Linking Social and Personal Preferences: Theory and Experiment*

William R. Zame, Bertil Tungodden, Erik Ø. Sørensen,
Shachar Kariv, and Alexander W. Cappelen†

August 13, 2024

Abstract

We provide necessary and sufficient conditions for linking preferences for personal and social consumption and attitudes toward risk. We also offer an experimental test of the theory in which subjects were confronted with risky personal choices, riskless social choices and risky social choices. Revealed preference tests show that subject choices are generally consistent *within* each choice domain but frequently involve at least some errors. We test for consistency *across* choice domains using a revealed preference test that accounts for these errors. The choices of a large majority of subjects are consistent with the predictions of our theory.
JEL Classification Numbers: C91, D63, D81.
Keywords: social preferences, risk preferences, revealed preference, experiments.

*We are grateful to Chris Chambers, Raymond Fisman, Douglas Gale, Edi Karni, Kristian López-Vargas, Daniel Markovits, Benjamin Polak, Matthew Polisson, Robert Powell, and Daniel Silverman for encouragement and helpful discussions, and to a number of seminar audiences for helpful suggestions. The experiments reported in this paper were conducted by the Choice Lab at the Centre for Experimental Research on Fairness, Inequality and Rationality (FAIR) at NHH Norwegian School of Economics. Funding for the experiments was provided by the Research Council of Norway through its Centres of Excellence Scheme, FAIR Project No. 262675 and Research Grant No. 236995. Financial support was also provided by the National Science Foundation, the Peder Sather Center for Advanced Study, the Geneva Finance Research Institute, and the Einaudi Institute for Economics and Finance. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of any funding agency.

†Authors in reverse alphabetical order. Zame: University of California, Los Angeles (zame@econ.ucla.edu); Tungodden: Norwegian School of Economics (bertil.tungodden@nhh.no); Sørensen: Norwegian School of Economics (erik.sorensen@nhh.no); Kariv: University of California, Berkeley (kariv@berkeley.edu); Cappelen: Norwegian School of Economics (alexander.cappelen@nhh.no).

1 Introduction

Many individuals make choices that have consequences only for themselves—choices in the *personal domain*—and also make choices that have consequences both for themselves and for others—choices in the *social domain*. Many of these choices involve risk, so a full understanding of choice behavior in these domains requires a commensurate understanding of both the individual's preferences over consequences and the individual's attitude toward risk. It is then natural to ask: Is there a connection between an individual's attitude toward risk in the personal domain and the same individual's attitude toward risk in the social domain? This paper offers formalizations of this question, theoretical responses to this question, and an experimental test of the theory.

Our motivation for asking (and answering) this question arises not only from intellectual curiosity but also from pragmatism because we often choose—or at least influence the choice of—those individuals who will be in a position to make choices that have consequences for us (and for themselves): Chairs, Deans, Mayors, Governors, Congresspersons, Senators—even Presidents. And we certainly care not only whether the President prefers peace to war and seeks tax and civil rights reforms, but what actions the President would be prepared to take to alter the risks of peace or war and the likelihood of achieving those reforms. Blockading Soviet ships bound for Cuba (as John Kennedy did) risks war, and putting forward both tax reform and civil rights legislation simultaneously (as Kennedy also did) risks accomplishing neither.[1]

But can we draw any inferences at all about the President's risky choices in the social domain from the fact that the President chooses to conduct an illicit affair or smoke in secret or invest aggressively or exaggerate his accomplishments (to mention just a few personal choices that have made headlines in recent memory)? What these personal choices have in common is that they involve (personal) risk—to the President's marriage or health or finances or reputation. Drawing inferences about (present and future) risky social choices from knowledge of (past and present) risky personal choices would seem useful, and also possible—provided that there is a *linkage* between attitude toward risk in the personal domain and attitude toward risk in the social domain.[2]

In this paper, we establish a theoretical linkage between preferences and risk attitudes in the social domain and in the personal domain, and provide an experimental test of this

[1] In fact, Kennedy accomplished neither tax reform nor civil rights legislation; both were pushed through by Lyndon Johnson after Kennedy's assassination.

