

Appendix I Theoretical Framework

Our experimental data are generated by individual subjects solving a series of randomly generated portfolio-choice problems. In our setting, there are S states of nature, denoted by $s = 1, \dots, S$. The probability of state s is commonly known to be $\pi_s > 0$, with $\sum_{s=1}^S \pi_s = 1$, so that $\boldsymbol{\pi} = (\pi_1, \dots, \pi_S) \gg \mathbf{0}$ denotes the vector of state probabilities.¹ For each state s , there is an Arrow security that pays one token (the experimental currency) in state s and nothing in the other state(s). The amount of consumption in state s is denoted by $x_s \geq 0$, and the portfolio of securities may be written as $\mathbf{x} = (x_1, \dots, x_S) \geq \mathbf{0}$.

In the experiment, each subject has a budget of 1, which has to be allocated among the Arrow securities, with $p_s > 0$ denoting the price of security s . Formally the subject chooses a portfolio $\mathbf{x} \geq \mathbf{0}$ among those which satisfy the constraint $\mathbf{p} \cdot \mathbf{x} = 1$, where $\mathbf{p} = (p_1, \dots, p_S) \gg \mathbf{0}$ denotes the vector of state prices. The subject can choose any portfolio \mathbf{x} satisfying the budget constraint.

Let $\mathcal{D} := (\mathbf{p}^i, \mathbf{x}^i)$ be the dataset generated by an individual subject's choices from these 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 by the subject. The subject's total expenditure is fixed at 1 throughout, so $\mathbf{p}^i \cdot \mathbf{x}^i = 1$ for all observations i . The experimental design required subjects to solve a sequence of 50 decision problems (so \mathcal{D} has 50 observations) involving three-dimensional budget sets ($S = 3$), and we also compare these results against the results from otherwise identical experiments involving two-dimensional budget lines ($S = 2$). In all of the two- and three-dimensional experiments that we consider, the states are equiprobable, though the theoretical results which we review below do not hinge on this feature.

Rationalizability (e^*) Recall, from the main paper, that we refer to a utility function $U : \mathbb{R}_+^S \rightarrow \mathbb{R}$ as *well-behaved* if it is continuous and increasing, where the latter means that $U(\mathbf{x}'') > U(\mathbf{x}')$ if $\mathbf{x}'' > \mathbf{x}'$. A utility function U rationalizes \mathcal{D} if $U(\mathbf{x}^i) \geq U(\mathbf{x})$ for all

$$\mathbf{x} \in \mathcal{B}^i = \{\mathbf{x} \in \mathbb{R}_+^S : \mathbf{p}^i \cdot \mathbf{x} \leq \mathbf{p}^i \cdot \mathbf{x}^i\}.$$

In other words, the utility of \mathbf{x}^i is weakly higher than that of any alternative that is weakly cheaper at the price vector \mathbf{p}^i . When a dataset \mathcal{D} can be rationalized by a well-behaved utility function U , we say that \mathcal{D} is *rationalizable by a well-behaved utility function*, or simply *rationalizable*. Afriat's (1967) Theorem characterizes rationalizable datasets via the Generalized Axiom of Revealed Preference (GARP).

Let $\mathcal{X} = \{\mathbf{x}^i\}$ be the set of portfolios observed across all observations i . For any $\mathbf{x}^i, \mathbf{x}^j \in \mathcal{X}$, we say that \mathbf{x}^i is *directly revealed preferred* to \mathbf{x}^j (and denote this relation by $\mathbf{x}^i R^D \mathbf{x}^j$) if $\mathbf{p}^i \cdot \mathbf{x}^i \geq \mathbf{p}^i \cdot \mathbf{x}^j$. GARP requires that if \mathbf{x}^i is revealed

¹As a matter of notation, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^S$, we say that $\mathbf{x} \geq \mathbf{y}$ if $x_s \geq y_s$ for all s ; $\mathbf{x} > \mathbf{y}$ if $\mathbf{x} \geq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$; and $\mathbf{x} \gg \mathbf{y}$ if $x_s > y_s$ for all s .

