$ is mixture continuous are requires tie-breaking.

The Additivity axiom in Theorem 2 can be naturally compared with GP's axioms used in their Theorems 2 and 3. In their framework, $X = \Delta^n$, and representation (9) satisfies all axioms of their Theorem 2 including mixture continuity. By Theorem 2, if Discreteness is imposed in GP's Theorem 2, then their finite additive representations over mixture continuous types can be replaced by discrete distributions over total types. By contrast, continuity that GP require in Theorem 3 must be violated by DREU whenever $X$ is non-collinear.

Additivity implies four conditions that GP assume for all $(x, A) \in \Omega$, $y \in X$, $B \in \mathcal{M}$, and $\gamma \in (0, 1]$.

**Axiom 3** (Monotonicity). $\rho(x, A) \geq \rho(x, A \cup B)$.

**Axiom 4** (Linearity). $\rho(x, A) = \rho(x, A, \gamma, y)$.

**Axiom 5** (Symmetric Extremeness (SE)). $\rho\left(0.5x + 0.5y, \{x, y, 0.5x + 0.5y\}\right) = 0$.

**Axiom 6** (Mixture Continuity (MC)). $\rho(x, A, \gamma, B)$ *is continuous over* $\gamma \in (0, 1]$.

Indeed, assume Additivity. Linearity is a special case where $B = y$ is a singleton. The function $\rho(x, A, \gamma, B)$ is constant and hence, continuous over $\gamma \in (0, 1]$. Monotonicity follows from

$$\rho(x, A) = \rho(x, A, 0.5, A \cup B) \geq \rho(x, A \cup B, 0.5, A) = \rho(x, A \cup B)$$

because $0.5x + 0.5(A \cup B) \supset 0.5x + 0.5A$. To show SE, let $B = \{x, y\}$. Then

$$1 = \rho(x, B) + \rho(y, B) = \rho(x, B, 0.5, B) + \rho(y, B, 0.5, B) = \\ 1 + \rho\left(0.5x + 0.5y, \{x, y, 0.5x + 0.5y\}\right)$$

and hence, $\rho\left(0.5x + 0.5y, \{x, y, 0.5x + 0.5y\}\right) = 0$.

Conversely, Additivity can be derived GP's list, but proving this claim requires a lot of effort and appears tangential to our desiderata. So we skip this proof. Besides simplifying the proofs, Additivity makes the statement of Theorem 2 more parsimonious. It also has a clear connection to the axiom that we use to characterize the general DREU next.

## 3.1 Characterization of DREU

When $\mu : \mathcal{M} \to \mathcal{D}_\Theta$ is menu-contingent, then (8) violates Additivity because it applies distinct distributions $\mu_A$ and $\mu_{\gamma A + (1-\gamma)B}$ to the two sides of the partition (6). However, a weaker form of Additivity still holds.

**Axiom 7** (Boundary Additivity (BA))**.** *For all $A, B \in \mathcal{M}$, $x \in A$ and $\gamma \in (0, 1]$,*

$$\rho(x, A) = 0 \iff \rho(x, A, \gamma, B) = 0.$$

Indeed, by (8) and partition (6),

$$\rho(x, A) = 0 \iff \mathcal{T}(x, A) \cap \Theta = \varnothing \iff \\ \mathcal{T}(\gamma x + (1-\gamma)y, \gamma A + (1-\gamma)B) \cap \Theta = \varnothing \text{ for all } \theta \in \Theta \text{ and } y \in B \iff \rho(x, A, \gamma, B) = 0.$$

In other words, BA requires Additivity only in the boundary case when either side in this equation is zero.

DREU satisfies Discreteness, but it may be desirable to strengthen this axiom so that types in the space $\Theta$ can be interpreted in terms of distinct expected utility functions.

**Axiom 8** (Stable Discreteness (SD))**.** *There is $A \in \mathcal{M}$ such that for all functions $f : A \to X$ and menus $B \in \mathcal{M}$,*

$$\lim_{\alpha \to 0} k_\rho \left( \bigcup_{x \in A} \alpha f(x) + (1-\alpha)x \right) \geq k_\rho(B).$$

Clearly, SD implies Discreteness because when $\alpha = 0$, then SD implies $k_\rho(A) \geq k_\rho(B)$. It requires that the maximal number of chosen elements in $A$ is unperturbed by small variations where each element $x \in A$ is mixed with $f(x)$.

A finite collection of total types $\Theta \subset \mathcal{T}$ is called *non-redundant* if for all types $\theta_1, \theta_2 \in \Theta$, their closures $\theta_1^*$ and $\theta_2^*$ are non-null, and

$$\theta_1^* = \theta_2^* \implies \theta_1 = \theta_2.$$

In other words, non-redundancy requires that distinct types should maximize distinct utility functions $u_1, u_2 \in \mathcal{U}$ rather than differ only in tie-breaking rules.

**Theorem 3.** *A stochastic choice rule $\rho$ satisfies BA and Discreteness if and only if $\rho$ has a discrete type space $\Theta \subset \mathcal{T}$. Such $\Theta$ is unique, and*

$$|\Theta| = \max_{A \in \mathcal{M}} k_\rho(A).$$

*Moreover, $\Theta$ is non-redundant if and only if $\rho$ satisfies SD.*

## 3.2 Monotonicity and Self-Selection

Monotonicity appears particularly important for stochastic choice models because it is convenient to test empirically and has a simple normative meaning. Monotonicity is proposed by Block and Marschak for their general RUM and holds in all of its refinements, providing that the distribution $\mu$ is invariant of the menu $A$. However, it can be violated by context dependence, where adding extra options to a menu $A$ can increase the probability of choosing $x \in A$ due to reason-based heuristics (e.g., Shafir, Simonson, and Tversky (1993)) and other behavioral reasons.

Next, we focus on the implications of Monotonicity for representation (8) and interpret them in terms of self-selection.

A function $\mu : \mathcal{M} \to \mathcal{D}_\Theta$ is called *selective* if for all menus $A, B \in \mathcal{M}$ and types $\theta \in \mathcal{T}$,

(i) the function $\mu_{\gamma A + (1-\gamma)B}(\theta)$ is continuous over $\gamma \in [0, 1]$, and

(ii) if $\theta$ has the same best element in $A$ and $A \cup B$, then

$$\mu_A(\theta) \geq \mu_{A \cup B}(\theta). \tag{12}$$

Condition (12) asserts that all types in $\Theta$ that preserve their best choice when $A$ is increased to the larger menu $A \cup B$ should not become more likely in $A \cup B$. This condition is plausible under self-selection because such types should not have any motivation to increase their participation in $A \cup B$, while other types can.

**Theorem 4.** *A stochastic choice rule $\rho$ satisfies Monotonicity, BA, Discreteness, and MC if and only if $\rho$ has representation* (8) *where $\Theta$ is a discrete type space, and the function $\mu : \Theta \times \mathcal{M} \to \mathcal{D}_\Theta$ is selective. Both $\Theta$ and $\mu$ are unique.*

This result suggests that self-selection is sufficient to explain deviations from Linearity while Monotonicity and MC are preserved.