[2] That voters do learn *something* from candidates' risky choices in the personal domain—and that they make use of what they learn—is perhaps most clearly illustrated by the case of Gary Hart. Hart was a leading candidate for the 1988 Democratic presidential nomination. In an attempt to dispel rumors that he was having an extra-marital affair, Hart challenged the press to "follow me around"; the press took up the challenge and promptly found him on his boat with a woman not his wife. Nelson (2018), among others, argued that the issue was less what Hart's behavior revealed about his morality and more what it revealed about his attitude toward risk: "Hart's extramarital escapades ... were politically harmful less because of his moral weakness than because of the recklessness the incidents illuminated in his character."

theory.[3] We formulate our problem abstractly by assuming that the individual Decision Maker ($\mathcal{DM}$) has preferences over risky choices in the social domain but that we can observe only preferences over risky choices in the personal domain and preferences over non-risky choices in the social domain. We then ask under what assumptions it will be possible to *deduce*—on the basis of these observations—preferences over risky choices in the social domain. Our theoretical results provide necessary and sufficient conditions that such deduction be possible. The required conditions depend on what is observed and what we assume about the $\mathcal{DM}$'s degree of rationality.

The formal model considers a $\mathcal{DM}$, characterized by a fixed preference relation $\succeq$ over the set $L(\Omega)$ of lotteries on a set Ω of *social states*.[4] A subset $P \subset \Omega$ consists of states that have consequences only for the $\mathcal{DM}$—these are *personal states*—while the others have consequences both for the $\mathcal{DM}$ and for others. It is convenient to view the personal states as a subset of the social states in which the consequences for others are fixed. We do not observe the entire preference relation $\succeq$ on $L(\Omega)$ but only some portion of it. In our main theoretical result, and in our experimental work, we assume that we can observe the restrictions $\succeq_0$ of $\succeq$ to Ω and to $L(P)$. That is, we observe comparisons between social states (including personal states) and comparisons between personal lotteries, but we *do not* observe comparisons between social states and personal lotteries.[5]

We ask: in what circumstances it is possible to *deduce* the entire preference relation $\succeq$ from the restriction $\succeq_0$? In other words, in what circumstances does $\succeq_0$ admit a *unique* extension to the preference relation $\succeq$ over the full domain of lotteries $L(\Omega)$ on social states? If we assume that the $\mathcal{DM}$'s preferences obey the usual axioms of individual choice under uncertainty— Completeness, Transitivity, and Continuity—together with a (relatively weak) axiom that we call *State Monotonicity*, then a necessary and sufficient condition that it be possible to deduce the preference relation $\succeq$ from the *partial* relation $\succeq_0$ is that the $\mathcal{DM}$ finds every social state to be indifferent to some personal state.[6]

[3]One important qualification needs to be remembered when interpreting our results. As the examples above illustrate, there is a question that seems both puzzling and important: what is the meaning of an individual's "attitude toward risk" in a domain in which the consequences are not monetary—or more generally, involve consequences other than consumption? We confess that we have no answer to offer to this question; indeed we suspect it has no entirely satisfactory answer. Keeping this in mind, we formulate an abstract model that avoids this issue entirely (we explain below) but we construct our experiment so that the consequences are *monetary* (for the subject and for one other).

[4]To make the analysis simpler and sharper, we assume that Ω is finite. This avoids subtle issues about the topology of Ω and the continuity of $\succeq$.

[5]For completeness, we do provide theoretical analysis of the setting in which we observe comparisons between social states and personal lotteries—that is, we observe the restriction $\succeq_1$ of $\succeq$ to $\Omega \cup L(P)$—but we do not have any experimental evidence in that setting.

[6]As we discuss in Section 3, State Monotonicity is a *much* weaker assumption than Independence because it compares only lotteries whose outcomes are primitives—social states—rather than lotteries whose outcomes are themselves lotteries. We show that, in the presence of the other Axioms, State Monotonicity is equivalent to respect for First Order Stochastic Dominance. Expected Utility, and almost all decision-theoretic models that have been proposed as alternatives to Expected Utility, have this property.

This theoretical result seems clean and satisfying but it is another question entirely whether is also descriptive of reality. To address this latter question, we design and execute an experiment in which we present subjects with a sequence of choices in three domains; each choice has consequences for *self* (the subject) and for an *other* (an anonymous other subject):

- Personal Risk The objects of choice are risky personal choices (equiprobable binary lotteries whose consequences are monetary outcomes for *self* alone).
- Social Choice The objects of choice are riskless social choices (deterministic divisions of money between *self* and one *other*).
- Social Risk The objects of choice are risky social choices (equiprobable binary lotteries whose consequences are divisions of money between *self* and *other*).

