Appendix : Table of Contents

Appendix I : Theoretical Framework

Our setting is a portfolio choice environment with S states of nature, with each state denoted by $s = 1, 2, \ldots, S$. For each state s, there is an Arrow security that pays one in state s and zero in the other state(s). We assume that each state s occurs with probability $\pi_s \in (0,1)$, so that $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_S) \in (0,1)^S$ is a vector of state probabilities where $\pi_1 + \pi_2 + \cdots + \pi_S = 1$; furthermore, the distribution $\boldsymbol{\pi}$ is commonly known among decision makers. In the experiment reported in the main paper, $S = 3$ and $\pi_s = \frac{1}{3}$ for $s = 1, 2, 3$. Let $x_s \geqslant 0$ denote the demand for the security that pays off in state s and $p_s > 0$ denote the corresponding price, so that $\mathbf{x} = (x_1, x_2, \ldots, x_S) \in \mathbb{R}^S_+$ is an asset demand allocation or contingent consumption bundle and $\mathbf{p} = (p_1, p_2, \ldots, p_S) \in \mathbb{R}^S_{++}$ is a vector of state prices.

A individual subject's preference over bundles in the consumption space $\mathbb{R}^S_+$ can be represented by a binary relation $\succsim$. This relation is *complete* if for any $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^S_+ \times \mathbb{R}^S_+$, either $\mathbf{x} \succsim \mathbf{y}$, $\mathbf{y} \succsim \mathbf{x}$, or both; in particular, a complete binary relation is *reflexive*, i.e., $\mathbf{x} \succsim \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^S_+$. The preference $\succsim$ is *transitive* if for any $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathbb{R}^S_+ \times \mathbb{R}^S_+ \times \mathbb{R}^S_+$, if $\mathbf{x} \succsim \mathbf{y}$ and $\mathbf{y} \succsim \mathbf{z}$, then $\mathbf{x} \succsim \mathbf{z}$. It is *continuous* if the set $\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^S_+ \times \mathbb{R}^S_+ : \mathbf{x} \succsim \mathbf{y}\}$ is closed; in particular, this implies that the *preferred set* of any bundle $\mathbf{x} \in \mathbb{R}^S_+$, the set $\mathcal{P}(\mathbf{x}) = \{\mathbf{x}' \in \mathbb{R}^S_+ : \mathbf{x}' \succsim \mathbf{x}\}$, is a closed set containing $\mathbf{x}$. We denote the asymmetric part of $\succsim$ by $\succ$, i.e., $\mathbf{x} \succ \mathbf{x}'$ if $\mathbf{x} \succsim \mathbf{x}'$ but $\mathbf{x}' \not\succsim \mathbf{x}$. If $\succsim$ is a complete preference, then $\mathcal{P}_0(\mathbf{x}) = \{\mathbf{x}' \in \mathbb{R}^S_+ : \mathbf{x}' \succ \mathbf{x}\}$ is an open set not containing $\mathbf{x}$. It is well known that if a preference $\succsim$ is complete, transitive, and continuous, then it is representable by a continuous utility function $U : \mathbb{R}^S_+ \to \mathbb{R}$ (see Debreu (1954)), i.e., $U(\mathbf{x}) \geqslant U(\mathbf{x}')$ if and only if $\mathbf{x} \succsim \mathbf{x}'$.

1 Basic Rationalizability

Let $\mathcal{D} := \{(\mathbf{p}^i, \mathbf{x}^i)\}_{i=1}^I$ be a finite dataset generated by an individual subject's choices from linear budget sets, where $\mathbf{p}^i$ denotes the i-th observation of the price vector and $\mathbf{x}^i$ denotes the corresponding demand allocation. We say that a dataset $\mathcal{D}$ can be *rationalized* by a complete preference $\succsim$ if (i) $\mathbf{x}^i \succsim \mathbf{x}$ for all $\mathbf{x} \in \mathcal{B}^i = \{\mathbf{x} \in \mathbb{R}^S_+ : \mathbf{p}^i \cdot \mathbf{x} \leqslant \mathbf{p}^i \cdot \mathbf{x}^i\}$ and (ii) $\mathbf{x}^i \succ \mathbf{x}$ for all $\mathbf{x} \in \mathcal{B}^i$ with $\mathbf{p}^i \cdot \mathbf{x} < \mathbf{p}^i \cdot \mathbf{x}^i$. It is straightforward to check that condition (ii) is a consequence of condition (i) if the preference is complete, transitive, and *locally nonsatiated* (which means that for any open neighborhood of $\mathbf{x} \in \mathbb{R}^S_+$, there is $\mathbf{x}'$ in the neighborhood such that $\mathbf{x}' \succ \mathbf{x}$). Similarly, condition (ii) is a consequence of (i) if the preference is complete (but not necessarily transitive) and has *extended local nonsatiation* in the sense that if $\mathbf{x} \sim \mathbf{x}'$ then, for any open neighborhood of $\mathbf{x}'$, there is $\mathbf{y}$ such that $\mathbf{y} \succ \mathbf{x}$.

Let $\mathcal{X} := \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^I\}$ be the finite set of bundles observed in the dataset $\mathcal{D}$. For any pair $(\mathbf{x}^i, \mathbf{x}^j) \in \mathcal{X} \times \mathcal{X}$, we say that $\mathbf{x}^i$ is *directly revealed preferred* to $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i \, R^D \, \mathbf{x}^j$) if $\mathbf{p}^i \cdot \mathbf{x}^i \geqslant \mathbf{p}^i \cdot \mathbf{x}^j$; and if this inequality is strict, we say that $\mathbf{x}^i$ is *directly revealed strictly preferred to* $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i \, P^D \, \mathbf{x}^j$).

Suppose that $\mathcal{D}$ can be rationalized by a complete preference $\succsim$. Then $\mathbf{x}^i \succsim \mathbf{x}^j$ if $\mathbf{x}^i \, R^D \, \mathbf{x}^j$ and $\mathbf{x}^i \succ \mathbf{x}^j$ if $\mathbf{x}^i \, P^D \, \mathbf{x}^j$. It follows that $\mathcal{D}$ cannot simultaneously have $\mathbf{x}^i \, R^D \, \mathbf{x}^j$ and $\mathbf{x}^j \, P^D \, \mathbf{x}^i$; in other words, $\mathcal{D}$ must obey the *weak axiom of revealed preference* (WARP).

The *revealed preferred* relation, denoted by R, is the transitive closure of the directly revealed preferred relation, i.e., $\mathbf{x}^i \, R \, \mathbf{x}^j$ if there exists a sequence of allocations $\{\mathbf{x}^k\}_{k=1}^K$ with $\mathbf{x}^1 = \mathbf{x}^i$ and $\mathbf{x}^K = \mathbf{x}^j$, such that $\mathbf{x}^k \, R^D \, \mathbf{x}^{k+1}$ for every $k = 1, \ldots, K-1$. The *generalized axiom of revealed preference* (GARP) requires that if $\mathbf{x}^i \, R \, \mathbf{x}^j$ then it is not the case that $\mathbf{x}^j \, P^D \, \mathbf{x}^i$. If $\mathcal{D}$ can be rationalized by a complete and transitive preference $\succsim$, then $\mathbf{x}^i \succsim \mathbf{x}^j$ if $\mathbf{x}^i \, R \, \mathbf{x}^j$ and $\mathbf{x}^i \succ \mathbf{x}^j$ if $\mathbf{x}^i \, P^D \, \mathbf{x}^j$; in other words, GARP holds. The substantive part of Afriat's (1967) Theorem says that the converse of this statement is also true: *if $\mathcal{D}$ satisfies GARP then $\mathcal{D}$ can be rationalized by a continuous and increasing utility function $U : \mathbb{R}^S_+ \to \mathbb{R}$.* We refer to any utility function that is continuous and increasing as *well-behaved.*

Figure 1 illustrates a straightforward violation of GARP (and WARP) involving only two observations: $\mathbf{p}^1 = \left(\frac{3}{9}, \frac{2}{9}, \frac{1}{9}\right)$, $\mathbf{x}^1 = (1, 2, 2)$; and $\mathbf{p}^2 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$, and $\mathbf{x}^2 = (0, 1, 5)$. GARP (and WARP) is (are) clearly violated since $\mathbf{p}^1 \cdot \mathbf{x}^1 > \mathbf{p}^1 \cdot \mathbf{x}^2$ and $\mathbf{p}^2 \cdot \mathbf{x}^2 > \mathbf{p}^2 \cdot \mathbf{x}^1$.

Rationalizability Index e^*. Afriat (1973) provides an extension to Afriat's (1967) Theorem which facilitates the calculation of the *critical cost-efficiency index* (CCEI). Recall (from the main paper) that we say that $\mathcal{D}$ *is rationalizable at cost-efficiency* e if there is a well-behaved utility function $U : \mathbb{R}^S_+ \to \mathbb{R}$ such that $U(\mathbf{x}^i) \geqslant U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{B}^i(e) = \{\mathbf{x} \in \mathbb{R}^S_+ : \mathbf{p}^i \cdot \mathbf{x} \leqslant e \, \mathbf{p}^i \cdot \mathbf{x}^i\}$. The CCEI (which we refer to as the rationalizability index e^* in the main paper) is the supremum of e such that $\mathcal{D}$ is rationalizable at cost-efficiency e.

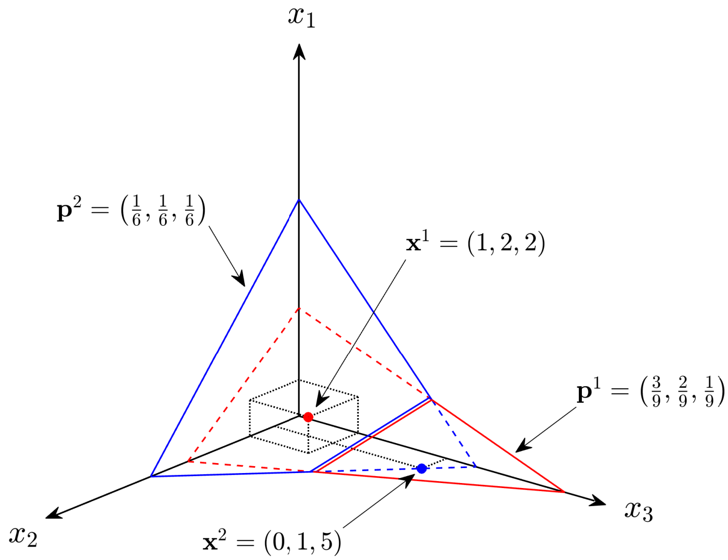


Figure 1: Violation of Basic Rationalizability

Whether or not a dataset $\mathcal{D}$ is rationalizable at cost-efficiency e can be checked by a suitable extension of GARP. Given $e \in (0, 1]$, for any pair $(\mathbf{x}^i, \mathbf{x}^j) \in \mathcal{X} \times \mathcal{X}$, we say that $\mathbf{x}^i$ is *directly revealed preferred at cost-efficiency* e to $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i \, R^D(e) \, \mathbf{x}^j$) if $e \, \mathbf{p}^i \cdot \mathbf{x}^i \geqslant \mathbf{p}^i \cdot \mathbf{x}^j$; if this inequality is strict, we say that $\mathbf{x}^i$ is *directly revealed strictly preferred at cost-efficiency* e to $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i \, P^D(e) \, \mathbf{x}^j$). We denote the transitive closure of $R^D(e)$ by $R(e)$.