Assume a selective environment as in condition (12). With the addition of $x$ to a menu $A$, agents that participated in $A$ would either make the same choice as in $A$ if $x$ is not the best in $A \cup x$, or switch to $x$ otherwise. This implies that no agent would switch to a different option in $A$. Agents that did not participate in $A$, but are participating in $A \cup x$ would only choose $x$. Therefore, no pre-existing option in $A$ would be chosen with a higher probability, which implies that self-selection does not violate Monotonicity.

In fact, we believe[1] that Block-Marschak polynomials constructed in such an environment would be non-negative. Falmagne (1978) stated that such a condition is equivalent to having a random utility representation.

Consider the following example. Let the world consist of two types: $a\theta b\theta c$ and $c\psi b\psi a$, each taking up half of the true population. Assume a type would participate for sure if their favorite option is available, and with probability 50% otherwise. In menu $A = \{a, b\}$, all type $\theta$ agents will show up whereas half of type $\psi$ agents will show up. $\theta$ agents will choose $a$ and $\psi$ agents will choose $b$, and the market will observe a stochastic choice rule of $\rho(a, A) = \frac{2}{3}$ and $\rho(b, A) = \frac{1}{3}$.[2]

However, researchers only observe the stochastic choice rule, and are not privy to the true distribution of types in the world. The following distribution would also provide the same $\rho$:[3]

$$\hat{\mu}(\theta) = \frac{5}{9}, \;\; \hat{\mu}(\psi) = \frac{1}{3}, \;\; \hat{\mu}(\xi) = \frac{1}{9}$$

where $b\xi c\xi a$. This is due to the non-uniqueness of Block-Marschak polynomials (Fishburn, 1998). As a result, the distribution of types can be misspecified in a random utility model where there is no parametric form, i.e., where agents preferences are arbitrary orders on the consumption space.

---

[1] via multiple computer simulations.

[2] $\rho(a, A) = \dfrac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{2} \times \frac{1}{2}}, \rho(b, A) = \dfrac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} + \frac{1}{2} \times \frac{1}{2}}$

[3] $\rho(a, A) = \dfrac{\frac{5}{9}}{\frac{5}{9} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{9}} = \dfrac{2}{3}, \rho(b, A) = \dfrac{\frac{1}{3} \times \frac{1}{2} + \frac{1}{9}}{\frac{5}{9} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{9}} = \dfrac{1}{3}.$

## 4 Discussion

First, we clarify the structure of total types that appear in DREU.

For any $u \in \mathcal{U}$, let

$$L_u = \{\alpha(x-y) : \alpha \geq 0 \text{ and } x, y \in X \text{ such that } u(x) = u(y)\}.$$

As $X$ is convex, then $L_u$ is a linear space.[4]

**Theorem 5.** *A type $\theta$ has a non-null closure $\theta^*$ if and only if there is a function $u \in \mathcal{U}$ and a type $\psi$ in $L_u$ such that for all $x, y \in X$,*

$$x\theta y \iff \begin{array}{l} u(x) > u(y), \text{ or} \\ u(x) = u(y) \text{ and } (x-y)\psi 0. \end{array} \tag{13}$$

*If (13) holds with another pair $(u', \psi')$, then $u' = \alpha u + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$, $L_u = L_{u'}$, and $\psi = \psi'$.*

The proof is in the appendix. Representation (13) asserts that any stable type $\theta$ maximizes some utility function $u \in \mathcal{U}$, which is unique up a positive linear transformation, and then uses the unique residual type $\psi$ on the linear subspace $L_u$ to compare any $x, y \in X$ such that $u(x) = u(y)$.

Representation (13) in Theorem 5 implies that for any stable type $\theta$,

$$\theta \text{ is total in } X \iff \psi \text{ is total in } L_u.$$

Indeed, if $a \sim_\psi b$ for some distinct elements $a, b \in L_u$, then by Independence, $\frac{a-b}{2} \sim_\psi 0$. Take $x, y \in X$ and $\alpha \geq 0$ such that $\alpha(x-y) = \frac{a-b}{2} \neq 0$ and $u(x) = u(y)$. By Independence, $x - y \sim_\psi 0$. By (13), $x \sim_\theta y$, and $\theta$ is not total.

For any $u \in \mathcal{U}$, let $\mathcal{T}_u$ be the set of all total types on $L_u$. Such types are called *tie-breakers.* To construct tie-breakers, take any *basis* $H$ in the linear space $L_u$. By definition, $H \subset L_u$ is a linearly independent system of vectors that spans $L_u$. This definition is standard if $L_u$ is finite dimensional. More generally, such $H$ is known as a *Hamel basis* when $L_u$ is infinite dimensional. The existence of a Hamel basis in any linear space is guaranteed by Zorn's

[4]Indeed, take any $\alpha \geq 0$ and $x, y \in X$ such that $u(x) = u(y)$. Then $\gamma\alpha(x-y) \in L_u$ for all $\gamma \in \mathbb{R}$ because if $\gamma < 0$, then $\gamma\alpha(x-y) = (-\gamma\alpha)(y-x)$. For any $\beta \geq 0$ and $x', y' \in X$ such that $u(x') = u(y')$, the sum

$$\alpha(x-y) + \beta(x'-y') = (\alpha+\beta)\left[\left(\tfrac{\alpha}{\alpha+\beta}x + \tfrac{\beta}{\alpha+\beta}x'\right) - \left(\tfrac{\alpha}{\alpha+\beta}y + \tfrac{\beta}{\alpha+\beta}y'\right)\right]$$

also belongs to $L_u$ because $u\left(\frac{\alpha}{\alpha+\beta}x + \frac{\beta}{\alpha+\beta}x'\right) = u\left(\frac{\alpha}{\alpha+\beta}y + \frac{\beta}{\alpha+\beta}y'\right)$.

lemma. For any $x \in L_u$ and $h \in H$, let $x(h)$ be the coefficient (coordinate) of the vector $h$ in the unique linear combination of basis vectors that equals $x$. Endow $H$ with any total order $\geq$. Define a binary relation $\psi_H$ for all $x, y \in L_u$ via

$$x\psi_H y \iff x = y \text{ or } x(h(x,y)) \geq y(h(x,y)), \tag{14}$$

where $h(x,y) \in H$ is the first vector in $H$ such that $x(h) \neq y(h)$. Then $\psi_H$ compares the coordinates of the vectors $x$ and $y$ *lexicographically* in the basis $H$. By definition, $\psi_H$ is a tie-breaker: it is complete, total, transitive, and satisfies Independence.

Thus a rich family of total types can be generated by representations (13) and (14). These representations can be further refined (e.g. Fishburn (1982), Myerson (1986), Blume, Branderburger, and Dekel (1991)) by imposing more structure on the consumption domain $X$, the functions $u \in \mathcal{U}$, and the basis $H$.

## 4.1 Multiple Risk Attitudes and Beliefs

Olschewski et al. (2022) discovered that the random utility model does not precisely differentiate between risk-averse and risk-seeking agents. This motivates us to next extend the discussion of menu dependence to risk attitudes and beliefs.

First, if $X$ is interpreted as a set of monetary lotteries or AA acts, then the discrete type space $\Theta$ can capture heterogeneity of risk attitudes and/or beliefs over an exogenous state space $S$.