preferred to \mathbf{x}^j (either directly or indirectly via a sequence of other portfolio choices), then \mathbf{x}^i must cost at least as much as \mathbf{x}^j at the prices prevailing when \mathbf{x}^j is chosen. To be precise, we define on \mathcal{X} the *revealed preference* relation, where \mathbf{x}^i is revealed preferred to \mathbf{x}^j (denoted by $\mathbf{x}^i R \mathbf{x}^j$) if there is a sequence of observations i_1, i_2, \dots, i_n such that

$$\mathbf{x}^i R^D \mathbf{x}^{i_1} R^D \mathbf{x}^{i_2} R^D \dots R^D \mathbf{x}^{i_n} R^D \mathbf{x}^j.$$

In other words, the relation R is the transitive closure of the relation R^D . We also define the *strict direct revealed preference* relation P^D , where $\mathbf{x}^i P^D \mathbf{x}^j$ if $\mathbf{p}^i \cdot \mathbf{x}^i > \mathbf{p}^i \cdot \mathbf{x}^j$. GARP requires that, for any $\mathbf{x}^i, \mathbf{x}^j \in \mathcal{X}$,

$$\text{if } \mathbf{x}^i R \mathbf{x}^j, \text{ then } \mathbf{x}^j P^D \mathbf{x}^i \text{ does not hold.}$$

The term “revealed preference” for the relation R is very intuitive, since if a dataset can be rationalized by some utility function U , then $U(\mathbf{x}^i) \geq U(\mathbf{x}^j)$ if $\mathbf{x}^i R \mathbf{x}^j$. Furthermore, it is not hard to show that if U is locally nonsatiated, then $U(\mathbf{x}^i) > U(\mathbf{x}^j)$ if $\mathbf{x}^i P^D \mathbf{x}^j$. It follows from these observations that if \mathcal{D} is rationalizable by a locally nonsatiated utility function then it must obey GARP, since it impossible for $U(\mathbf{x}^i) \geq U(\mathbf{x}^j)$ and for $U(\mathbf{x}^j) > U(\mathbf{x}^i)$ to hold simultaneously.² The substantive part of Afriat’s Theorem says that if \mathcal{D} obeys GARP then it is rationalizable by a concave and well-behaved utility function. Notice that the two statements are not completely symmetric: GARP holds whenever a dataset is generated by a locally nonsatiated utility function, but whenever GARP holds on a dataset, it can also be rationalized by a utility function with properties that are stronger than local nonsatiation.

Figure 1 illustrates a simple violation of GARP involving two budget sets $\mathbf{p}^1 = (\frac{3}{9}, \frac{2}{9}, \frac{1}{9})$ and $\mathbf{p}^2 = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, and two portfolio allocations $\mathbf{x}^1 = (1, 2, 2)$ and $\mathbf{x}^2 = (0, 1, 5)$. It is clear that $\mathbf{x}^1 P^D \mathbf{x}^2$ and $\mathbf{x}^2 P^D \mathbf{x}^1$ since $\mathbf{p}^1 \cdot \mathbf{x}^1 > \mathbf{p}^1 \cdot \mathbf{x}^2$ and $\mathbf{p}^2 \cdot \mathbf{x}^2 > \mathbf{p}^2 \cdot \mathbf{x}^1$.

GARP provides an exact test of utility maximization (either the data satisfy GARP or they do not). To account for the possibility of errors, we assess how close a dataset is to being rationalizable by using Afriat’s (1972, 1973) Critical Cost Efficiency Index (CCEI), which we shall now explain.

Given a number $e \in (0, 1]$, a dataset \mathcal{D} is *rationalizable at cost efficiency* e if there is a well-behaved utility function U such that $U(\mathbf{x}^i) \geq U(\mathbf{x})$ for all

$$\mathbf{x} \in \mathcal{B}^i(e) = \{\mathbf{x} \in \mathbb{R}_+^S : \mathbf{p}^i \cdot \mathbf{x} \leq e \mathbf{p}^i \cdot \mathbf{x}^i\}.$$