We say that $\mathcal{D}$ satisfies *GARP at cost-efficiency* e if, whenever $\mathbf{x}^i \, R(e) \, \mathbf{x}^j$, then it is not the case that $\mathbf{x}^j \, P^D(e) \, \mathbf{x}^i$. Afriat (1973) shows that $\mathcal{D}$ *is rationalizable at cost-efficiency* e *if and only if it satisfies GARP at cost-efficiency* e.

Removing WARP Violations. In Section 5.2 of the main paper, we refer to "the CCEI which measures the amount by which each budget constraint needs to be reduced in order to remove all violations of WARP." We now explain precisely what we mean. For each $e \in (0, 1]$, we can check whether there is $\mathbf{x}^i$ and $\mathbf{x}^j$ in $\mathcal{D}$ such that $\mathbf{x}^i \, R^D(e) \, \mathbf{x}^j$ and $\mathbf{x}^j \, P^D(e) \, \mathbf{x}^i$; we can think of such a pair of $\mathbf{x}^i$ and $\mathbf{x}^j$ as forming a generalized WARP violation defined on the modified revealed preference relations. We let $\tilde{e}^\star$ be the supremum value of e at which $R^D(e)$ and $P^D(e)$ are free of such WARP violations.

We say that $\mathcal{D}$ can be *rationalized by a complete (but not necessarily transitive) preference*

$\succsim$ *satisfying extended local nonsatiation at cost-efficiency* e if $\mathbf{x}^i \succsim \mathbf{x}$ for all $\mathbf{x} \in \mathcal{B}^i(e)$; let $e^\star$ be the supremum value of e at which $\mathcal{D}$ can be rationalized in this manner. Notice that if $\mathcal{D}$ satisfies this property, then $R^D(e)$ and $P^D(e)$ must be free of WARP violations as defined above. Thus we may conclude that $\tilde{e}^\star \geqslant e^\star$; more generally, we have

$$\tilde{e}^\star \geqslant e^\star \geqslant e^*.$$

While we do not know of a way to calculate $e^\star$, it is straightforward to calculate $\tilde{e}^\star$, and this constitutes an upper bound of $e^\star$. In particular, if the difference between $\tilde{e}^\star$ and e^* is small or nonexistent for a particular dataset $\mathcal{D}$, then we know that a model of complete preference allowing for nontransitivity cannot not do a better job of explaining $\mathcal{D}$ than a model of complete and transitive preference.

2 FOSD-Rationalizability

In a contingent consumption environment, it is natural to strengthen the increasing requirement on U so that it is increasing with respect to first-order stochastic dominance (FOSD); by this we mean that $U(\mathbf{x}'') \geqslant U(\mathbf{x}')$ whenever $\mathbf{x}''$ (considered as a distribution through $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_S)$) first-order stochastically dominates $\mathbf{x}'$, with the inequality being strict if the dominance is strict. We refer to a utility function as FOSD-increasing whenever it is continuous and increasing with respect to FOSD; note that the latter implies that the utility function is increasing.

A dataset $\mathcal{D}$ is said to be *FOSD-rationalizable* (with respect to a given $\boldsymbol{\pi}$) if it can be rationalized by a utility function that is FOSD-increasing. An extension of Afriat's (1967) theorem by Nishimura, Ok, and Quah (2017) states that a dataset $\mathcal{D}$ is FOSD-rationalizable if and only if it satisfies a property which we refer to as FOSD-GARP. This property is a stronger version of GARP which rules out a larger set of revealed preference cycles.

For the bundles $\mathbf{x}^i$ and $\mathbf{x}^j$ in the set $\mathcal{X}$, we say that $\mathbf{x}^i$ is *directly revealed FOSD-preferred* to $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i \, R_F^D \, \mathbf{x}^j$) if there is a bundle $\mathbf{y}$ which first-order stochastically dominates $\mathbf{x}^j$ with $\mathbf{p}^i \cdot \mathbf{x}^i \geqslant \mathbf{p}^i \cdot \mathbf{y}$. If $\mathbf{y}$ can be chosen such that the latter inequality is strict, then we say that $\mathbf{x}^i$ is *directly revealed strictly FOSD-preferred to* $\mathbf{x}^j$ (and denote it

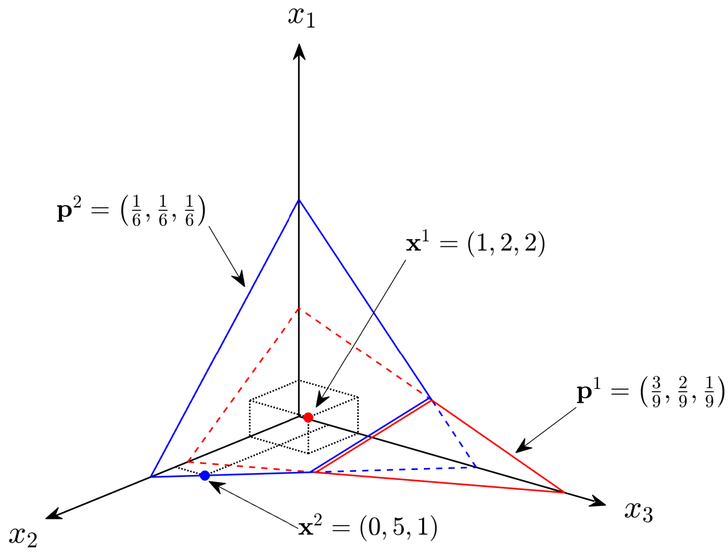


Figure 2: Violation of FOSD-Rationalizability

by $\mathbf{x}^i \, P_F^D \, \mathbf{x}^j$). We denote the transitive closure of R_F^D by R_F.

A dataset $\mathcal{D}$ satisfies FOSD-WARP if $\mathbf{x}^i \, R_F^D \, \mathbf{x}^j$ then it is not the case that $\mathbf{x}^j \, P_F^D \, \mathbf{x}^i$, and it satisfies FOSD-GARP if $\mathbf{x}^i \, R_F \, \mathbf{x}^j$ then it is not the case that $\mathbf{x}^j \, P_F^D \, \mathbf{x}^i$. Note that if $\mathbf{x}^i$ is directly revealed preferred (directly revealed strictly preferred) to $\mathbf{x}^j$, then $\mathbf{x}^i$ is also directly revealed FOSD-preferred (directly revealed strictly FOSD-preferred) to $\mathbf{x}^j$ (since we can choose $\mathbf{y} = \mathbf{x}^j$). For this reason, a dataset that satisfies FOSD-WARP must also satisfy WARP, and if it satisfies FOSD-GARP then it must satisfy GARP.

To illustrate in simple terms how FOSD-GARP works, consider Figure 2 which depicts the same two budget sets as in Figure 1, with price vectors $\mathbf{p}^1 = \left(\frac{3}{9}, \frac{2}{9}, \frac{1}{9}\right)$ and $\mathbf{p}^2 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$, and portfolio allocations $\mathbf{x}^1 = (1, 2, 2)$ and $\mathbf{x}^2 = (0, 5, 1)$. These two observations form a rationalizable but not FOSD-rationalizable dataset (assuming equiprobable states) because FOSD-GARP is violated. Indeed, $\mathbf{p}^2 \cdot \mathbf{x}^2 > \mathbf{p}^2 \cdot \mathbf{x}^1$, but it is also the case that $\mathbf{p}^1 \cdot \mathbf{x}^1 > \mathbf{p}^1 \cdot \mathbf{y}$ where $\mathbf{y} = (0, 1, 5)$ is stochastically equivalent to $(0, 5, 1) = \mathbf{x}^2$.

Rationalizability Index e^{**}. We now briefly explain how to calculate e^{**}, the CCEI for the family of FOSD-increasing utility functions. A dataset $\mathcal{D}$ is said to be *FOSD-rationalizable at cost-efficiency* e if there is an FOSD-increasing utility function $U : \mathbb{R}_+^S \to \mathbb{R}$

such that $U(\mathbf{x}^i) \geqslant U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{B}^i(e)$; by definition,

$$e^{**} = \sup\{e \in (0,1] : \mathcal{D} \text{ is FOSD-rationalizable at cost-efficiency } e\}.$$

FOSD-rationalizability at cost-efficiency e is characterized by FOSD-GARP at cost-efficiency e, the definition of which is a straightforward modification of the definition for FOSD-GARP.

Given $e \in (0,1]$, for any pair $(\mathbf{x}^i, \mathbf{x}^j) \in \mathcal{X} \times \mathcal{X}$, we say that $\mathbf{x}^i$ is *directly revealed FOSD-preferred at cost-efficiency* e *to* $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i\, R_F^D(e)\, \mathbf{x}^j$) if there is a bundle $\mathbf{y} \in \mathbb{R}^S_+$ which first-order stochastically dominates $\mathbf{x}^j$ with $e\, \mathbf{p}^i \cdot \mathbf{x}^i \geqslant \mathbf{p}^i \cdot \mathbf{y}$; and if $\mathbf{y}$ can be chosen such that the inequality is strict, then we say that $\mathbf{x}^i$ is *directly revealed strictly FOSD-preferred at cost-efficiency* e *to* $\mathbf{x}^j$ (and denote it by $\mathbf{x}^i\, P_F^D(e)\, \mathbf{x}^j$). We denote the transitive closure of $R_F^D(e)$ by $R_F(e)$.

A dataset $\mathcal{D}$ satisfies *FOSD-GARP at cost-efficiency* e if $\mathbf{x}^i\, R_F(e)\, \mathbf{x}^j$ then it is not the case that $\mathbf{x}^j\, P_F^D(e)\, \mathbf{x}^i$. Appealing to Nishimura, Ok, and Quah (2017), we know that a dataset $\mathcal{D}$ is FOSD-rationalizable at cost-efficiency e if and only if it satisfies FOSD-GARP at cost-efficiency e. This result provides a practical way of checking FOSD-GARP at cost-efficiency e and thus of computing e^{**}.