Payoffs $q \in Q$ and lotteries $x \in X$ are called *monetary* if $Q \subseteq \mathbb{R}$ consists of monetary rewards. In this case, let $\mathcal{T}_+ \subseteq \mathcal{T}$ bet the set of all types such that $q > r$ implies $q\theta r$ for all $q, r \in Q$. Each type $\theta \in \mathcal{T}_+$ can be interpreted in terms of the *risk attitude* that is revealed by its increasing utility index $u_\theta$, and some endogenous tie-breaking rule.

If $Q \subseteq \mathbb{R}$ is monetary, then less money should never be chosen over more money.

**Axiom 9** (No Free Lunch (NFL))**.** *For all $q, r \in Q$, if $q > r$, then $\rho(r, \{q, r\}) = 0$.*

An easy corollary of Theorem 3 is that $\rho$ satisfies BA, SD, and NFL if and only if $\rho$ is represented by DREU with a discrete type space $\Theta \subseteq \mathcal{T}_+$. Indeed, NFL is obvious if $\Theta \subseteq \mathcal{T}_+$. Conversely, suppose that $\Phi$ satisfies these axioms. By Theorem 3, $\Phi$ is represented by some non-redundant pair $(\Theta, \mu)$. Suppose that there is $\theta \in \Theta$ that does not belong to $\mathcal{T}_+$. Then $r\theta q$ for some $q > r$. By DREU, $\rho(r, \{q, r\}) > 0$, which contradicts NFL.

Peng (2025) explores the effects of a certain option (degenerate lottery) on risk behavior in an experimental setting. Let there be two types: a risk averse type $\theta$ with a Bernoulli utility index of $u_\theta(x) = \sqrt{x}$ and a risk neutral type $n$ that always chooses the option with a higher expected value. Assume the following observations on two menus:

$$A = \left\{ r \equiv \begin{pmatrix} \$10 & \text{w.p.} & 0.5 \\ \$0 & \text{w.p.} & 0.5 \end{pmatrix}, l \equiv \begin{pmatrix} \$8 & \text{w.p.} & 0.5 \\ \$1 & \text{w.p.} & 0.5 \end{pmatrix} \right\} \text{ and } A' = A \cup \{c \equiv (\$3 \text{ w.p. } 1)\}.$$

A risk averse type will choose $l$ in both cases whereas a risk neutral type will chose $r$. Let the data show that $\rho(r, A) = \rho(l, A) = 0.5$ but $\rho(r, A') = 0.25$ and $\rho(l, A') = 0.75$. This would suggest that in menu $A$, the population consists of half risk neutral agents and half risk averse agents. However, in menu $A'$, risk neutral agents only takes up a quarter of the population while the rest are risk averse. This is a violation of Monotonicity as $l$ is chosen with a higher likelihood in a superset menu. Hence, GP's standard REU would fail to model this phenomenon, and menu-dependence is needed.

We can explain this phenomenon by treating the certain option as a decoy, where its presence will make some agents more risk averse. Another working paper with a similar design (Lim and R., 2021) confirms that adding a safer option increases risk aversion. They also observed that adding a riskier option does not have a noticeable change in risk attitude. The latter was also noticed by Chen et al. (2024). Select slides for the pilot study of the experiment can be found in Appendix B.

This can be extended to more types. Specifically, we can use discrete Arrow-Pratt indices by Baillon and L'Haridon (2021) to capture the finite nature of the type space. A discrete lognormal distribution may be appropriate for $\mu(\cdot, A)$ as CRRA, in the format of $u(x) = \frac{x^r}{1-r}$, indices are positive and skewed right, with most agents exhibiting some level of risk aversion. The mode of $\mu(\cdot, A)$ will move left (towards 0) when a certain option is present.

More broadly, $X$ can be a set $\Delta^{Q \times S}$ of all *acts*—functions $x : S \to \Delta^Q$ on a finite state space $S$. This domain is introduced by Anscombe and Aumann (1963), and can be embedded as a convex compact subset of $L = \mathbb{R}^{Q \times S}$.

When $X$ is the AA domain, then the heterogeneity of risk attitudes is combined with heterogeneity of beliefs on the likelihoods of each state. Epstein and Schneider (2007) found that beliefs can change depending on the environment.

Ambiguity aversion and the multiple-priors model can be captured using AA. Let $L_p$ be a shorthand of the lottery $\begin{pmatrix} \$100 & \text{w.p.} & p \\ \$0 & \text{w.p.} & 1-p \end{pmatrix}$. Consider first the following two acts with two states $\{s_1, s_2\}$: $x_1(s_1) = x_1(s_2) = L_{0.5}$ and $x_2(s_1) = L_{0.55}, x_2(s_2) = L_{0.4}$. Since all lotteries presented are unambiguous, it is reasonable for an agent to set the same beliefs on each state. A Bernoulli utility index of $u(m) = \sqrt{m}$ would dictate the act $x_1$ to be chosen.[5]

[5] The utility for act $x_1$ would be $0.5 \times (0.5 \times \sqrt{100} + 0.5 \times \sqrt{0}) + 0.5 \times (0.5 \times \sqrt{100} + 0.5 \times \sqrt{0}) = 5$ and for act $x_2$ is $0.5 \times (0.55 \times \sqrt{100} + 0.45 \times \sqrt{0}) + 0.5 \times (0.4 \times \sqrt{100} + 0.6 \times \sqrt{0}) = 4.75$.

Let a third act is added: $x_3(s_1) = L_{0.5}, x_3(s_2) = L_p$ where $p$ is determined based on a separate max-min expected utility. Since an ambiguous lottery is presented, the agent may now assign instead a lower belief to $s_2$, perhaps to 0.3. The utility from act $x_2$ would now be higher and hence be the one chosen.[6] The observation of this preference reversal in some agents, as well as a violation of Monotonicity.

## 4.2 Filters and Noisy Data

Theorem 3 can be reinterpreted in terms of filters (or choice functions) rather than stochastic choice rules. A function $\Phi : \mathcal{M} \to \mathcal{M}$ is called a *filter* if $\Phi(A) \subseteq A$ for all $A \in \mathcal{M}$. A filter $\Phi$ is *single-valued* if $|\Phi(A)| = 1$ for all $A \in \mathcal{M}$.

For any type $\theta \in \mathcal{T}$ and menu $A \in \mathcal{M}$, let

$$\Phi_\theta(A) = \{x \in A : x\theta y \text{ for all } y \in A\}$$

be the set of all maximal elements of $\theta$ in $A$. As $\theta$ is total and menus $A$ are finite, then $\Phi_\theta$ is a single-valued filter.