²A utility function $U : \mathbb{R}_+^S \rightarrow \mathbb{R}$ is locally nonsatiated if, in any open ball around $\mathbf{x} \in \mathbb{R}_+^S$, there is some \mathbf{x}' such that $U(\mathbf{x}') > U(\mathbf{x})$. The eagle-eyed reader may notice that in our experiments each subject at observation i chooses from the budget boundary $\bar{\mathcal{B}}^i = \{\mathbf{x} \in \mathbb{R}_+^S : \mathbf{p}^i \cdot \mathbf{x} = 1\}$ rather than from the budget set \mathcal{B}^i , so that we ought to check that \mathcal{D} satisfies GARP if \mathbf{x}^i is a utility-maximizing choice from $\bar{\mathcal{B}}^i$. This is indeed the case provided that U is continuous and locally nonsatiated; these assumptions on U guarantee that $\arg \max_{\mathbf{x} \in \bar{\mathcal{B}}^i} U(\mathbf{x}) = \arg \max_{\mathbf{x} \in \mathcal{B}^i} U(\mathbf{x})$ so that $U(\mathbf{x}^i) \geq U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{B}^i$ and $U(\mathbf{x}^i) > U(\mathbf{x})$ if $\mathbf{x} \in \mathcal{B}^i \setminus \bar{\mathcal{B}}^i$. In particular, this implies that $U(\mathbf{x}^i) \geq U(\mathbf{x}^j)$ if $\mathbf{x}^i R \mathbf{x}^j$ and $U(\mathbf{x}^i) > U(\mathbf{x}^j)$ if $\mathbf{x}^i P^D \mathbf{x}^j$.

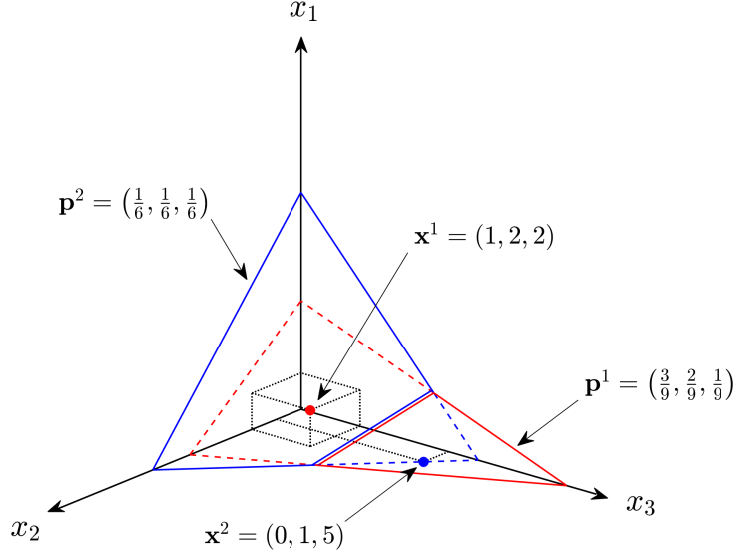


Figure 1: Violation of Rationalizability

It is not difficult to see that *every* dataset \mathcal{D} could be rationalized by a well-behaved utility function at an efficiency level e for some $e \in (0, 1]$ that is sufficiently close to zero. Afriat's CCEI, denoted by e^* , is the largest value of e associated with the dataset \mathcal{D} ; formally,

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

A subject with a CCEI of $e^* < 1$ makes mistakes, in the sense that there is at least one observation k for which $U(\mathbf{x}^k) < U(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{B}^k$, but the cost inefficiency is bounded in the sense that $\mathbf{p} \cdot \mathbf{x} \geq e^*$; thus the subject could switch to a bundle \mathbf{x} that gives the same utility as \mathbf{x}^k and spend less, but the savings is no more than $1 - e^*$.