Equiprobable States. In our experiment, the states of the world are equally likely. In this case, the bundles $\mathbf{x}$ and $\mathbf{x}'$ are stochastically equivalent (i.e., have the same distribution) if the entries in one bundle are a permutation of the entries in the other, and $\mathbf{x}'$ strictly dominates $\mathbf{x}$ by FOSD if and only if there is some $\mathbf{y}$, the entries of which are a permutation of the entries in $\mathbf{x}'$, such that $\mathbf{y} > \mathbf{x}$ (in the sense of the product order).[1]

It follows from this observation that $\mathbf{x}^i$ is directly revealed FOSD-preferred (directly revealed strictly FOSD-preferred) to $\mathbf{x}^j$ at cost-efficiency e if and only if there is some $\mathbf{y}$, the entries of which are a permutation of the entries in $\mathbf{x}^j$, such that $e\, \mathbf{p}^i \cdot \mathbf{x}^i \geqslant (>)\ \mathbf{p}^i \cdot \mathbf{y}$. As a further consequence, a dataset $\mathcal{D}$ satisfies FOSD-GARP at cost-efficiency e if and only if another dataset $\mathcal{D}'$ (a derivative of $\mathcal{D}$) satisfies GARP at cost-efficiency e. The dataset $\mathcal{D}'$ is the *symmetric augmentation of* $\mathcal{D}$ and it is defined as follows: $(\breve{\mathbf{p}}^k, \breve{\mathbf{x}}^k) \in \mathcal{D}'$ if the entries of $\breve{\mathbf{p}}^k$ and $\breve{\mathbf{x}}^k$ are corresponding permutations of the entries in $\mathbf{p}^i$ and $\mathbf{x}^i$ for some

[1] For any $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^S_+ \times \mathbb{R}^S_+$, we say that $\mathbf{x} \geqslant \mathbf{y}$ if $x_s \geqslant y_s$ for all s; and $\mathbf{x} > \mathbf{y}$ if $\mathbf{x} \geqslant \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$.

observation i (in $\mathcal{D}$). For example, if $S = 2$ and $\mathcal{D}$ has 50 observations, then its symmetric augmentation $\mathcal{D}'$ will have 100 observations consisting of the elements in $\mathcal{D}$ as well as the elements $((p_2^i, p_1^i), (x_2^i, x_1^i))$ for $i = 1, 2, \ldots, 50$; $\mathcal{D}$ satisfies FOSD-GARP at cost-efficiency e if and only if $\mathcal{D}'$ satisfies GARP at cost-efficiency e.

Equivalence of FOSD-WARP and FOSD-GARP. We know that, in general, $\mathcal{D}$ is FOSD-rationalizable if and only if it satisfies FOSD-GARP. We claim that when there are just two states of the world (so $S = 2$), *$\mathcal{D}$ is FOSD-rationalizable if and only if it satisfies FOSD-WARP*. Indeed, if $\mathcal{D}$ satisfies FOSD-WARP, then its symmetric augmentation $\mathcal{D}'$ must satisfy WARP. By Banerjee and Murphy (2006) (which builds on Rose (1958)), this guarantees that $\mathcal{D}'$ satisfies GARP and thus $\mathcal{D}$ satisfies FOSD-GARP.

Symmetry and Local Nonsatiation. We say that a preference $\succsim$ is *symmetric* if whenever the entries of $\mathbf{y}$ are a permutation of the entries in $\mathbf{x}$, then $\mathbf{x} \succsim \mathbf{y}$ and $\mathbf{y} \succsim \mathbf{x}$; in other words, the agent with such a preference is indifferent between the bundles $\mathbf{x}$ and $\mathbf{y}$. It is clear that, when the states are equiprobable, the preference induced by an FOSD-increasing utility function must be increasing and symmetric; the latter holds because any two bundles related by a permutation are stochastically equivalent and must have the same utility. It is not hard to check that the converse is also true. Thus *when the states are equiprobable, U is FOSD-increasing if and only if it is symmetric and increasing.*

In the main paper (see Footnote 12), we claim that the explanatory power of the choice acclimating personal equilibrium model of Kőszegi and Rabin (2007) does not go beyond that of the family of FOSD-increasing functions when the states are equiprobable. To understand why, we first consider a preference $\succsim$ that is complete, transitive, symmetric, and locally nonsatiated. Then it is straightforward to check that $\mathbf{x}^i \, R_F^D(e) \, \mathbf{x}^j$ implies $\mathbf{x}^i \succsim \mathbf{x}^j$ and $\mathbf{x}^i \, P_F^D(e) \, \mathbf{x}^j$ implies $\mathbf{x}^i \succ \mathbf{x}^j$. This in turn implies that if a dataset $\mathcal{D}$ is generated by such a preference, then it must satisfy FOSD-GARP at cost-efficiency e. On the other hand, we know (from Nishimura, Ok, and Quah (2017)) that whenever $\mathcal{D}$ satisfies FOSD-GARP at cost-efficiency e then it is rationalizable by an FOSD-increasing utility function—such a function is symmetric and increasing, and not just locally nonsatiated.[2]

[2]The inability to separate locally nonsatiated from increasing utility functions in budgetary data is a well-known phenomenon in the revealed preference literature, and it arises (in essence) from the fact the

In choice acclimating personal equilibrium (Kőszegi and Rabin, 2007), even though the utility function need not be FOSD-increasing, it is locally nonsatiated and, when the states are equiprobable, it is also symmetric. If a data set $\mathcal{D}$ can be rationalized at cost-efficiency e by a utility function in this family, then it will obey FOSD-GARP at cost-efficiency e, and thus it can also be rationalized by an FOSD-increasing utility function at cost-efficiency e.

FOSD-Rationalizability with Two Asymmetric States. In the experiment conducted by Choi *et al.* (2007), some subjects were required to make choices in a consumption space where the two states have known but unequal probabilities. We have opted not to include such a design in our experiment. The main reason for this is that when there are two (or more) asymmetric states, there is a sense in which the test for FOSD-rationalizability is overly permissive. To be precise, in the two-state case, a dataset is FOSD-rationalizable with state 1 having probability $\beta > \frac{1}{2}$ if and only if it is FOSD-rationalizable with state 1 having *any* probability $\beta' > \frac{1}{2}$. This is because a continuous function U is FOSD-increasing with state 1 having probability $\beta > \frac{1}{2}$ if and only if U is increasing and $U(a,b) > U(b,a)$ whenever $a > b$; the latter condition is obviously independent of the precise value of β. On the other hand, the test of EUT-rationaliability in this environment *can* distinguish between state 1 having probability (say) $\frac{2}{3}$ and state 1 having probability (say) $\frac{3}{4}$. This makes comparing the relative performance of the two models problematic. That said, we have examined the data with asymmetric probabilities in two states and the results are broadly similar to the case of symmetric probabilities (see Appendix II, Section 1.2).

3 EUT-Rationalizability

A utility function $U : \mathbb{R}^S_+ \to \mathbb{R}$ has the expected utility form if there is a continuous and strictly increasing Bernoulli index $u : \mathbb{R}_+ \to \mathbb{R}$ such that

$$U(\mathbf{x}) = \pi_1 u(x_1) + \pi_2 u(x_2) + \cdots + \pi_S u(x_S),$$

for all $\mathbf{x} \in \mathbb{R}^S_+$. A dataset $\mathcal{D}$ is said to be EUT-rationalizable if it is rationalizable by a utility function taking the expected utility form.

budget sets are downward comprehensive, i.e., $\mathcal{B}^i(e)$ has the property that if $\mathbf{x} \in \mathcal{B}^i(e)$ then $\mathbf{y} \in \mathcal{B}^i(e)$ for all $\mathbf{y} \in \mathbb{R}^S_+$, such that $\mathbf{y} \leqslant \mathbf{x}$.

Rationalizability Index e^{***}. A dataset $\mathcal{D}$ is said to be *EUT-rationalizable at cost-efficiency* e if there is a utility function $U : \mathbb{R}^S_+ \to \mathbb{R}$ with the expected utility form such that $U(\mathbf{x}^i) \geqslant U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{B}^i(e)$. By definition, e^{***} is the CCEI when we confine rationalization to the family of utility functions of the expected utility form, i.e.,

$$e^{***} = \sup\{e \in (0,1] : \mathcal{D} \text{ is EUT-rationalizable at cost-efficiency } e\}.$$

The index e^{***} can be calculated so long as there is a computationally straightforward way of testing whether $\mathcal{D}$ is EUT-rationalizable at cost-efficiency e. Such a test is provided by Polisson, Quah, and Renou (2020), which shows that, for U having certain properties (the expected utility form is a prime example of such properties), to determine whether or not there is such a U rationalizing the dataset, it suffices to check the optimality of the chosen bundle $\mathbf{x}^i$ against a *finite* set of alternative bundles in $\mathcal{B}^i(e)$ (rather than against all alternative bundles in $\mathcal{B}^i(e)$). For this reason, the method is referred to as the *generalized restriction of infinite domains* (GRID). Unlike GARP or FOSD-GARP tests, GRID tests do not involve checking for the absence of (variously defined) revealed preference cycles, but they are fully nonparametric, in the sense that the Bernoulli index u is allowed to be any strictly increasing and continuous function.

Given a dataset $\mathcal{D}$, let $\mathcal{Y}$ be the finite set that contains any demand level observed in $\mathcal{D}$ plus zero, that is

$$\mathcal{Y} := \{x \in \mathbb{R}_+ : x = x^i_s \text{ for some } i \text{ and } s\} \cup \{0\}.$$

We then form the finite grid $\mathcal{G} = \mathcal{Y}^S \subset \mathbb{R}^S_+$ which is a restriction of the consumption space $\mathbb{R}^S_+$ to allocations comprised of demand levels that have been observed in the dataset $\mathcal{D}$.

We claim that if $\mathcal{D}$ is EUT-rationalizable at cost-efficiency e then there are real numbers $\bar{u}(y)$ (associated with each $y \in \mathcal{Y}$), with

$$\bar{u}(y') > \bar{u}(y) \text{ whenever } y' > y, \tag{1}$$

and at each observation i,

$$\sum_{s=1}^{S} \pi_s \bar{u}(x_s^i) \geqslant \sum_{s=1}^{S} \pi_s \bar{u}(x_s) \text{ for any } \mathbf{x} \in \mathcal{G} \text{ such that } \mathbf{p}^i \cdot \mathbf{x} \leqslant e\, \mathbf{p}^i \cdot \mathbf{x}^i, \tag{2}$$

$$\sum_{s=1}^{S} \pi_s \bar{u}(x_s^i) > \sum_{s=1}^{S} \pi_s \bar{u}(x_s) \text{ for any } \mathbf{x} \in \mathcal{G} \text{ such that } \mathbf{p}^i \cdot \mathbf{x} < e\, \mathbf{p}^i \cdot \mathbf{x}^i. \tag{3}$$

Indeed, suppose $\mathcal{D}$ can be EUT-rationalized at cost-efficiency e by an expected utility function with the continuous and increasing Bernoulli index u, then these conditions must hold if we choose $\bar{u}(y) = u(y)$ for each $y \in \mathcal{Y}$ since, in the case of the first condition, $\mathbf{x}$ is in $\mathcal{B}^i(e)$ and in the case of the second condition, $\mathbf{x}$ is in the interior of $\mathcal{B}^i(e)$.

An application of the main result of Polisson, Quah, and Renou (2020) says that these conditions are also *sufficient* for EUT-rationalizability at cost-efficiency e; more formally, if, for a dataset $\mathcal{D}$, there exists $\bar{u}(y)$ for each $y \in \mathcal{Y}$ such that (1), (2), and (3) hold, then $\mathcal{D}$ is EUT-rationalizable at cost-efficiency e. Note that the conditions constitute a finite set of linear inequalities and ascertaining whether or not it has a solution is computationally straightforward. This gives us a feasible way of determining whether a dataset is EUT-rationalizable at cost-efficiency e and thus a way to calculate e^{***}.