For any stochastic choice rule $\rho$ and menu $A \in \mathcal{M}$, let

$$\Phi_\rho(A) = \{x \in A : \rho(x, A) > 0\}.$$

Define stable choices in any filter $\Phi$ similarly to stochastic choice rules. For any $A \in \mathcal{M}$, let $\Phi^*(A)$ be the set of all elements $x \in A$ such that for any function $f : A \to X$, there is $\varepsilon > 0$ such that for all $\alpha \in [0, \varepsilon)$,

$$\alpha f(x) + (1-\alpha)x \in \Phi\left(\bigcup_{y \in A} \alpha f(y) + (1-\alpha)y\right).$$

Consider four conditions for a filter $\Phi$:

(F1) for all $(x, A) \in \Omega$, $y \in X$,

$$x \in \Phi(A \cup y) \implies x \in \Phi(A),$$

[6]The utility for act $x_1$ would be $0.7 \times (0.5 \times \sqrt{100} + 0.3 \times \sqrt{0}) + 0.5 \times (0.5 \times \sqrt{100} + 0.5 \times \sqrt{0}) = 5$ and for act $x_2$ is $0.7 \times (0.55 \times \sqrt{100} + 0.45 \times \sqrt{0}) + 0.3 \times (0.4 \times \sqrt{100} + 0.6 \times \sqrt{0}) = 5.05$. Due to the max-min expected utility in act $x_3$, it's utility must be lower than act $x_1$.

(F2) for all $(x, A) \in \Omega$, $B \in \mathcal{M}$, and $\gamma \in (0, 1)$,

$$x \in \Phi(A) \iff [\gamma x + (1-\gamma)B] \cap \Phi(\gamma A + (1-\gamma)B) \neq \emptyset,$$

(F3) there is $A \in \mathcal{M}$ such that $|\Phi(A)| \geq |\Phi(B)|$ for all $B \in \mathcal{M}$,

(F4) there is $A \in \mathcal{M}$ such that $|\Phi^*(A)| \geq |\Phi(B)|$ for all $B \in \mathcal{M}$.

Note that BA, Discreteness, and SD for stochastic choice rules are identical to F2, F3, F4 imposed on the filter $\Phi_\rho$. By contrast, Additivity cannot be reduced to any such condition.

Accordingly, DREU is equivalent to representation

$$\Phi_\rho(A) = \bigcup_{\theta \in \Theta} \Phi_\theta(A) \text{ for all } A \in \mathcal{M}. \tag{15}$$

Thus $\Phi_\rho(A)$ consists of all elements that maximize at least one type $\theta \in \Theta$ in the menu $A$. This representation refines the *multiple utility* model proposed by Aizerman and Malishevski Aizerman and Malishevski (1981) (henceforth AM) by restricting the type space $\Theta$ to $\mathcal{T}$ rather than to arbitrary total orders.

Accordingly, Theorem 3 can be rewritten as

**Theorem 6.** *A filter $\Phi : \mathcal{M} \to \mathcal{M}$ satisfies F1–F3 if and only if $\Phi$ is represented by* (15) *for some finite $\Theta \subset \mathcal{T}$. Such $\Theta$ is unique, and*

$$|\Theta| = \max_{A \in \mathcal{M}} |\Phi(A)|.$$

*Moreover, $\Phi$ satisfies F4 if and only if $\Theta$ is non-redundant.*

Note that representation (15) is well-defined even if the types $\theta \in \Theta$ are not total, and $\Phi_\theta$ are multi-valued. Heller Heller (2015) characterizes a special case of representation (15) where the set $\Theta$ consists of mixture continuous types that share some $a, b \in X$ such that $a\theta b$ for all $b \in \Theta$. This unanimity constraint is restrictive for expected utility representations. Besides the differences in the type spaces, Heller's characterization requires $X$ to be a finite dimensional simplex, and one of Heller's main assumptions (CARNI) is existential and cannot be tested in empirical settings.

# A APPENDIX: PROOFS

Prove Theorem 5. In this proof, write any type $\theta$ as a preference relation $\succeq$ with asymmetric and symmetric parts $\succ$ and $\sim$ respectively. Suppose that $\succeq$ is represented by (13) for some

$u \in \mathcal{U}$ and type $\psi$ in $L_u$. Completeness is obvious. To show transitivity, take any $x, y, z \in X$ such that $x \succeq y \succeq z$. Then $u(x) \geq u(y) \geq u(z)$. If $u(x) > u(z)$, then $x \succeq z$ by (13). If $u(x) = u(z)$, then $x\psi y\psi z$ and hence, $x\psi z$. By (13), $x \succeq z$. Next, $\theta$ satisfies Independence because $u$ is mixture linear, and the tie-breaker $\psi$ obeys Independence. Stability has been already derived from representation (4) and a fortiori, from (13).

Conversely, take any stable type $\succeq$ and derive representation (13). Take $x, y$ that satisfy the stability condition (3). Then $x \succ y$ because $y \succeq x$ would imply by Independence that for any $z_x \prec z_y$ and $\alpha \in (0,1)$, $\alpha z_x + (1-\alpha)x \prec \alpha z_y + (1-\alpha)y$. Thus (3) would not hold.

Take any menu $A = \{a_1, a_2, \ldots, a_n\}$ and index so that $a_1 \succeq a_2 \succeq \cdots \succeq a_n$. Let $Z$ be the convex hull of $A$ in $L$.

Show that for all $z \in Z$,

$$a_1 \succeq z \succeq a_n. \tag{16}$$

The proof is by induction with respect to $n \geq 2$. If $n = 2$, then $z = \gamma a_1 + (1-\gamma)a_2$ for some $\gamma \in (0,1)$, and (16) follows from $a_1 \succ a_2$ by Independence. Let $n > 2$ and $Z'$ be the convex hull of $A \setminus a_n$. Take any $z \in X$ and let $\gamma = z(a_n)$. Then $z = \gamma a_n + (1-\gamma)y$ for some $y \in X'$. By the inductive assumption, $a_1 \succeq y \succeq a_{n-1} \succeq a_n$. By Independence,

$$a_1 \succeq \gamma a_n + (1-\gamma)a_1 \succeq \gamma a_n + (1-\gamma)y \succeq \gamma a_n + (1-\gamma)a_n = a_n$$

and hence, (16) holds.

For each $\alpha \in [0,1]$, let $z_\alpha = \alpha a_1 + (1-\alpha)a_n$. Show that for all $\alpha, \beta \in [0,1]$,

$$\alpha \geq \beta \iff z_\alpha \succeq z_\beta. \tag{17}$$

By Independence,

$$\alpha > \beta \implies z_\alpha = \alpha a_1 + (1-\alpha)a_n \succ a_n \implies$$
$$z_\alpha = \tfrac{\beta}{\alpha} z_\alpha + \tfrac{\alpha-\beta}{\alpha} z_\alpha \succ \tfrac{\beta}{\alpha} z_\alpha + \tfrac{\alpha-\beta}{\alpha} a_n = z_\beta.$$

If $\alpha \geq \beta$, then either $z_\alpha = z_\beta$ or $z_\alpha \succ z_\beta$. Thus (17) holds. For all $x \in X$, let

$$u_\theta(x) = \inf\{\alpha \in [0,1] : z_\alpha \succeq x\}$$
$$v_\theta(x) = \sup\{\beta \in [0,1] : x \succeq z_\beta\}.$$

By (16), these values are well-defined. Show that $u_\theta = v_\theta$. Indeed, $z_\alpha \succeq x \succeq z_\beta$ implies $\alpha \geq \beta$ and hence, $u_\theta(x) \geq v_\theta(x)$. Suppose that $u_\theta(x) > v_\theta(x)$. Then there is $\gamma$ such that

$u_\theta(x) > \gamma > v_\theta(x)$. By definition of $u_\theta$, $u_\theta(x) > \gamma$ implies $x \succ z_\gamma$. By definition of $v_\theta$, $\gamma > v_\theta(x)$ implies $z_\gamma \succ x$. By contradiction, $u_\theta(x) = v_\theta(x)$.