In the experimental setting, we present each decision problem as a choice from a budget line using a graphical interface developed by Choi et al. (2007b). The Personal Risk domain is identical to the domain in the (symmetric) risk experiment of Choi et al. (2007a). The Social Choice domain is identical to the domain in the (linear) two-person dictator experiment of Fisman et al. (2007). The Social Risk domain is new; it represents the choice problem over lotteries over pairs of consumption for *self* and for *other*.

To test the predictions of our theory, we need to know the *particular* personal state to which each social state is indifferent—not just the *fact* that every social state is indifferent to some personal state. For some of our subjects, this would require making additional assumptions about the form of the underlying preferences. However, for two important classes of subjects, the predictions of our theory *are* directly testable:

- For subjects who are *selfish*—those who, in the Social Choice domain, give nothing to *other*—the theory predicts that choice behavior in the Personal Risk domain should *coincide* with choice behavior in the Social Risk domain.
- For subjects who are *impartial*—those who, in the Social Choice domain, treat *other* symmetrically to *self*— the theory predicts that choice behavior in the Social Choice domain should *coincide* with choice behavior in the Social Risk domain.

Among our subjects, we find many who are completely (or at least extremely) selfish and a number who are completely (or at least extremely) impartial.[7] For these subjects,

[7]The objects of choice in the Social Choice domain are payout pairs (x, y) where x is the payout to *other* and y is the payout to *self*. In the experimental setting, we offer choices from linear budget lines $px+qy = w$. Behavior in the Social Choice domain is *selfish* if the choice subject to the budget constraint $px + qy = w$ is always of the form $(0, y)$. Behavior in the Social Choice domain is *symmetric* if (a, b) is chosen subject to the budget constraint $px+qy = w$ if and only if (b, a) is chosen subject to the mirror-image budget constraint $qx + py = w$. We will provide a formal nonparametric test of impartial behavior in the Social Choice domain.

the theory of revealed preference allows us to provide an individual-level nonparametric *permutation test* of these predictions, based on the observation that if preferences in two domains are the same then choice behavior in those two domains should also be the same, and hence any random selection of choices from the *union* of the choices in those domains should be *indistinguishable* from the actual choices in the two domains.

The analysis is complicated by the fact that individual choices frequently involve at least some errors: subjects may compute incorrectly, or execute intended choices incorrectly, or err in other less obvious ways. Because of these "mistakes" subjects' preferences need not be consistent *within* a choice domain—their choices need not conform perfectly with the Generalized Axiom of Revealed Preference (GARP)—so these "mistakes" must be taken into account in the permutation test for consistency *across* the various choice domains; the discussion in Section 7 explains how we do this. The important feature of our test is that it is purely *nonparametric*, in the sense that we make no functional form assumptions about subjects' underlying preferences and the specification of the error structure.

We note that there is considerable heterogeneity in the choice behaviors both among selfish subjects and among impartial subjects. Despite this heterogeneity, our theoretical predictions are well supported by the experimental data for the large majority of selfish or impartial subjects.[8] We emphasize that a minority of selfish (resp. impartial) subjects do display *different* choice behavior in the Personal Risk (resp. Social Choice) and Social Risk domains. According to the theory, the preferences of these subjects are *not consistent* across the various choice domains. This might be because the preferences of these subjects violate one or more of our axioms or because we incorrectly identify these subjects as selfish or impartial.

The remainder of the paper is organized as follows. Section 2 provides the theoretical framework for our analysis. Section 3 contains our main theoretical result. Section 4 provides the transition from the general theory to the implications for the experimental setting. Section 5 describes the experiment, Section 6 describes the data, and Section 7 presents the tests of the theory. Section 8 presents an extension of the theoretical analysis to the setting in which we can observe comparisons between social states and personal lotteries. Section 9 describes how the paper is related to prior research, and Section 10 provides some concluding remarks. All proofs are gathered in the Appendix. Some additional discussion and all technical details are relegated to an Online Appendix.