To illustrate how this test works, we consider the dataset depicted in Figure 3. The budget sets are the same as in Figures 1 and 2, with price vectors $\mathbf{p}^1 = \left(\frac{3}{9}, \frac{2}{9}, \frac{1}{9}\right)$ and $\mathbf{p}^2 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$, and portfolio allocations $\mathbf{x}^1 = (1, 2, 2)$ and $\mathbf{x}^2 = (3, 1, 2)$. Assuming that the three states are equiprobable, it is easy to verify that these choices are FOSD-rationalizable, but we claim that they are not EUT-rationalizable. To see this, consider the portfolio allocations $\mathbf{y} = (1, 1, 3)$ and $\mathbf{z} = (2, 2, 2)$ and notice that

$$\mathbf{p}^1 \cdot \mathbf{x}^1 > \mathbf{p}^1 \cdot \mathbf{y} \text{ and } \mathbf{p}^2 \cdot \mathbf{x}^2 = \mathbf{p}^2 \cdot \mathbf{z}.$$

EUT-rationalizability requires that there is a Bernoulli index u such that

$$\tfrac{1}{3}u(1) + \tfrac{1}{3}u(2) + \tfrac{1}{3}u(2) = U(\mathbf{x}^1) > U(\mathbf{y}) = \tfrac{1}{3}u(1) + \tfrac{1}{3}u(1) + \tfrac{1}{3}u(3),$$

$$\tfrac{1}{3}u(3) + \tfrac{1}{3}u(1) + \tfrac{1}{3}u(2) = U(\mathbf{x}^2) \geqslant U(\mathbf{z}) = \tfrac{1}{3}u(2) + \tfrac{1}{3}u(2) + \tfrac{1}{3}u(2),$$

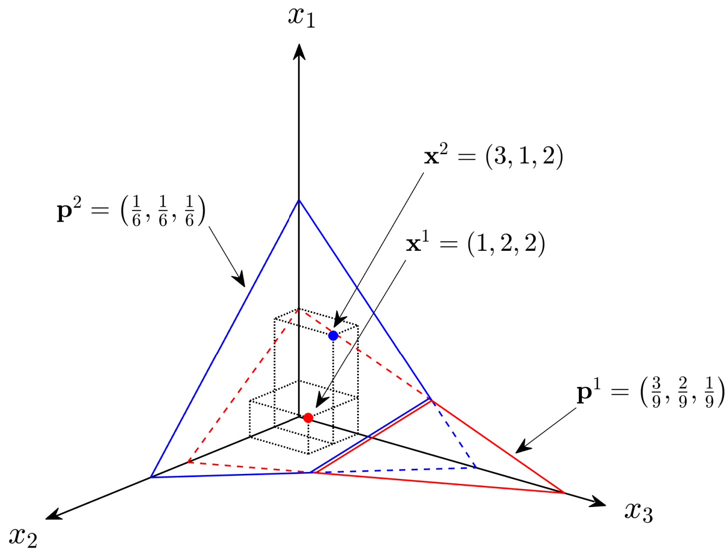


Figure 3: Violation of EUT-Rationalizability

implying that $2u(2) > u(1) + u(3)$ and $u(3) + u(1) \geqslant 2u(2)$, a contradiction. The GRID test would also reveal this violation of EUT-rationalizability, since $\mathcal{Y} = \{0, 1, 2, 3\}$ and $\mathbf{y}$ and $\mathbf{z}$ are both in $\mathcal{G}$.

4 Rank-Dependent Utility and Disappointment Aversion Models

We briefly describe the rank-dependent utility and disappointment aversion models in the case where the states are equally likely. In our experiment, there are three equally likely states and a preference over bundles in the contingent consumption space is consistent with expected utility theory if and only if there is a continuous and increasing *Bernoulli index* $u : \mathbb{R}_+ \to \mathbb{R}$, such that the preference is represented by $U : \mathbb{R}^3_+ \to \mathbb{R}$, where $U(\mathbf{x}) = \sum_{s=1}^{3} u(x_s)$. In the rank-dependent utility model (Quiggin, 1982, 1993), the utility of $\mathbf{x} = (x_1, x_2, x_3)$ when $x_1 \leqslant x_2 \leqslant x_3$ is

$$U(\mathbf{x}) = \alpha' u(x_1) + \alpha'' u(x_2) + \alpha''' u(x_3),$$

where α', α'', $\alpha''' > 0$; more generally, $U(\mathbf{x}) = U(\zeta(\mathbf{x}))$, where $\zeta(\mathbf{x})$ refers to a permutation of the entries of $\mathbf{x}$, so that the entries of $\zeta(\mathbf{x})$ are weakly increasing. With three or more states, the disappointment aversion model (Gul, 1991) defines $U(\mathbf{x})$ implicitly as the unique solution to

$$U(\mathbf{x}) = \frac{\sum_{\{s:u(x_s)\geqslant U(\mathbf{x})\}} u(x_s) + \beta \sum_{\{s:u(x_s)<U(\mathbf{x})\}} u(x_s)}{\#\{s : u(x_s) \geqslant U(\mathbf{x})\} + \beta\#\{s : u(x_s) < U(\mathbf{x})\}}$$

for $\beta > 1$. This utility function is increasing and symmetric, hence FOSD-increasing.

As we note in the main paper, the rank-dependent utility and disappointment aversion models are identical when there are two equiprobable states in the sense that they represent the same family of preferences over (two-state) contingent consumption bundles. Indeed, it is not difficult to check that, in both cases, the preference has the representation

$$U(x_1, x_2) = u(\max\{x_1, x_2\}) + \beta u(\min\{x_1, x_2\}).$$

When there are three states of the world, these two models generate distinct families of preferences on the contingent consumption space. For a quick way of seeing this, consider the utility function in a neighborhood of the bundle $(1, 3, 3)$. In the case of rank-dependent utility, the probabilities are distorted and the bundle $(1, y, z)$ where y and z are close to 3 will have utility $\alpha' u(1) + \alpha'' u(y) + \alpha''' u(z)$ when $y < z$ and utility $\alpha' u(1) + \alpha''' u(y) + \alpha'' u(z)$ when $y > z$. On the plane $\mathcal{P} = \{(1, y, z) : y,\ z \in \mathbb{R}_+\}$, the indifference curve of the rank-dependent utility function will have a *kink* at $(1, 3, 3)$, provided $\alpha'' \neq \alpha'''$. On the other hand, in the case of disappointment aversion, the utility of $(1, y, z)$ has the form $\beta u(1) + [u(y) + u(z)]$ provided y and z are sufficiently close to 3 and irrespective of whether y is greater or smaller than z. Thus, on the plane $\mathcal{P}$, the indifference curve through $(1, 3, 3)$ is *smooth* at $(1, 3, 3)$, provided u is smooth.

References

Afriat, S. N. 1967. "The Construction of Utility Functions from Expenditure Data." *International Economic Review* 8(1): 67–77.

———. 1973. "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method." *International Economic Review* 14(2): 460–472.

Banerjee, S. and J. H. Murphy. 2006. "A Simplified Test for Preference Rationality of Two-Commodity Choice." *Experimental Economics* 9: 67–75.

Choi, S., R. Fisman, D. Gale, and S. Kariv. 2007. "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review* 97(5): 1921–1938.

Debreu, G. 1954. "Representation of a Preference Ordering by a Numerical Function." In *Decision Processes*, edited by R. M. Thrall, C. H. Coombs, and R. L. Davis. New York: John Wiley and Sons, 159–165.

Gul, F. 1991. "A Theory of Disappointment Aversion." *Econometrica* 59(3): 667–686.

Kőszegi, B. and M. Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review* 97(4): 1047–1073.

Nishimura, H., E. Ok, and J. K.-H. Quah. 2017. "A Comprehensive Approach to Revealed Preference Theory." *American Economic Review* 107(4): 1239–1263.

Polisson, M., J. K.-H. Quah, and L. Renou. 2020. "Revealed Preferences over Risk and Uncertainty." *American Economic Review* 110(6): 1782–1820.

Quiggin, J. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3(4): 323–343.

———. 1993. *Generalized Expected Utility Theory: The Rank-Dependent Model.* Dordrecht: Kluwer.

Rose, H. 1958. "Consistency of Preference: The Two-Commodity Case." *Review of Economic Studies* 25(2): 124–125.

Appendix II : Further Empirical Analysis

The majority of the empirical analysis in Appendix II essentially replicates the empirical findings in the main paper, but using data drawn from two-dimensional (rather than three-dimensional) choice environments. For the vast majority of two-dimensional subjects, the states of the world occur with equal probabilities (are symmetric), but we also analyze the choices of a small cohort of two-dimensional subjects for whom the state probabilities occur with unequal probabilities (are asymmetric). In addition, we include the details of a limited empirical analysis using distance-based indices, which provide an alternative notion of error to the more standard cost-efficiency indices—despite using a different metric, the results from this analysis are very much in line with the empirical findings and conclusions drawn using cost-efficiency indices. Finally, we provide some further analysis on revealed preference cycles, in particular cycles with respect to FOSD; the results reinforce the conclusions drawn in the main paper involving more standard revealed preference cycles.

1 Two-Dimensional Data

In this section, we report our analysis of the data from 1,002 subjects making portfolio decisions in two-dimensional environments. The data include data collected by Choi *et al.* (2007), similar data using subject pools collected by Zame *et al.* (2020) and Cappelen *et al.* (2023), as well as new data. These experiments are identical to the one in the main paper, except that there are two states rather than three. For each subject, we have a set of 50 observations $\mathcal{D} := \{(\mathbf{p}^i, \mathbf{x}^i)\}_{i=1}^{50}$, where $\mathbf{p}^i = (p_1^i, p_2^i)$ denotes the i-th observation of the price vector and $\mathbf{x}^i = (x_1^i, x_2^i)$ denotes the corresponding allocation. For the vast majority of subjects—956 subjects to be precise—the two states are equally likely. There is also a small pool of 46 subjects (Choi *et al.*, 2007) where the states occur with unequal probabilities.

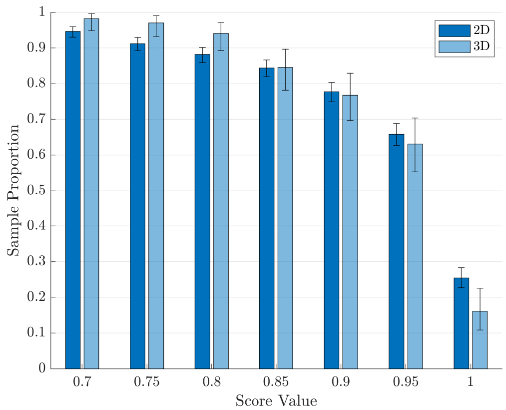
The main conclusion is that the results obtained from the two-dimensional choice environment are broadly similar to those obtained from the three-dimensional choice environment. In particular, violations of basic ordering and monotonicity (with respect to FOSD) are just as prominent, and in many cases much more so, than violations of properties specific to EUT (such as the independence axiom).