The coefficient e^* can be straightforwardly obtained through a binary search, once there is a way to check if a dataset is rationalizable at cost efficiency e for any given value of e . Very conveniently, rationalizability at cost efficiency e can be characterized by a generalized version of GARP. We define the direct revealed preference relation at efficiency e (denoted by $R^D(e)$) as follows: $\mathbf{x}^i R^D(e) \mathbf{x}^j$ if $e \mathbf{p}^i \cdot \mathbf{x}^i \geq \mathbf{p}^i \cdot \mathbf{x}^j$. The revealed preference relation $R(e)$ is the transitive closure of $R^D(e)$. Similarly, the strict direct revealed preference relation at efficiency e (denoted by $P^D(e)$) is defined as follows: $\mathbf{x}^i P^D(e) \mathbf{x}^j$ if $e \mathbf{p}^i \cdot \mathbf{x}^i > \mathbf{p}^i \cdot \mathbf{x}^j$. e -GARP requires that, for any $\mathbf{x}^i, \mathbf{x}^j \in \mathcal{X}$,

$$\text{if } \mathbf{x}^i R(e) \mathbf{x}^j, \text{ then } \mathbf{x}^j P^D(e) \mathbf{x}^i \text{ does not hold.}$$

It is straightforward to check that if a dataset \mathcal{D} can be rationalized at cost

efficiency e by a locally nonsatiated utility function, then it will satisfy e -GARP; conversely, if \mathcal{D} satisfies e -GARP, then it is rationalizable at efficiency e by a concave and well-behaved utility function (see Afriat (1973)).

FOSD-Rationalizability (e^{})** Nishimura et al. (2017) shows that a further modification of GARP can be used to test whether a dataset \mathcal{D} is rationalizable (at cost efficiency e) by a continuous utility function that is increasing with respect to a given preorder \succeq on the choice space. This result is convenient for our purposes because for a utility function U to be monotone with respect to FOSD simply means that it is increasing with respect to the preorder \succeq , where $\mathbf{x}'' \succeq \mathbf{x}'$ if \mathbf{x}' and \mathbf{x}'' (when considered as distributions given the vector of state probabilities $\boldsymbol{\pi}$) have the property that \mathbf{x}'' first-order stochastically dominates \mathbf{x}' . In our experiments, each state is equally likely; thus, $\mathbf{x}'' \succeq \mathbf{x}'$ if there is some permutation of the entries in \mathbf{x}'' such that the permuted allocation is entry-by-entry weakly greater than \mathbf{x}' . For example, $(1, 0, 1) \succeq (0, 1, 0)$ since $(1, 1, 0) \geq (0, 1, 0)$. In this case, a well-behaved utility function is monotone with respect to FOSD if and only if it is symmetric.

We say that a dataset \mathcal{D} is *FOSD-rationalizable at cost efficiency e* if it can be rationalized at cost efficiency e by a well-behaved utility function that is monotone with respect to FOSD. The rationalizability score e^{**} is given by

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

The FOSD-rationalizability at cost efficiency e of a dataset \mathcal{D} can be characterized by a generalized notion of GARP which we shall now explain.

We define the direct revealed preference relation at efficiency e (denoted by $R_{\succeq}^D(e)$) as follows: $\mathbf{x}^i R_{\succeq}^D(e) \mathbf{x}^j$ if there exists some \mathbf{y} such that $e \mathbf{p}^i \cdot \mathbf{x}^i \geq \mathbf{p}^i \cdot \mathbf{y}$ and $\mathbf{y} \succeq \mathbf{x}^j$. The revealed preference relation $R_{\succeq}(e)$ is the transitive closure of $R_{\succeq}^D(e)$. Similarly, the strict direct revealed preference relation at efficiency e (denoted by $P_{\succeq}^D(e)$) is defined as follows: $\mathbf{x}^i P_{\succeq}^D(e) \mathbf{x}^j$ if there exists some \mathbf{y} such that $e \mathbf{p}^i \cdot \mathbf{x}^i \geq \mathbf{p}^i \cdot \mathbf{y}$ and $\mathbf{y} \succeq \mathbf{x}^j$ but $\mathbf{x}^j \not\succeq \mathbf{y}$ (in other words, \mathbf{y} strictly first-order stochastically dominates \mathbf{x}^j). e -GARP(\succeq) requires that, for any $\mathbf{x}^i, \mathbf{x}^j \in \mathcal{X}$,

$$\text{if } \mathbf{x}^i R_{\succeq}(e) \mathbf{x}^j, \text{ then } \mathbf{x}^j P_{\succeq}^D(e) \mathbf{x}^i \text{ does not hold.}$$

A dataset \mathcal{D} satisfies e -GARP(\succeq) if and only if it is FOSD-rationalizable at cost efficiency e .