2 Framework

We consider a $\mathcal{DM}$ and a given set of outcomes Ω with a distinguished proper subset $P \subset \Omega$. For convenience, we refer to elements of Ω as *social states* and to elements of P

[8]To provide a benchmark—and for the sake of completeness—we also test the same predictions of our theory for non-selfish and non-impartial subjects, as well as linkages between preferences on which our theory has no testable implications.

as *personal states*. In the interpretation discussed in the Introduction, a social state has consequences for society (of which the $\mathcal{DM}$ is a member) as a whole; a personal state has consequences only for the $\mathcal{DM}$. We stress, however, that this is only an interpretation: our abstract formalization is quite general and encompasses many other interpretations. We are agnostic about the specific natures of Ω, P in part because outside observers may differ with respect to their knowledge of the relevant social and personal states and with respect to what they observe.

We assume Ω is finite; this avoids subtle issues about the topology of Ω and the continuity properties of preferences. We assume that $P \subset \Omega$ contains at least two states that the $\mathcal{DM}$ does not find indifferent (there are states $A, B \in P$ with $A \succ B$) and that $\Omega \setminus P \neq \emptyset$; this avoids degeneracy. For any subset $\Theta \subset \Omega$, we write $L(\Theta)$ for the set of lotteries over states in Θ. We frequently write $\sum_i p_i\omega_i$ for the lottery whose outcome is the state ω_i with probability p_i. We refer to lotteries in $L(P)$ as *personal lotteries* and to lotteries in $L(\Omega)$ as *social lotteries*.

We assume that the $\mathcal{DM}$ has a preference relation $\succeq$ on $L(\Omega)$ that satisfies the familiar requirements of Completeness, Transitivity, and Continuity, and that the $\mathcal{DM}$ is indifferent between any two lotteries $\sum_i p_i\omega_i$ and $\sum_j q_j\theta_j$ that assign the same total probabilities to each state $\omega \in \Omega$.[9] Throughout, we also assume that $\succeq$ obeys the following requirement, which we call *State Monotonicity*.

State Monotonicity If $\omega_i, \omega_i' \in \Omega$ for $i = 1, \ldots k$, $\omega_i \succeq \omega_i'$ for each i and $p = (p_1, \ldots, p_k)$ is a probability vector, then

$$\sum_{i=1}^{k} p_i\omega_i \succeq \sum_{i=1}^{k} p_i\omega_i'$$

State Monotonicity is *equivalent* to a condition that Grant et al. (1992) call *Degenerate Independence* and analogous to a condition that Savage (1954) calls *Event Monotonicity*.[10] As we show below, in the presence of the other axioms, requiring that the preference relation obeys State Monotonicity is equivalent to requiring that it respects First Order Stochastic Dominance (FOSD). The formulation in terms of State Monotonicity permits a clearer comparison with the familiar (von Neumann and Morgenstern, 1947) Independence Axiom.

Independence If $W_i, W_i' \in L(\Omega)$ for $i = 1, \ldots k$, $W_i \succeq W_i'$ for each i and $p =$

[9]We caution the reader that, in the absence of the Independence Axiom, which we do not assume, the Continuity Axiom is stronger than the Archimedean Axiom.

[10]As we do here, Grant et al. (1992) asks to what extent *all* preference comparisons can be deduced from a *subset* of preference comparisons. However, because our intent is different from Grant et al. (1992) we face quite different issues so the differences are greater than the similarities. We elaborate on this in the discussion of the related literature in Section 9.

$(p_1, \ldots, p_k)$ is a probability vector, then

$$\sum_{i=1}^{k} p_i W_i \succeq \sum_{i=1}^{k} p_i W_i'$$

We have formulated State Monotonicity in terms of weak preference, rather than indifference, because the two formulations are not equivalent. We have formulated Independence in terms of weak preference, rather than indifference, despite the fact that the two formulations *are* equivalent, in order to highlight the difference between State Monotonicity and Independence. Notice that the difference between these two axioms is precisely that Independence posits comparisons between *lotteries over lotteries*, while State Monotonicity only posits comparisons between *lotteries over states*. As we discuss below, the difference is enormous.