1.1 Symmetric States

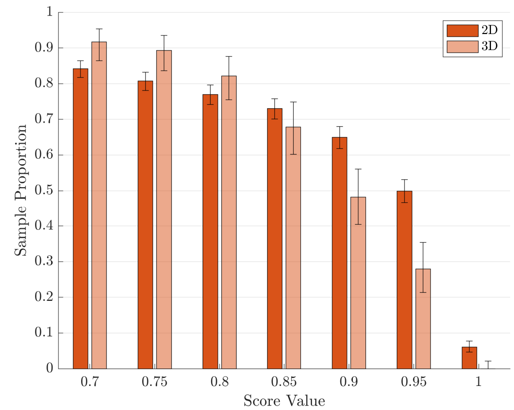
Rationalizability Scores. Figure 1 compares the rationalizability scores across two-dimensional and three-dimensional choice experiments for e^* (Figure 1a), e^{**} (Figure 1b), and e^{***} (Figure 1c). It might seem intuitive that making portfolio decisions across three states is more cognitively challenging than making such decisions across two states, and so one could reasonably expect rationalizability scores to be lower in three-dimensional settings than in two-dimensional settings. Broadly speaking, this is what we find, especially at higher rationalizability score values. However, it is somewhat reassuring that the fall in rationalizability scores is not too large—this suggests that subjects did not have major difficulties in understanding the three-dimensional procedures or using the computer interface (over and above those in the two-dimensional context, which have been implemented and tested extensively across many domains of choice).

With respect to basic rationalizability (e^*), we find that the results are broadly similar across experiments. In the three-dimensional (resp. two-dimensional) experiment, 63.1 (resp. 65.8) percent of subjects have e^* scores above 0.95, and 76.8 (resp. 77.7) percent have scores above 0.9. On the other hand, we find that choices from three-dimensional budget sets are distinctly less FOSD-rationalizable (e^{**}) and EUT-rationalizable (e^{***}) than choices from two-dimensional budget lines (Figures 1b and 1c). In the three-dimensional experiment, 28.0 (resp. 16.1) percent of subjects have e^{**} (resp. e^{***}) scores above 0.95, and 48.2 (resp. 36.9) percent have scores above 0.9. In the two-dimensional experiment, the corresponding percentages are 49.9 (resp. 46.4) and 65.0 (resp. 63.4).

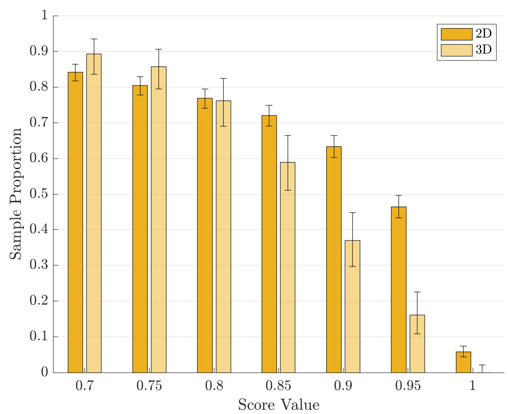
Difference-in-Differences. In the two-dimensional data, as in the three-dimensional data, the loss of consistency arising from EUT is relatively small after accounting for consistency with basic ordering and monotonicity with respect to FOSD. To be precise, $1 - e^{**} > e^{**} - e^{***}$ for 827 out of 956 subjects (86.5 percent). Following the analysis in the main paper, for each subject we randomly draw 1,000 subsets of 25 (out of 50) observations where each subset is drawn without replacement; we then calculate (across 25 observations) 1,000 pairs of scores $(e^{\dagger\dagger}, e^{\dagger\dagger\dagger})$ for FOSD-rationalizability ($e^{\dagger\dagger}$) and EUT-rationalizability ($e^{\dagger\dagger\dagger}$) and denote the mean scores within this sample by $\bar{e}^{\dagger\dagger}$ and $\bar{e}^{\dagger\dagger\dagger}$, respectively.



(a) Distributions of e^*



(b) Distributions of e^{**}



(c) Distributions of e^{***}

Figure 1: Distributions of Rationalizability Scores

The plots depict distributions of rationalizability scores across the two-dimensional (2D) and three-dimensional (3D) experiments for the rationalizability scores e^* (a), e^{**} (b), and e^{***} (c). The bars represent one minus the cumulative distributions of e^*, e^{**}, and e^{***} at each score value. The braces represent exact 95 percent confidence intervals on the proportions.

As a prelude to the statistical analysis, Figure 2 shows scatterplots of e^{**} against $\bar{e}^{\dagger\dagger}$ (Figure 2a), and e^{***} against $\bar{e}^{\dagger\dagger\dagger}$ (Figure 2b). The scores depicted in each panel of Figure 2 are very highly correlated—the correlation coefficients between e^{**} and $\bar{e}^{\dagger\dagger}$, and between e^{***} and $\bar{e}^{\dagger\dagger\dagger}$, are 0.977 and 0.978, respectively. Following the analysis in the main paper, for each subject, the null hypothesis is given by

$$H_0 : D_1 = (1 - \mu_{e^{\dagger\dagger}}) - (\mu_{e^{\dagger\dagger}} - \mu_{e^{\dagger\dagger\dagger}}).$$

That is, that there is no difference between $1 - \mu_{e^{\dagger\dagger}}$ (the mean difference between perfect

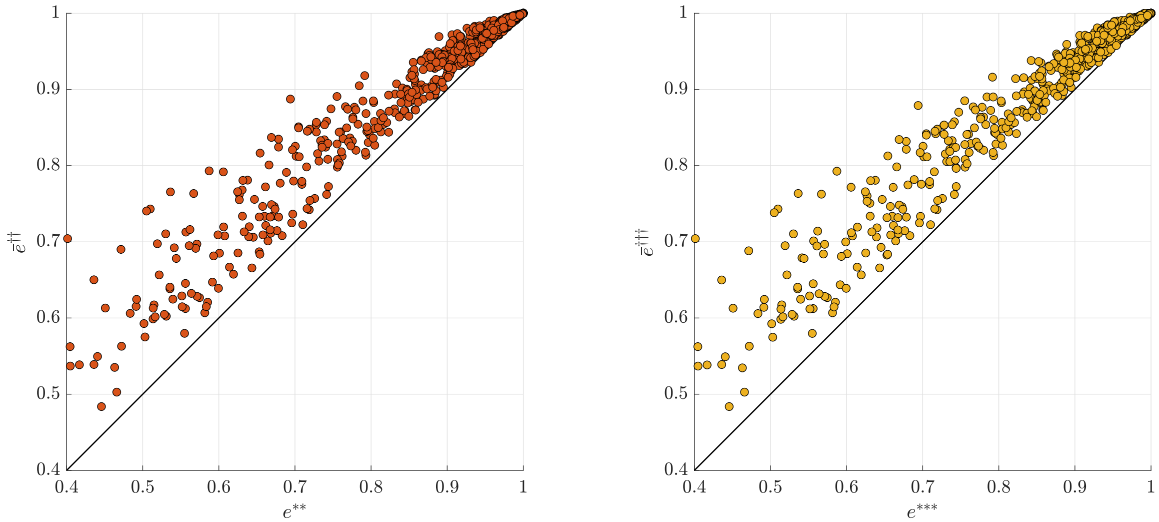


Figure 2: Rationalizability Score Correlations

The plots depict rationalizability scores for individual subjects. By definition, $\bar{e}^{\dagger\dagger} \geqslant e^{**}$ and $\bar{e}^{\dagger\dagger\dagger} \geqslant e^{***}$, so all points in the scatterplots must lie on or above the 45-degree line.

rationalizability and FOSD-rationalizability) and $\mu_{e^{\dagger\dagger}} - \mu_{e^{\dagger\dagger\dagger}}$ (the mean difference between FOSD-rationalizability and EUT-rationalizability). For each subject, we use the 1,000 pairs of scores $(e^{\dagger\dagger}, e^{\dagger\dagger\dagger})$ to calculate the difference between $1 - \bar{e}^{\dagger\dagger}$ and $\bar{e}^{\dagger\dagger} - \bar{e}^{\dagger\dagger\dagger}$, denoted by $\bar{D}_1$, which is an estimator for D_1. To obtain the distribution of the estimator $\bar{D}_1$, we adopt a simple bootstrapping procedure. This involves re-sampling (with replacement) from the original sample consisting of 1,000 pairs of scores $(e^{\dagger\dagger}, e^{\dagger\dagger\dagger})$ to create a bootstrapped sample of the same size; we repeat this procedure 10^6 times.

Out of 956 subjects, $\bar{D}_1$ is positive and statistically significant at the 1 percent level for 850 subjects (88.9 percent). The $\bar{e}^{\dagger\dagger}$ scores for 686 (71.8 percent) and 570 (59.6 percent) of our subjects are above 0.9 and 0.95, respectively. Even out of those highly FOSD-rationalizable subjects, $\bar{D}_1$ is positive and statistically significant at the 1 percent significance level for 580 (84.5 percent) and 465 (81.6 percent) subjects, respectively. Finally, we again strengthen our difference-in-differences estimator by using a "double-differencing" strategy under the null hypothesis is that $1 - e^{\dagger\dagger}$ is *twice* as large as $e^{\dagger\dagger} - e^{\dagger\dagger\dagger}$. We let $\bar{D}_2$ be the mean of this difference-in-differences (which is an estimator for D_2 defined analogously to D_1) and find that it is positive and statistically significant at the 1 percent level for 819 out of 956 subjects (85.7 percent). In Figure 3, individual subjects are depicted in red if both $\bar{D}_1$ and $\bar{D}_2$ are

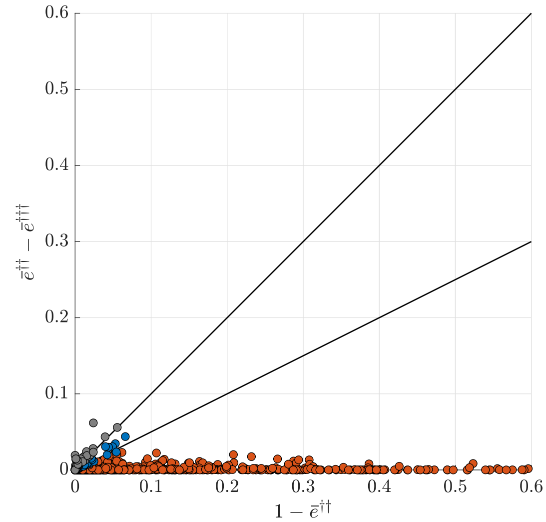


Figure 3: Scatterplot of Rationalizability Score Differences

The plot depicts rationalizability score differences for individual subjects. The differences between $1 - \bar{e}^{\dagger\dagger}$ and $2(\bar{e}^{\dagger\dagger} - \bar{e}^{\dagger\dagger\dagger})$ are positive and statistically significant for 85.7 percent of the sample (red). The differences between $1 - \bar{e}^{\dagger\dagger}$ and $\bar{e}^{\dagger\dagger} - \bar{e}^{\dagger\dagger\dagger}$ are positive and statistically significant for 88.9 percent of the sample (red and blue combined).

positive and statistically significant at the 1 percent level, in blue if only $\bar{D}_1$ is positive and statistically significant, and in gray if neither.