Show that $u_\theta$ satisfies the mixture linearity. By Independence, for any $x, y \in X$ and $\alpha, \beta, \gamma \in [0,1]$ such that $z_\alpha \succeq x$ and $z_\beta \succeq y$,

$$z_{\gamma\alpha+(1-\gamma)\beta} = \gamma z_\alpha + (1-\gamma)z_\beta \succeq \gamma x + (1-\gamma)z_\beta \succeq \gamma x + (1-\gamma)y.$$

Thus $u_\theta(\gamma x + (1-\gamma)y) \leq \gamma\alpha + (1-\gamma)\beta$. By definition of $u_\theta$, $\alpha$ and $\beta$ can be arbitrarily close to $u_\theta(x)$ and $u_\theta(y)$ respectively. Thus

$$u_\theta(\gamma x + (1-\gamma)y) \leq \gamma u_\theta(x) + (1-\gamma)u_\theta(y).$$

Analogously, $v_\theta(\gamma x+(1-\gamma)y) \geq \gamma v_\theta(x)+(1-\gamma)v_\theta(y)$. As $u_\theta = v_\theta$, then $u_\theta$ is mixture linear.

Suppose that $x$ does not maximize $u_\theta$ in $A$. Take $y \in A$ such that $u_\theta(y) > u_\theta(x)$. Take $\gamma$ such that $u_\theta(y) = v_\theta(y) > \gamma > u_\theta(x)$. Then $y \succ z_\gamma \succ x$ and hence, $x$ does not maximize $\succeq$ in $A$. Thus $\Phi_\theta(A) \in \Phi_{u_\theta}(A)$.

## A.1 Main Proofs

Theorems 3–4 are shown together. Suppose that $\rho$ is represented by some discrete map $\mu$ with a finite support $\Theta = \{\theta_1, \ldots, \theta_k\} \subset \mathcal{T}$. Construct a menu $B = \{b_1, \ldots, b_k\}$ such that

$$\theta_i \in \mathcal{T}(b_i, B).$$

Let $I$ be the set of all pairs $i, j \in \{1, \ldots, k\}$ such that $i > j$. For each pair $(i,j) \in I$, the types $\theta_i$ and $\theta_j$ are distinct and total. Thus there are $x_{ij}, x_{ji} \in X$ such that $x_{ij}\theta_i x_j$ and $x_{ji}\theta_j x_i$. For any $n \in \{1, \ldots, k\}$, let

$$b_n = \sum_{(i,j)\in I} \tfrac{2}{k(k-1)} x_{ij}^n,$$

where $x_{ij}^n = x_{ij}$ if $x_{ij}\theta_p x_{ji}$ and $x_{ij}^n = x_{ji}$ otherwise. For any $m, n \in \{1, \ldots, k\}$, the types $\theta_n, \theta_m$ satisfy Independence and hence,

$$b_n = \left(\sum_{(i,j)\in I} \tfrac{2}{k(k-1)} x_{ij}^n\right) \theta_n \left(\sum_{(i,j)\in I} \tfrac{2}{k(k-1)} x_{ij}^m\right) = b_m$$

because $x_{ij}^n \theta_n x_{ij}^m$ for all $(i,j) \in I$.

## A.2 Proof of Theorem 6

We start by proving Theorem 6 for filters.

Consider a filter $\Phi : \mathcal{M} \to \mathcal{M}$ and a finite type space $\Theta \subset \mathcal{T}$ such that

$$\Phi(A) = \bigcup_{\theta \in \Theta} \Phi_\theta(A). \tag{18}$$

This representation implies F1, F2, and F3. Indeed, take any type $\theta \in \Theta$, $(x, A) \in \Omega$, $B \in \mathcal{M}$, and $\gamma \in (0, 1)$. Then $x = \Phi_\theta(A)$ if and only if $\gamma x + (1-\gamma)\Phi_\theta(B) = \Phi_\theta(\gamma x + (1-\gamma)B)$. Thus F1 holds. If $x = \Phi_\theta(A \cup B)$, then $x \in \Phi_\theta(A)$. Thus F2 holds. As all types $\theta \in \Theta$ are total, then $|\Phi(A)| \leq |\Theta|$. Thus F3 holds.

Conversely, let $\Phi$ satisfy F1, F2, F3. Suppose first that $k = 1$, and $\Phi$ is single-valued. Define a binary relation $\theta$ as

$$x\theta y \iff x = \Phi(\{x, y\}).$$

As $\Phi$ is a single-valued filter, then $\theta$ is complete and total. Show that $\theta$ is transitive. Suppose that $x\theta y\theta z\theta x$. By F2, $x = \Phi(\{x, y, z\})$ implies $x = \Phi(\{x, z\})$, which contradicts $z\theta x$. Similarly, $y = \Phi(\{x, y, z\})$ or $z = \Phi(\{x, y, z\})$ contradict $x\theta y$ and $y\theta z$ respectively. Thus $\theta$ is transitive. Next, by F1, for all $x, y, z \in X$ and $\gamma \in (0, 1]$,

$$\gamma\Phi(\{x, y\}) + (1-\gamma)z = \Phi(\{\gamma x + (1-\gamma)z, \gamma y + (1-\gamma)z\}).$$

Thus $\theta$ satisfies Independence, and hence $\theta \in \mathcal{T}$ is a type. Show that $\Phi = \Phi_\theta$. Both $\Phi$ and $\Phi_\theta$ satisfy F2. Thus for all $A \in \mathcal{M}$,

$$\Phi(A) = \Phi(\{\Phi(A), \Phi_\theta(A)\}) =_{\text{by definition of } \theta} \Phi_\theta(\{\Phi(A), \Phi_\theta(A)\}) = \Phi_\theta(A).$$

Suppose that $\Phi(A) \neq \Phi_\theta(A)$ for some $A \in \mathcal{M}$. By F2, $\Phi(A) = \Phi(\{\Phi(A), \Phi_\theta(A)\})$. Thus $\Phi_\theta(A)$ is not maximal for $\theta$ in $A$. By contradiction, $\Phi(A) = \Phi_\theta(A)$ for all $A \in \mathcal{M}$.