The following Lemma tells us that in the presence of the other axioms a preference relation obeys State Monotonicity exactly when it respects FOSD. Given a preference relation $\succeq$, say that the lottery X *first-order stochastic dominates* the lottery X', and write $X \geq_{\text{FOSD}} X'$, if, for all states $\omega \in \Omega$ we have

$$\sum \{p_i : \omega_i \succeq \omega\} \geq \sum \{p_i' : \omega_i' \succeq \omega\}$$

That is: $X \geq_{\text{FOSD}} X'$ if, for all states ω, the probability that the realization of X is weakly preferred to ω is at least as great as the probability that the realization of X' is weakly preferred to ω.[11] We say that the preference relation $\succeq$ *respects FOSD (on lotteries)* if

$$X \geq_{\text{FOSD}} X' \implies X \succeq X'.$$

Lemma *In the presence of the other axioms (Continuity, Completeness, Transitivity), the preference relation $\succeq$ on $L(\Omega)$ satisfies State Monotonicity if and only if it respects FOSD.*

A choice that does not respect FOSD may be regarded as an error—a failure to recognize that some allocation yields a payoff distribution with unambiguously lower returns (Hadar and Russel, 1969)—so this principle seems compelling and is generally accepted in decision theory. All decision-theoretic models that have been proposed as alternatives to Expected Utility of which we are aware obey FOSD, including Weighted Expected Utility (Dekel, 1986; Chew, 1989), Rank Dependent Utility (Quiggin, 1982, 1993), and (much of) Prospect Theory (Tversky and Kahneman, 1992). Indeed, various theories of choice under uncertainty have been amended explicitly to avoid failure to respect FOSD (Quiggin, 1990; Wakker, 1993).

[11] If the states yielded monetary outcomes, with preferred states yielding more money, this would reduce to the usual definition of FOSD; see Quirk and Saposnik (1962), for example. Mas-Colell et al. (1995) defines FOSD in terms of ranking by all increasing utility functions and establishes via a Proposition that the two definitions are equivalent.

Example 1 To understand the relationship of our system of axioms to others, assume that Ω consists of three mutually non-indifferent states $\Omega = \{A, X, B\}$ with $A \succ X \succ B$, and picture the familiar Marschak–Machina triangle depicted in the panels of Figure 1 in which each point represents a lottery $aA + xX + bB$ over the states A, X, B ($a = 0$ on the horizontal edge, $x = 0$ on the hypotenuse, and $b = 0$ on the vertical edge). Continuity implies that X is indifferent to some lottery over A, B; say $X \sim \frac{1}{2}A + \frac{1}{2}B$. Assuming the other axioms, Independence implies that the preference relation $\succeq$ admits an Expected Utility representation, so that the indifference curves in the triangle are parallel straight lines. Hence, knowledge that $X \sim \frac{1}{2}A + \frac{1}{2}B$ completely determines $\succeq$ on the entire triangle. In particular, $\frac{1}{2}A + \frac{1}{2}X \sim \frac{3}{4}A + \frac{1}{4}B$, $\frac{1}{2}X + \frac{1}{2}B \sim \frac{1}{4}A + \frac{3}{4}B$ and so forth (see Figure 1A).

Betweenness, which is a weaker axiom than Independence (and is the central axiom in Weighted Expected Utility), implies that all indifference curves are again straight lines but they need not be parallel; in particular, it may be that $\frac{1}{2}A + \frac{1}{2}X \sim \frac{1}{8}A + \frac{7}{8}B$ and $\frac{1}{2}X + \frac{1}{2}B \sim \frac{1}{10}A + \frac{9}{10}B$ (see Figure 1B). In this example, the indifference curves "fan out," becoming steeper (corresponding to higher risk aversion) when moving northeast in the triangle. Rank Dependent Utility and Prospect Theory allow for indifference curves that are not straight lines and can "fan out" or "fan in," especially near the triangle boundaries (see Figure 1C).

Our system of axioms is weaker than any of these, but our system still has bite. To see this, consider a lottery $aA + xX + bB$ and the lottery $(a+\varepsilon)A + (x-\varepsilon)X + bB$ obtained by shifting ε of the probability mass from X to A. Because State Monotonicity—together with the other axioms—implies respect for FOSD, we see that

$$(a+\varepsilon)A + (x-\varepsilon)X + bB \succeq aA + xX + bB$$

Thus, our axioms imply that the preference relation $\succeq$ is increasing (from bottom to top) along vertical lines. Similarly, $\succeq$ is increasing (from right to left) along horizontal lines. It follows that the indifference curves of $\succeq$ must be "upward sloping" (pointing northeast in the triangle)—but they can otherwise be quite arbitrary (see Figure 1D).