1.2 Asymmetric States

We also replicate our empirical analysis on the 46 subjects in the *asymmetric* treatment of the two-dimensional experiment in Choi *et al.* (2007). This treatment was identical to the symmetric treatment, except that one of the states occurred with probability 2/3 (instead of 1/2). Note that, for each subject, this asymmetric distribution was maintained throughout the experiment, i.e., if, for a given subject, state 1 occurred with probability 2/3 and state 2 with probability 1/3, then it remained that way across all 50 observations.[1]

At a score value of 0.95, the number of subjects (out of 46) that are FOSD-rationalizable and EUT-rationalizable is 26 (56.5 percent) and 12 (26.1 percent), respectively. While the overall number of subjects in the asymmetric treatment is small, it is still useful to draw a comparison with the (much larger) symmetric counterpart. In the symmetric case, the proportions that are FOSD-rationalizable and EUT-rationalizable at a score value of

[1]To be precise, $\boldsymbol{\pi} = (1/3, 2/3)$ for 17 subjects and $\boldsymbol{\pi} = (2/3, 1/3)$ the other 29 subjects.

0.95 are 65.0 percent and 63.4 percent, respectively—so the asymmetric treatment has a *higher* pass rate for FOSD-rationalizability and a *lower* pass rate for EUT-rationalizability.[2] Thus, in contrast to the symmetric case, there is a significant proportion of subjects in the asymmetric treatment that are FOSD-rationalizable but not EUT-rationalizable. However, while the proportion of FOSD-rationalizable subjects is higher in the asymmetric treatment, there is still a sizable proportion that are not FOSD-rationalizable.

To investigate the issue more carefully, we again perform a difference-in-differences analysis. For 16 out of 46 asymmetric subjects, $\bar{D}_1$ is positive and statistically significant at the 1 percent level. Moreover, for 6 subjects, $\bar{D}_1$ is statistically insignificant at the 1 percent level. Therefore, for roughly half of all asymmetric subjects (22 out of 46 to be precise), the additional error requited to guarantee EUT-rationalizability is less than or no different from what is requied to guarantee FOSD-rationalizability in the first place. Although less stark than in the symmetric case, our main qualitative conclusion from the analysis of the symmetric experiments continues to hold: for around half of all subjects, violations of ordering and monotonicity with respect to FOSD are just as significant as departures from the independence axiom in explaining departures from EUT.

To conclude this subsection, we note that a choice environment with asymmetric state probabilities is a very permissive one in which to test FOSD-rationalizability—for example, if state 1 had not occurred with probability 2/3 but had instead occurred with some other probability $\pi_1 > 1/2$, this distinction could not have been detected by a test of FOSD-rationalizability (as we discuss in detail in Appendix I). This likely explains why (as reported above) the distribution of e^{**} scores is higher in the asymmetric case than in the symmetric case. On the other hand, a data set that is EUT-rationalizable with $\pi_1 = 2/3$ is not necessarily rationalizable for any other $\pi_1 > 1/2$. So the choice environment remains highly discriminating with respect to EUT-rationalizability. For this reason, the wider gap between e^{**} and e^{***} in the asymmetric environment is not entirely surprising and may be an artifact of the asymmetric environment.[3]

[2]If instead we use a score value of 0.9, the pattern is similar. In the asymmetric treatment, the number of subjects that are FOSD-rationalizable and EUT-rationalizable is 33 (71.7 percent) and 18 (39.1 percent), respectively. In the symmetric case, the same proportions are 65.0 percent and 63.4 percent, respectively. Once again, the asymmetric treatment gives a higher pass rate for FOSD-rationalizability and a lower pass rate for EUT-rationalizability.

[3]As an illustration, recall that 18 out of 46 subjects (39.1 percent) have e^{***} exceeding 0.9. We could

2 Distance-Based Indices

In this section, we provide a limited empirical implementation using distance-based indices (leveraging the recent results in Hu *et al.* (2021)). The approach could be thought of as allowing for literal "hand trembles", or perturbations to portfolio allocations along the budgets—so a data set may fail to comply with basic rationalizability (or FOSD-rationalizability/EUT-rationalizability), but if we put a "ball" around each observed portfolio allocation (intersected with the corresponding budget plane), then for a sufficiently large ball we can find *another data set* which *is* rationalizable (or FOSD-rationalizable/EUT-rationalizable). We should note that this concept corresponds to the classical demand approach which incorporates an additive error component—see, for example, Varian (1985) which characterizes the hard (in a computational sense) problem of measurement error.

Relying upon Hu *et al.* (2021), we are able to calculate distance-based indices for rationalizability (k^*) and FOSD-rationalizability (k^{**}), which is an improvement given the difficulty of the existing problem. However, we are unable to calculate distance-based indices for EUT-rationalizability (k^{***}), which we require for our difference-in-differences analysis. But we can calculate upper bounds on these indices, through an intelligent "guess and check" approach. These upper bounds work against the conclusions that we ultimately draw.

Due to computational constraints, we divide each subject's 50 observations into five subsets of 10 observations each. For each of these "mini datasets" we calculate an upper bound on the index for EUT-rationalizability. We then obtain, for each subject and on each subset of 10 observations, the index for FOSD-rationalizability (k^{**}) and an upper bound on the index for EUT-rationalizability (k_u^{***}). Taking an average over the five subsets of 10 observations, we obtain for each subject, $\bar{k}^{**}$ and $\bar{k}_u^{***}$ (where $\bar{k}^{***} \leqslant \bar{k}_u^{***}$).

The results are shown in Figure 4. Out of our 168 subjects, we find that $2(\bar{k}_u^{***} - \bar{k}^{**}) < \bar{k}^{**}$ (hence $2(\bar{k}^{***} - \bar{k}^{**}) < \bar{k}^{**}$ since $\bar{k}^{***} \leqslant \bar{k}_u^{***}$) for 124 subjects (73.8 percent),

similarly calculate the EUT rationalizability score, not at the true probability distribution, but at some other distribution where the state occurring with probability 2/3 has a probability strictly greater than 1/2. We perform such an exercise: for each subject, we calculate the EUT-rationalizability score for a range of values. We denote the highest value among these scores by $\breve{e}^{***}$. By definition, we must have $e^{**} \geqslant \breve{e}^{***} \geqslant e^{***}$ for each subject (where e^{***} is calculated at the true distribution). Indeed, 33 subjects (71.7 percent) have $\breve{e}^{***}$ exceeding 0.9 as opposed to 18 subjects (39.1 percent) with e^{***} exceeding 0.9; note that 33 is also the number of subjects with e^{**} exceeding 0.9.

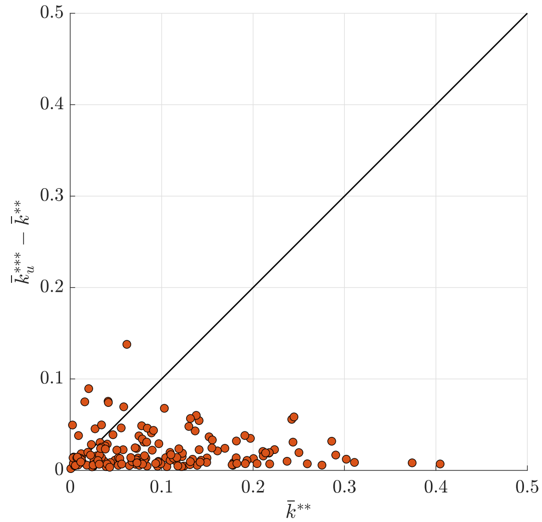


Figure 4: Scatterplot of Distance-Based Indices

The plot depicts distance-based indices for individual subjects: the index for FOSD-rationalizability ($\bar{k}^{**}$) against the difference between (the upper bound on) EUT-rationalizability and FOSD-rationalizability ($\bar{k}_u^{***} - \bar{k}^{**}$).

and $4(\bar{k}_u^{***} - \bar{k}^{**}) < \bar{k}^{**}$ for 99 subjects (58.9 percent). In other words, for the typical subject, the additional distance required to guarantee EUT-rationalizability beyond FOSD-rationalizability is relatively modest.

3 FOSD-WARP and FOSD-GARP

To complement the analysis of WARP and GARP in the main paper, we provide a basic description of individual-level revealed FOSD-preference violations. Table 1 reports percentile values of the numbers of FOSD-WARP violations, FOSD-GARP violations, and FOSD-GARP violations that do not contain a FOSD-WARP violation, alongside percentile values of the CCEI scores required to remove all violations of FOSD-WARP and FOSD-GARP. The number of FOSD-GARP violations is the number of distinct revealed FOSD-preference cycles. A FOSD-WARP violation is a special case of a FOSD-GARP violation involving a pairwise or within-observation revealed FOSD-preference cycle.

We see from Table 1 that the median number of FOSD-WARP violations is 30 and the median number of FOSD-GARP violations is 1,646. While the number of FOSD-GARP violations is very high among a considerable number of subjects, the vast majority of these FOSD-GARP violations contain FOSD-WARP violations. In fact, only 6 subjects (3.6 per-

		Violations			CCEI Scores	
		FOSD-WARP	FOSD-GARP	FOSD-GARP \ FOSD-WARP	FOSD-WARP	FOSD-GARP
Percentile Values	1	3	5	0	0.501	0.501
	5	7	13	0	0.656	0.656
	10	9	19	0	0.722	0.722
	25	18	45	0	0.830	0.830
	50	30	1,646	0	0.898	0.898
	75	50	$\geqslant 10^7$	0	0.957	0.957
	90	82	$\geqslant 10^7$	0	0.981	0.981
	95	124	$\geqslant 10^7$	0	0.989	0.989
	99	223	$\geqslant 10^7$	0	0.999	0.999

Table 1: FOSD-WARP and FOSD-GARP

The table reports percentile values of the numbers of FOSD-WARP violations, FOSD-GARP violations, and FOSD-GARP violations that do not contain a FOSD-WARP violation, alongside percentile values of the CCEI scores required to remove all violations of FOSD-WARP and FOSD-GARP. Due to computational constraints, we are unable to compute the number of FOSD-GARP violations that do not contain a FOSD-WARP violation for 6 subjects (3.6 percent), each of whom has more than 10^7 FOSD-GARP violations (even after removing FOSD-WARP violations); for these subjects, we obtain a lower bound on the number of FOSD-GARP violations that do not contain a FOSD-WARP violation.

cent) have violations of FOSD-GARP that do not contain a violation of FOSD-WARP. More generally, we can (for each subject) calculate the (usual) CCEI corresponding to FOSD-rationalizability/FOSD-GARP and also the CCEI which measures the amount by which each budget constraint needs to be reduced in order to remove all violations of FOSD-WARP. The latter must (by definition) be weakly greater than the former, but the scores turn out to be identical for all subjects.