Take $B \in \mathcal{M}$ such that $|\Phi(B)| = k$. Let $C = \Phi(B) = \{c_1, \ldots, c_k\}$. By Chernoff, $\Phi(C) = C$ because $C \subset B$. Let $R = \max_{i=1\ldots,k} ||x_i||$. Take $\alpha \in (0, 1)$ such that

$$2\alpha R < (1-\alpha) \min_{1 \leq i < j \leq k} ||x_i - x_j||.$$

Then for any $\alpha \in (0, \alpha_{xy})$, the menu $\alpha\{x, y\} + (1-\alpha)B$ has $2k$ distinct elements. To show this, it is enough to check that for any distinct $i, j \in \{1, \ldots, k\}$, the menu

$$\alpha\{x, y\} + (1-\alpha)\{x_i, x_j\}$$

has four distinct elements. Here the inequalities

$$\alpha x + (1-\alpha)x_i \neq \alpha y + (1-\alpha)x_i \neq \alpha y + (1-\alpha)x_j \neq \alpha x + (1-\alpha)x_j \neq \alpha x + (1-\alpha)x_i$$

are obvious, the other two inequalities $\alpha x + (1-\alpha)x_i \neq \alpha y + (1-\alpha)x_j$ and $\alpha x + (1-\alpha)x_j \neq \alpha y + (1-\alpha)x_i$ follow from

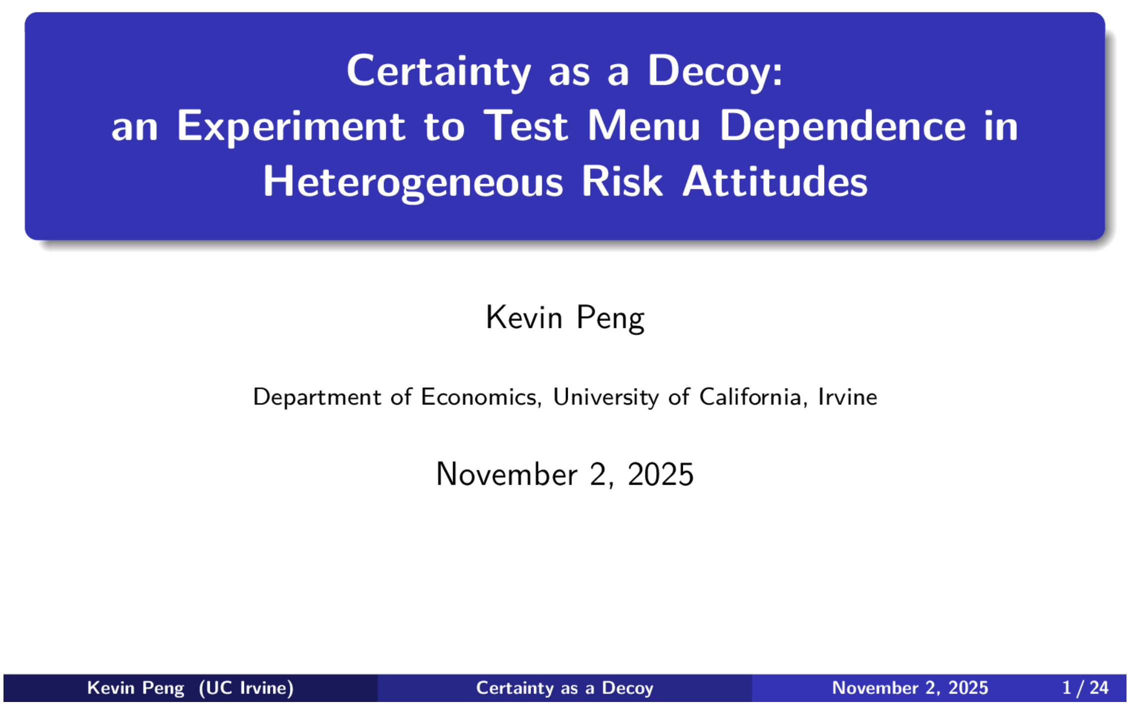
$$(1-\alpha)||x_i - x_j|| - \alpha||x - y|| > 0.$$

As $x_i \in \Phi(B)$ for all $i$, then by Type Independence, $\Phi(\alpha\{x, y\} + (1-\alpha)B)$ overlaps with each of the $k$ distinct sets

$$A_i = \alpha\{x, y\} + (1-\alpha)x_i.$$

Each of these $k$ overlaps contains precisely one element, $\alpha x + (1-\alpha)x_i$ or $\alpha y + (1-\alpha)x_i$.

# B Select Slides for Experiment



## Research Question and Contribution

On the aggregate level, *the distribution of risk attitudes can vary depending on what options are presented to the decision makers.*

- *Harrison, et al., 12*: individuals are risk neutral when lotteries are presented in a binary format $L = \begin{pmatrix} \$x & \text{w.p. } \pi \\ \$y & \text{w.p. } 1-\pi \end{pmatrix}$.
- *Hermanns & Kokot, 23*: framing can affect levels of risk aversion.

**Question**: does the presence of a certain option in the menu make agents on average more risk averse?

- Most literature focuses on **framing** effects (*Crosta, et al., 23*).
- *Chen, et al., 24*: introducing a high-risk-high-reward lottery does not decrease risk aversion.

$\rightarrow$ I designed the menus so a no-risk-low-reward is introduced.

$\rightarrow$ **Risk aversion is observed to increase**.



## Monotonicity

When additional options are added to a menu, the probability of choosing a **pre-existing** option should **not increase**: for all $x \in A$, $A$, $y$,

$$\rho(x, A) \geq \rho(x, A \cup \{y\})$$

- A stochastic choice rule that has a random utility representation is **monotone**.
- However, if agents become more risk averse with the expansion of a menu, *they may switch to a pre-existing option that is of lower risk.*
- **Goal**: violations of monotonicity in some observations.

$\Longrightarrow$ a menu-dependent version of random utility is needed to model the choice data.

## Hypotheses

Introduce a no-risk-low-reward (certain) option:

$$A \equiv \{\text{low-risk } (l), \text{high-risk } (h)\} \to \{l, h, \text{certain } (c)\} \equiv A'.$$

***H1*:** the proportion of agents that choose $h$ will decrease.
***H2*:** the proportion of agents that choose $l$ will increase.
***H3*:** the relative proportion of agents that chose $l$ will increase over $h$:

$$\rho(l, A) \leq \frac{\rho(l, A')}{\rho(l, A') + \rho(h, A')}.$$

Note that $H2 \implies H3 \implies H1$.

$\rightarrow$ **Decoy effects** and **extemeness aversion** suggest at least $H1$.

***H4*:** exposure to $c$ has an extended effect, even after it is removed.

$\rightarrow$ *Hedgcock, Rao & Chen, 16*: decoy effects can last beyond removal.

$\rightarrow$ *Pettibone & Wedell, 07*: phantom decoys (not selectable).



## Menu Without a Certain Option

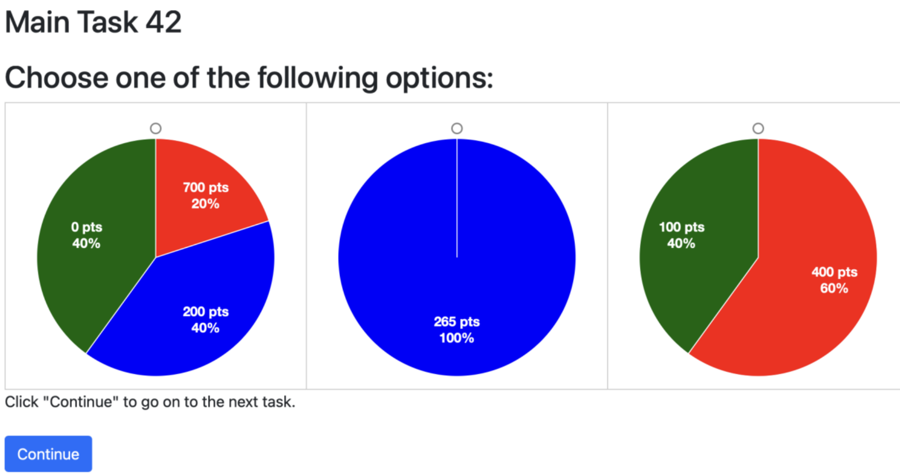
Main Task 2

Choose one of the following options:


Click "Continue" to go on to the next task.