To illustrate in simple terms how the test works, Figure 2 depicts the same two budget sets as in Figure 1, $\mathbf{p}^1 = (\frac{3}{9}, \frac{2}{9}, \frac{1}{9})$ and $\mathbf{p}^2 = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, with the portfolio allocations $\mathbf{x}^1 = (1, 2, 2)$ and $\mathbf{x}^2 = (0, 5, 1)$. These choices are rationalizable but not FOSD-rationalizable (with equiprobable states) because $\mathbf{x}^1 P_{\succeq}^D(1) \mathbf{x}^2$ and $\mathbf{x}^2 P_{\succeq}^D(1) \mathbf{x}^1$. It is clear that $\mathbf{p}^2 \cdot \mathbf{x}^2 > \mathbf{p}^2 \cdot \mathbf{x}^1$, but it is also the case that $\mathbf{p}^1 \cdot \mathbf{x}^1 > \mathbf{p}^1 \cdot \mathbf{y}$ where $\mathbf{y} = (0, 1, 5) \succeq (0, 5, 1) = \mathbf{x}^2$.

Violations of FOSD are typically regarded as errors, regardless of risk attitude — that is, as a failure to recognize that some allocations yield payoff distributions with unambiguously lower returns. As a result, the most prominent non-EUT models have been constructed/amended to avoid violations of

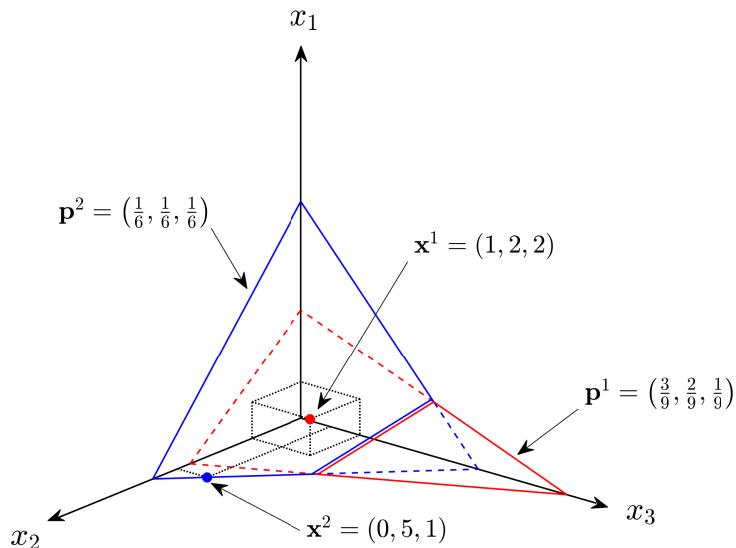


Figure 2: Violation of FOSD-Rationalizability

FOSD. There are, however, some notable exceptions. For example, Kőszegi and Rabin's (2007) reference-dependent risk preferences may violate FOSD due to (excessive) loss aversion — see Masatlioglu and Raymond's (2016) characterization. However, the Kőszegi and Rabin (2007) utility function U is locally nonsatiated and, in the case where states are equiprobable (as in our experiments), it must respect symmetry. It is straightforward to check that a subject who maximizes a symmetric and locally nonsatiated utility function at cost efficiency e would generate a dataset \mathcal{D} satisfying e -GARP(\succeq) (with \succeq being the preorder corresponding to equiprobable states) and thus \mathcal{D} is FOSD-rationalizable at cost efficiency e . In other words, in the context of our experiments, reference-dependent risk preferences cannot do better in explaining a subject's data than the family of utility functions that are monotone with respect to FOSD.