Within-Decision FOSD Violations. Finally, since all three states are equally likely, any allocation of fewer tokens to a cheaper security ($x_s > x_{s'}$ where $p_s > p_{s'}$ for any two states s and s') results in a within-decision FOSD violation (which is a special case of a FOSD-GARP/FOSD-WARP violation). All subjects have such within-decision FOSD violations but the vast majority are small due to their imprecision in handling the mouse when trying to demand nearly equal amounts in all states $x_1 = x_2 = x_3$ (suggesting infinite risk aversion) or nearly equal amounts in two out of the three states $x_s = x_{s'}$ (suggesting probability weighting).[4]

[4]The indifference curves of a rank-dependent utility function (Quiggin, 1982, 1993) have “kinks” where $x_s = x_{s'}$ so allocations that satisfy $x_s = x_{s'}$ are chosen for a non-negligible set of price vectors. Recall that

		(1)	(2)	(3)	(4)	(5)	(6)
Mean		22	0.897	0.991	0.999	0.874	0.022
Std Dev		9	0.106	0.011	0.002	0.108	0.045
Percentile Values	1	5	0.501	0.938	0.991	0.501	0.000
	5	8	0.663	0.971	0.997	0.656	0.000
	10	10	0.776	0.981	0.998	0.722	0.000
	25	15	0.865	0.987	1.000	0.830	0.000
	50	22	0.925	0.994	1.000	0.898	0.000
	75	27	0.975	0.998	1.000	0.957	0.024
	90	34	0.986	0.999	1.000	0.981	0.083
	95	37	0.994	1.000	1.000	0.989	0.117
	99	40	0.999	1.000	1.000	0.999	0.189

Table 2: FOSD Violations

The table reports summary statistics and percentile values for individual-level FOSD violations. Column (1) reports the number of within-decision FOSD violations. Columns (2), (3), and (4) report the minimum, mean, and median within-decision FOSD cost-efficiency scores, respectively. Column (5) reports the FOSD-rationalizability score e^{**}, and column (6) reports the difference between e^{**} and the minimum within-decision FOSD cost-efficiency score.

Table 2 provides a summary of the individual-level FOSD violations by reporting summary statistics and percentile values. Column (1) reports the distribution among subjects of the number of within-decision FOSD violations. For each subject, we can also compute the 50 cost-efficiency scores (the amount by which a budget constraint must be adjusted, one for each of the 50 budget constraints) required to remove all within-decision violations of FOSD; from this, we can then calculate (for each subject) the minimum, mean, and median cost-efficiency scores; the distributions are reported in columns (2), (3), and (4), respectively. For example, the mean (across subjects) of the minimum cost-efficiency score is 0.897. Column (5) reports the distribution of our measure of FOSD-rationalizability (e^{**}), and column (6) reports the difference between e^{**} and the minimum within-decision FOSD cost-efficiency score, which can be no smaller than e^{**}.

There is a large amount of heterogeneity across subjects, which is characteristic of all our data. In column (1), we see that the (rounded) average and median number of within-decision FOSD violations is 22 (out of 50). Furthermore, as reported in column (2), the average and median of the minimum within-decision FOSD cost-efficiency scores are 0.897

price vectors are randomly generated and give rise to allocations satisfying $x_s = x_{s'}$ with probability zero if preferences are smooth.

and 0.925, respectively. However, the average and median cost-efficiency scores for the within-decision FOSD violations are near 1 for the vast majority of subjects, as can be seen in columns (3) and (4). In comparison, as reported in column (5), the average and median FOSD-rationalizability (e^{**}) scores are 0.874 and 0.898, respectively. For 68 subjects (40.5 percent), there is a difference between e^{**} and the minimum within-decision FOSD cost-efficiency score. Among these subjects, there is marked heterogeneity in the magnitude of the difference, as reported in column (6).

We also pursued this question further in a complementary way. For each subject, we removed the 5 worst within-decision violations of FOSD (as measured by the within-decision FOSD cost-efficiency scores), and then recalculated the overall scores for FOSD-rationalizability (e^{**}) and EUT-rationalizability (e^{***}) on the remaining 45 observations. We note that after removing the 5 worst observations, 148 subjects (88.1 percent) had no within-decision FOSD cost-efficiency scores below 0.95, and so we can think of the remaining within-decision FOSD violations across 45 observations as minor "hand trembles" for the vast majority of subjects. Interestingly, we find that the differences between e^{**} and the minimum within-decision FOSD cost-efficiency scores (both calculated using 45 observations) are more pronounced than when using all 50 observations: there is a difference for 122 subjects (72.6 percent), and the magnitudes of these differences also tend to be larger (0.042 versus 0.022, on average). Thus we conclude that violations of FOSD-rationalizability do not simply arise from within-decision violations of FOSD.

Once the data had been "cleaned" of within-decision violations of FOSD in this way, we again performed a difference-in-differences analysis. We find that the difference between perfect rationalizability and FOSD-rationalizability ($1 - e^{**}$) is greater than the difference between FOSD-rationalizability and EUT-rationalizability ($e^{**} - e^{***}$) for 102 subjects (60.7 percent). Hence, our main result holds for the majority of subjects, even after removing within-decision violations of FOSD which may or may not have been involved in across-decision violations.

4 Power Simulations

The power simulation corresponding to Figure 3 adapts the Bronars (1987) procedure. To be precise, a dataset is generated in the following manner. To construct observation 1, we first generate a budget set as in the experiment. We then select a bundle randomly (uniformly) from this budget and check e^{**}. If $e^{**} = 1$, we proceed to observation 2; otherwise, we select another bundle at observation 1. To construct observation 2, we again generate a budget set and select a bundle randomly (as in observation 1); we then check whether $e^{**} = 1$ for observations 1 and 2 taken together. If $e^{**} = 1$, we proceed to observation 3; otherwise, we select another bundle at observation 2. We repeat this process until we obtain a dataset with 50 observations and $e^{**} = 1$. Overall, we simulate 1,000 such datasets.

The power simulation corresponding to Figure 9 also adapts the Bronars (1987) procedure. To be precise, a dataset is generated in the following manner. Consider the interval $(0.91, 0.92]$. To construct observation 1, we first generate a budget set as in the experiment. We then select a bundle randomly (uniformly) from this budget and check $e^{\dagger\dagger}$. If $e^{\dagger\dagger} > 0.91$, we proceed to observation 2; otherwise, we select another bundle at observation 1. To construct observation 2, we again generate a budget set and select a bundle randomly (as in observation 1); we then check whether $e^{\dagger\dagger} > 0.91$ for observations 1 and 2 taken together. If $e^{\dagger\dagger} > 0.91$, we proceed to observation 3; otherwise, we select another bundle at observation 2. We repeat this process until we obtain a dataset with 25 observations and $e^{\dagger\dagger} > 0.91$. Finally, we check whether $e^{\dagger\dagger} \leqslant 0.92$; the simulated dataset is retained if that is the case and it is discarded otherwise. Overall, we simulate 1,000 such datasets on each interval. In order to generate the simulated bars in Figure 9, we aggregate across intervals.

References

Bronars, S. G. 1987. "The Power of Nonparametric Tests of Preference Maximization." *Econometrica* 55(3): 693–698.

Cappelen, A. W., S. Kariv, E. Ø. Sørensen, and B. Tungodden. 2023. "The Development Gap in Economic Rationality of Future Elites." *Games and Economic Behavior* 142: 866–878.

Choi, S., R. Fisman, D. Gale, and S. Kariv. 2007. "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review* 97(5): 1921–1938.

Hu, G., J. Li, J. K.-H. Quah, and R. Tang. 2021. "A Theory of Revealed Indirect Preference." Unpublished paper.

Varian, H. R. 1985. "Non-Parametric Analysis of Optimizing Behavior with Measurement Error." *Journal of Econometrics* 30(1-2): 445–458.

Zame, W. R., B. Tungodden, E. Ø. Sørensen, S. Kariv, and A. W. Cappelen. 2020. "Linking Social and Personal Preferences: Theory and Experiment." Unpublished paper.

Appendix III : Experimental Instructions

Introduction. This is an experiment in decision-making. Research foundations have provided funds for conducting this research. Your payoffs will depend partly only on your decisions and partly on chance. It will not depend on the decisions of the other participants in the experiments. Please pay careful attention to the instructions as a considerable amount of money is at stake.

The entire experiment should be complete within an hour and a half. At the end of the experiment you will be paid privately. At this time, you will receive $5 as a participation fee (simply for showing up on time). Details of how you will make decisions and receive payments will be provided below.

During the experiment we will speak in terms of experimental tokens instead of dollars. Your payoffs will be calculated in terms of tokens and then translated at the end of the experiment into dollars at the following rate:

2 Tokens = 1 Dollar

A Decision Problem. In this experiment, you will participate in 50 independent decision problems that share a common form. This section describes in detail the process that will be repeated in all decision problems and the computer program that you will use to make your decisions.

In each decision problem you will be asked to allocate tokens between three accounts, labeled x, y and z. Each choice will involve choosing a point on a three-dimensional graph representing possible token allocations, x / y / z. The x account corresponds to the x-axis, the y account corresponds to the y-axis and the z account corresponds to the z-axis in a three-dimensional graph. In each choice, you may choose any combination of x / y / z that is on the plane that is shaded in gray. Examples of planes that you might face appear in Figure 1.

Each decision problem will start by having the computer select such a plane randomly from the set of planes that intersect with at least one of the axes (x, y or z) at 50 tokens or more but with no intercept exceeding 100 tokens. The planes selected for you in different decision problems are independent of each other and independent of the planes selected for any of the other participants in their decision problems.

For example, as illustrated in Figure 2, choice A represents an allocation in which you allocate approximately 20 tokens in the x account, 21 tokens in the y account, and 30 tokens in the z account. Another possible allocation is B, in which you allocate approximately 40 tokens in the x account, 17 tokens in the y account, and 11 tokens in the z account.

To choose an allocation, use the mouse to move the pointer on the computer screen to the allocation that you desire. On the right hand side of the program dialog window, you will be informed of the exact allocation that the pointer is located. When you are ready to make your decision, left-click to enter your chosen allocation. After that, confirm your decision by clicking on the Submit button. Note that you can choose only x / y / z combinations that are on the gray plane. To move on to the next round, press the OK button. The computer program dialog window is shown in Figure 3.

Your payoff at each decision round is determined by the number of tokens in each account. At the end of the round, the computer will randomly select one of the accounts, x, y or z. For each participant, account x will be selected with 1/3 chance, account y will be selected with 1/3 chance and account z will be selected with 1/3 chance. You will only receive the number of tokens you allocated to the account that was chosen.

Next, you will be asked to make an allocation in another independent decision. This process will be repeated until all 50 rounds are completed. At the end of the last round, you will be informed the experiment has ended.