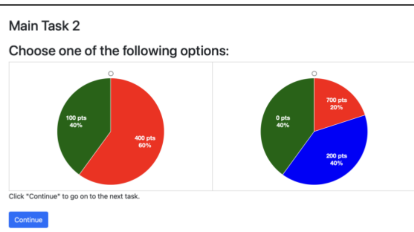
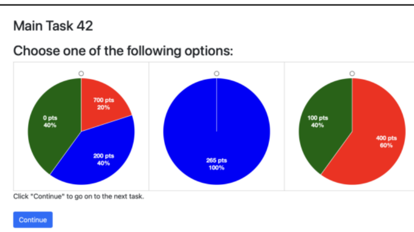
Continue

# Menu With a Certain Option






# Main Tasks and Treatments

| Menu $A$ | Menu $A'$ |
|---|---|
|  |  |





- Each Menu $A$ has a corresponding Menu $A' \supseteq A$.
- There are 26 pairs, with 52 total main tasks.

**T1**: Subjects face the 26 $A$ menus first, and then move on to 26 $A'$ menus.

**T2**: Subjects face the 26 $A'$ menus first, and then move on to 26 $A$ menus.

**T3**: Subjects face alternating menus $A \to A' \to A \to \cdots$.

## Pilot Study

23 test subjects (UCI economics graduate students).
48 choices each, with a total of 1104 observations.

| (#) | high-risk ($h$) option no $c$ | high-risk ($h$) option with $c$ | low-risk ($l$) option no $c$ | low-risk ($l$) option with $c$ | certain option |
|---|---|---|---|---|---|
| *T1* (8) | 95 (49%) | 77 (40%) | 97 (51%) | 84 (43%) | 31 (16%) |
| *T2* (7) | 77 (46%) | 58 (35%) | 91 (54%) | 102 (61%) | 8 (5%) |
| *T3* (8) | 72 (38%) | 70 (36%) | 120 (62%) | 92 (48%) | 30 (16%) |
| Pool | 244 (44%) | 205 (37%) | 308 (56%) | 278 (50%) | 69 (13%) |
| Cond. | 44% | 43% | 56% | 57% | |

- *H1* (decrease in high-risk choice) holds here.

→ On average, agents become more risk averse with the presence of $c$.

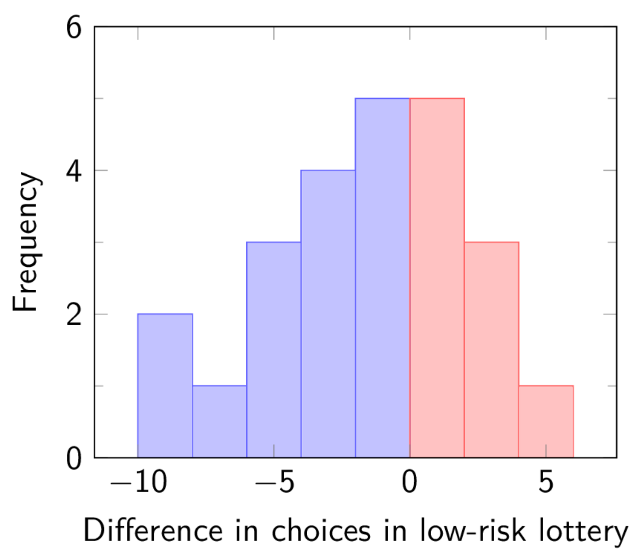
- *H3* does not seem statistically significant. (*H2* does not hold.)

→ Some people switch from $l$ to $c$.



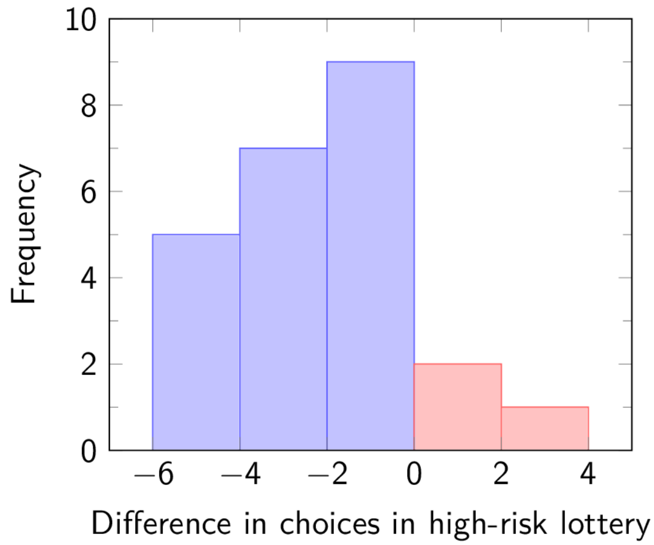
## Monotonicity: Low-Risk Option

When separately discussing each menu pair, monotonicity is violated in 9 of the 24 pairs when comparing the number of people that chose the low-risk option with or without the certain option ($c$).



## Monotonicity: High-Risk Option

Monotonicity is violated in 3 of the 24 pairs when comparing the number of people that chose the *high*-risk option with or without the certain option.






## Logistic Regression with Random Effects

$$\log\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right) =$$
$$\beta_0 + \beta_1 \mathit{threeHigh}_{ij} + \beta_2 \mathit{threeLow}_{ij} + \beta_3 \mathit{switch}_{ij} + \beta_4 \mathit{certain}_{ij} + b_i$$

- $Y_{ij}$: whether subject $i$ chooses the riskier lottery in task $j$.
- $\mathit{threeHigh}_{ij}$: whether the high-risk lottery in the task has three states.
- $\mathit{threeLow}_{ij}$: whether the low-risk lottery in the task has three states.
- $\mathit{switch}_{ij}$: the level of risk aversion (category) that would switch from high- to low-risk.
- $\mathit{certain}_{ij}$: whether a certain lottery is present in the task.
- $b_i \sim \mathcal{N}(0, \sigma_b^2)$: random intercept for subject $i$.

## Estimation

| | **Estimate** | **Std. Error** | $z$ **value** | **Pr**$(> \lvert z \rvert)$ |
|---|---|---|---|---|
| (Intercept) | -3.562 | 0.319 | -11.150 | $<$ 2e-16$^{***}$ |
| threeHigh | -0.165 | 0.140 | -1.174 | 0.240 |
| threeLow | 0.049 | 0.140 | 0.353 | 0.724 |
| switch | 0.603 | 0.046 | 13.155 | $<$ 2e-16$^{***}$ |
| certain | -0.382 | 0.141 | -2.719 | 0.007$^{**}$ |

*Signif. codes:* 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Table 1:** Fixed effect estimates from mixed-effects logistic regression predicting whether a subject will choose the high-risk lottery.