EUT-Rationalizability (e^{*})** Polisson et al. (2020) develops a revealed preference method to test whether choice data under risk are consistent with maximizing a utility function that has some special structure. The method restricts an infinite choice set to a finite grid, and is thus called the method of Generalized Restriction of Infinite Domains (GRID). GRID tests are mechanically distinct from GARP tests (in the sense that they do not involve constructing revealed preference relations and checking for strict cycles), but they are fully nonparametric (within the specified class of utility functions) and can also be used to measure inconsistencies. This is the approach that we use to test expected utility.

We say that a dataset \mathcal{D} is *EUT-rationalizable at cost efficiency e* if it can be rationalized at cost efficiency e by a well-behaved utility function U taking the expected utility form, i.e., if there is a continuous and increasing Bernoulli index $u : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that $U(\mathbf{x}) = \sum_{s=1}^S \pi_s u(x_s)$. Following Polisson et al. (2020), let \mathcal{Y} be the set that contains any demand level observed in a given dataset \mathcal{D} plus zero, that is

$$\mathcal{Y} := \{x \in \mathbb{R}_+ : x = x_s^i \text{ for some } (i, s)\} \cup \{0\}.$$

We then form the finite grid $\mathcal{G} = \mathcal{Y}^S \subset \mathbb{R}_+^S$ which is a restriction of the choice space \mathbb{R}_+^S to allocations comprised of demand levels that have been observed in the dataset \mathcal{D} . We claim that EUT-rationalizability at cost efficiency e requires the existence of a real number $\bar{u}(y)$ associated with each $y \in \mathcal{Y}$, with $\bar{u}(y') > \bar{u}(y)$ whenever $y' > y$, such that at each observation of $(\mathbf{p}^i, \mathbf{x}^i)$

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

and

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

Indeed, if a dataset \mathcal{D} can be EUT-rationalized at cost efficiency e by a 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 important application of the main result of Polisson et al. (2020) is that these conditions are also sufficient for EUT-rationalizability 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 way of determining whether a dataset \mathcal{D} is EUT-rationalizable at cost efficiency e and thus allows us to calculate its rationalizability score

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

To illustrate, Figure 3 depicts the same two budget sets as in Figures 1 and 2, $\mathbf{p}^1 = (\frac{3}{9}, \frac{2}{9}, \frac{1}{9})$ and $\mathbf{p}^2 = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, with the 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

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

But EUT-rationalizability requires that

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

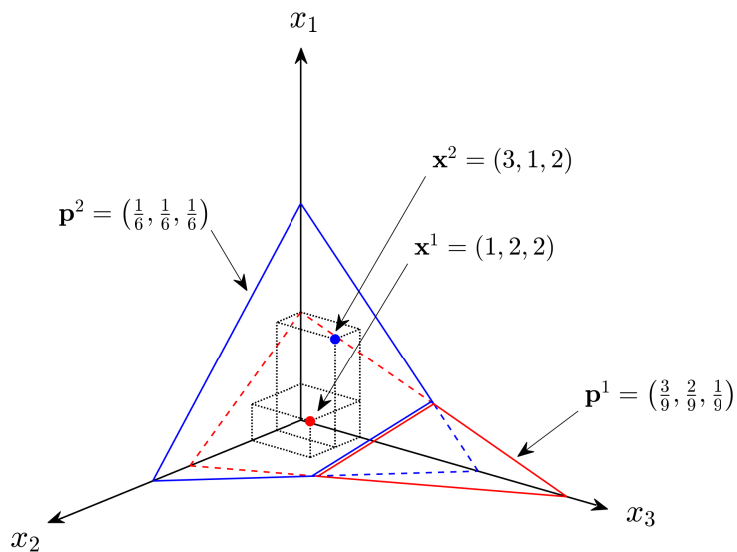


Figure 3: Violation of EUT-Rationalizability

and

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

implying that $2u(2) > u(1) + u(3)$ and $u(3) + u(1) \geq 2u(2)$, a contradiction. The GRID procedure would also reveal this violation of EUT-rationalizability. To see this, note there must exist real numbers $\bar{u}(1) < \bar{u}(2) < \bar{u}(3)$ satisfying

$$2\bar{u}(2) > \bar{u}(1) + \bar{u}(3) \text{ and } \bar{u}(3) + \bar{u}(1) \geq 2\bar{u}(2),$$

which is an impossibility.