Earnings. Your earnings in the experiment are determined as follows. At the end of the experiment, the computer will randomly select one decision round from each participant to carry out (that is, 1 out of 50). The round selected depends solely upon chance. For each participant, it is equally likely that any round will be chosen.

The round selected, your choice and your payment will be shown in the large window that appears at the center of the program dialog window. At the end of the experiment, the tokens will be converted into money. Each token will be worth 0.50 Dollars. Your final earnings in the experiment will be your earnings in the round selected plus the $5 show-up fee. You will receive your payment as you leave the experiment.

Rules. Your participation in the experiment and any information about your payoffs will be kept strictly confidential. Your payment-receipt and participant form are the only places in which your name and social security number are recorded.

You will never be asked to reveal your identity to anyone during the course of the experiment. Neither the experimenters nor the other participants will be able to link you to any of your decisions. In order to keep your decisions private, please do not reveal your choices to any other participant.

Please do not talk with anyone during the experiment. We ask everyone to remain silent until the end of the last round. If there are no further questions, you are ready to start. An instructor will approach your desk and activate your program.

Figure 1

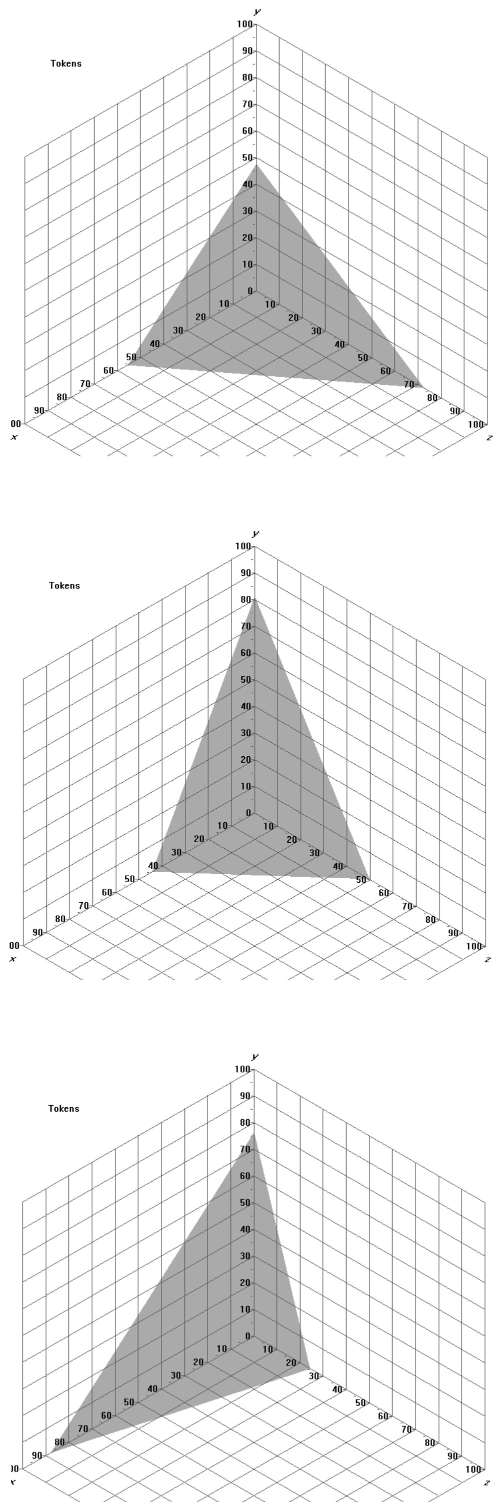
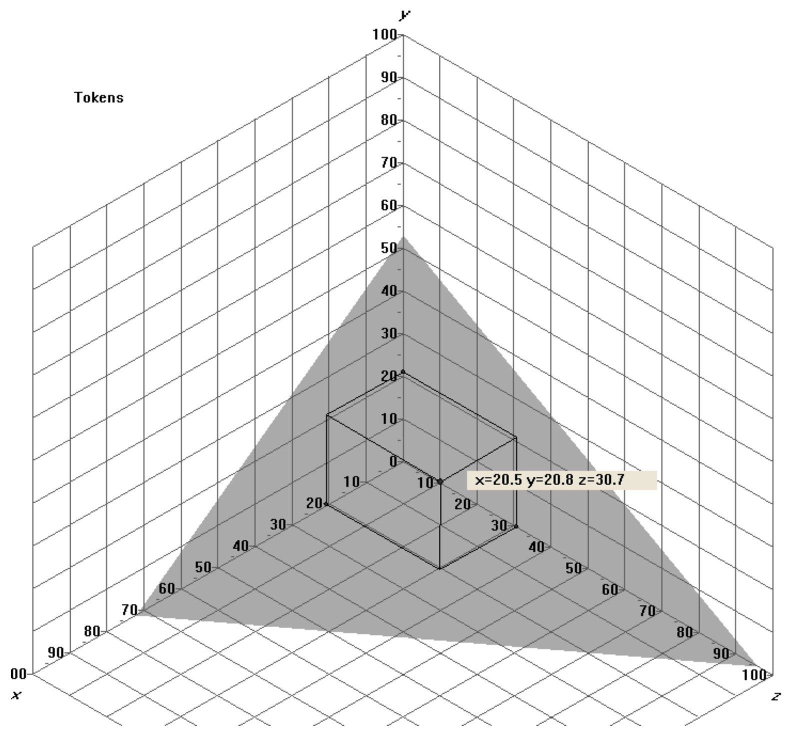


Figure 2

Choice A



Choice B

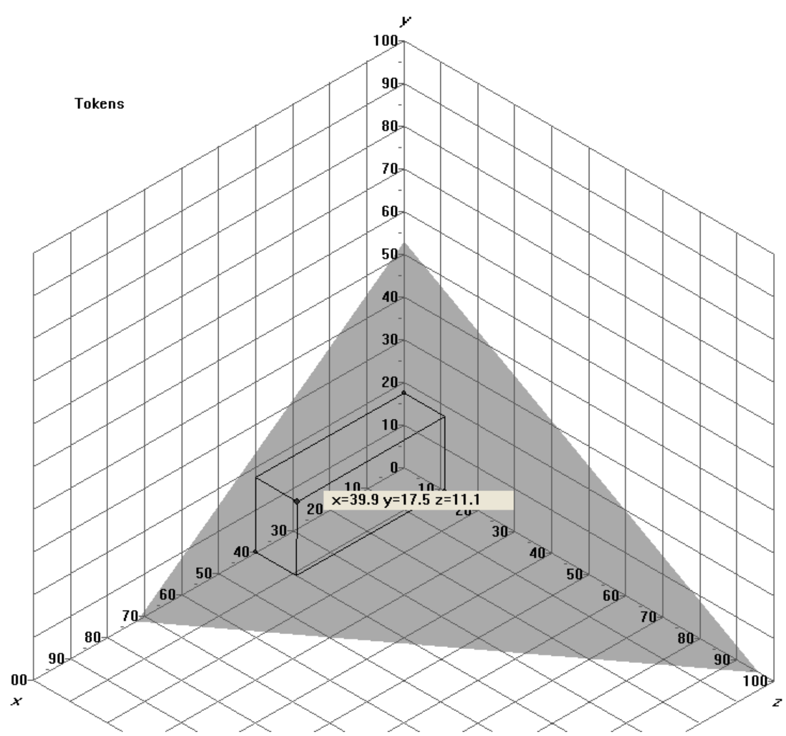
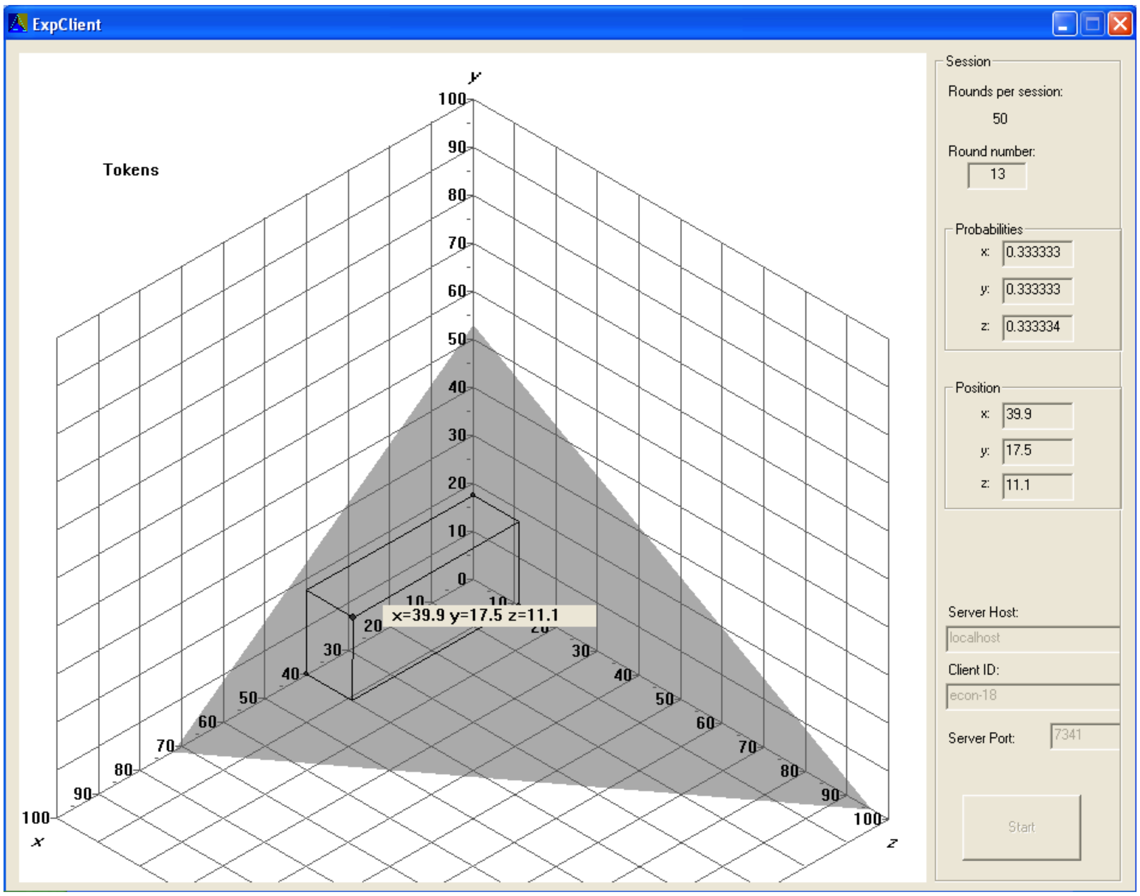


Figure 3


ExpClient
Tokens
x=39.9 y=17.5 z=11.1
Session
Rounds per session:
50
Round number:
13
Probabilities
x: 0.333333
y: 0.333333
z: 0.333334
Position
x: 39.9
y: 17.5
z: 11.1
Server Host:
localhost
Client ID:
econ-18
Server Port:
7341
Start