## Persistent Effect of Exposure (*H4*)

There should be higher level of risk aversion in the $A$ portion (no $c$) for subjects faced with $A'$ (with $c$) first (*T2*) than those with $A$ first (*T1*).

| | high-risk option | | low-risk option | | certain |
|---|---|---|---|---|---|
| | no $c$ | with $c$ | no $c$ | with $c$ | option |
| *T1* | 95 (**49%**) | 77 (40%) | 97 (51%) | 84 (43%) | 31 (16%) |
| *T2* | 77 (**46%**) | 58 (35%) | 91 (54%) | 102 (61%) | 8 (5%) |

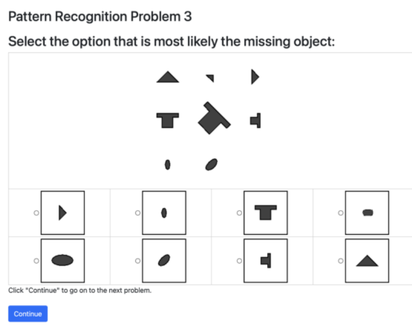
- *T3*'s level of risk aversion should increase over time. The starting level should be similar to *T1* and end similar to *T2*.

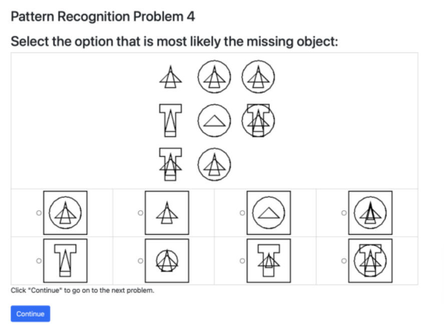
$\rightarrow$ Plan: compare first 10 tasks with last 10 tasks.

## Cognitive Test: Pattern Recognition

*Matzen, et al., 10*: a non-branded version of Raven's Progressive Matrices.

Patterns | Logic gates




- 8 problems with progressive difficulty.
- Plan: find relationship between cognition and menu-dependent risk attitudes. Cognizant agents less likely to be affected by the decoy?



## Conclusion

Findings in the pilot study:

- Agents are on average **more risk averse** when a certain option is presented.
- *Monotonicity* is **violated** in a third of the scenarios.

To dos for the main study (around 270 subjects, to be collected in ESSL):

- Observe what types of menus are more likely to violate monotonicity.
- Analysis on persistence of decoy effects (*H4*).
- Relate cognition to sensitivity to decoy effects.

# References


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

D. S. Ahn and H. Ergin. Framing contingencies. 78(2):655–695, 2010.

M. Aizerman and A. Malishevski. General theory of best variants choice. *IEEE Trans. Automat. Control*, pages 1030–1041, 1981.

F. Anscombe and R. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

J. Apesteguia, M. Ballester, and J. Lu. Single-crossing random utility models. *Econometrica*, 85:661–674, 2017.

A. Baillon and O. L'Haridon. Discrete arrow-pratt indexes for risk and uncertainty. *Economic Theory*, 72(4):1375–1393, 2021.

H. Block and J. Marschak. Random orderings and stochastic theories of response. In Olkin, Ghurye, Hoeffding, Madow, and Mann, editors, *Contributions to Probability and Statistics*, pages 97–132. Stanford University Press, Palo Alto, 1960.

L. Blume, A. Branderburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, pages 61–79, 1991.

Z. Chen, R. Golman, and J. Somerville. Menu-dependent risk attitudes: Theory and evidence. *Journal of Risk and Uncertainty*, 68(1):77–105, 2024.

A. Daly and S. Zachery. Improved multiple choice models. In D. Hensher and Q. Dalvi, editors, *Identifying and Measuring the Determinants of Mode Choice*. London: Teakfield, 1979.

L. G. Epstein and M. Schneider. Learning under ambiguity. *The Review of Economic Studies*, 74(4):1275–1303, 2007.

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

P. C. Fishburn. *The Foundations of Expected Utility*. D. Reidel Publishing Company, Dordrecht, 1982.

P. C. Fishburn. Stochastic utility. In S. Barbera, P. J. Hammond, and C. Seidl, editors, *Handbook of Utility Theory*. Springer, 1998.

E. E. Garnett, A. Balmford, C. Sandbrook, M. A. Pilling, and T. M. Marteau. Impact of increasing vegetarian availability on meal selection and sales in cafeterias. *Proceedings of the National Academy of Sciences*, 116(42):20923–20929, 2019.

F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74:121–146, 2006.

G. W. Harrison, J. MartÃnez-Correa, and J. T. Swarthout. Inducing risk neutral preferences with binary lotteries: A reconsideration. *Journal of Economic Behavior & Organization*, 94:145–159, 2013.

Y. Heller. Justifiable choice. *Games and Economic Behavior*, 76:375–390, 2015.

I. Herstein and J. Milnor. An axiomatic approach to measurable utility. *Econometrica*, 21: 291–297, 1953.

N. Hing, M. Smith, M. Rockloff, H. Thorne, A. M. T. Russell, N. A. Dowling, and B. Breen. How structural changes in online gambling are shaping the contemporary experiences and behaviours of online gamblers: an interview study. *BMC Public Health*, 22(1620), 2022.

X. Lim and S. R. Avoiding risk in the lab: An experiment on avoidable risk. Working paper, 2021. URL `https://silvioravaioli.github.io/website-documents/Research/Avoidable%20Risk/LimRavaioli2021AvoidableRisk.pdf`.

J. Lu. Random ambiguity. *Theoretical Economics*, 16(2):539–570, 2021.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Economics*. Academic Press, New York, 1973.

D. McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001. doi: 10.1257/aer.91.3.351.

R. Myerson. Multistage games with communication. *Econometrica*, pages 323–358, 1986.

S. Olschewski, P. Sirotkin, and J. Rieskamp. Empirical underidentification in estimating random utility models: The role of choice sets and standardizations. *British Journal of Mathematical and Statistical Psychology*, 75(2):252–292, 2022.

K. Peng. Certainty as a decoy: an experiment to test menu dependence in heterogeneous risk attitudes. Working paper, 2025.

E. Piermont. Disentangling strict and weak choice in random expected utility models. *Journal of Economic Theory*, 202:201–230, 2022.

M. Rabin. Psychology and economics. *Journal of Economic Literature*, pages 11–46, 1998.

E. Shafir, I. Simonson, and A. Tversky. Reason-based choice. *Cognition*, 49:11–36, 1993.

K. M. Sharpe, R. Staelin, and J. Huber. Using extremeness aversion to fight obesity: Policy implications of context dependent demand. *Journal of Consumer Research*, 35(3):406–422, 2008.

K. E. Train. Recreation demand models with taste differences over people. *Land economics*, pages 230–239, 1998.

H. Williams. On the formulation of travel demand models and economic measures of user benefit. *Environment and Planning A*, 9:285–344, 1977.

E. Yang and I. Kopylov. Random quasi-linear utility. *Journal of Economic Theory*, 209, 2023. doi: 10.1016/j.jet.2023.105650.