As a concluding point, suppose we are given *any* family $\mathcal{C}$ of non-self-intersecting curves that fill out the triangle and enjoy the following properties: no two curves intersect; each curve connects a point on one of the sides AX, XB with the hypotenuse AB. Any such family $\mathcal{C}$ constitutes the indifference curves of a Complete, Transitive, Continuous preference relation $\succeq_{\mathcal{C}}$ defined on lotteries (points in the triangle) by: $\Phi \succeq_{\mathcal{C}} \Psi$ if and only if the indifference curve through Φ is above the indifference curve through Ψ. Moreover, the preference relation

> $\succeq_{\mathcal{C}}$ satisfies State Monotonicity (respects FOSD) if and only if each of the curves $C \in \mathcal{C}$ is upward sloping: for each pair of points c, c' on the same curve C, either c is above and to the right of c' or vice versa (see also Figure 1D).

[Figure 1 here]

The $\mathcal{DM}$'s preference relation $\succeq$ over *all* lotteries $L(\Omega)$ is fixed, but not known (to us). We seek to *deduce* $\succeq$ but must base this deduction on observation/knowledge/inference of only a *subset* of all comparisons. Our main focus is on a setting in which we observe only the $\mathcal{DM}$'s comparisons between pairs of social states and comparisons between pairs of personal lotteries—but not comparisons between social states and personal lotteries. That is, we observe the *sub-relation* $\succeq_0$ whose graph is:

$$\text{graph}(\succeq_0) = \text{graph}(\succeq) \cap \Big(\big[L(P) \times L(P)\big] \cup \big[\Omega \times \Omega\big] \Big)$$

In Section 8 we consider a setting in which we also observe the $\mathcal{DM}$'s comparisons between social states and personal lotteries. That is, we observe the *sub-relation* $\succeq_1$ whose graph is:

$$\text{graph}(\succeq_1) = \text{graph}(\succeq) \cap \Big(\big[\Omega \cup L(P)\big] \times \big[\Omega \cup L(P)\big] \Big)$$

Observing $\succeq_0$ means observing the restriction of $\succeq$ to Ω *and* the restriction of $\succeq$ to $L(P)$, whereas observing $\succeq_1$ means observing the restriction of $\succeq$ to $\Omega \cup L(P)$. Our main focus is on the setting in which we observe $\succeq_0$ rather than $\succeq_1$ because the former seems more natural and can more easily be presented in an experimental setting.

> **Example 2** To illustrate the difference between observing $\succeq_0$ and $\succeq_1$, consider once again the setting in which Ω consists of the three states A, X, B. In addition, assume that A, B are personal states and X is a social state for which $A \succ X \succ B$ (so that X is a social state that is not indifferent to any personal state). We picture this in the Marschak–Machina triangle depicted in the panels of Figure 2. The areas in the triangle shaded gray represent the ordering of lotteries that can be inferred.
>
> If we observe $\succeq_0$ we observe the ordering $A \succ_0 X \succ_0 B$ and the ordering of lotteries between A, B but no others (see Figure 2A). State Monotonicity assures us that from these observations we can *infer* the ordering of lotteries between A, X and lotteries between X, B (see Figure 2B). The Continuity Axiom assures us that X is indifferent to *some* lottery $\alpha A + (1-\alpha)B$, but we do not observe *which lottery*. If we observe $\succeq_1$ then we *do* observe which lottery—but that is all (see Figure 2C). However, if we observe $\succeq_1$ and we *assume* that $\succeq$ obeys Independence—and hence has an Expected Utility representation—then observing *which lottery* completely determines $\succeq$ (see Figure 2D).

[Figure 2 here]

This example illustrates that the possibility of deducing the $\mathcal{DM}$'s entire preference relation $\succeq$ from a sub-relation depends not only on the amount that can be observed about $\succeq$ but *also* on the degree of rationality the observer ascribes to the $\mathcal{DM}$—in particular on whether the observer believes/assumes that the $\mathcal{DM}$'s preferences obey the Independence Axiom.

3 Deducing Preferences

We can now formalize our question in the following way: If we observe the sub-relation $\succeq_0$ can we *deduce* (infer) the entire preference relation $\succeq$? In different words: is $\succeq$ (over *all* social lotteries) the *unique* complete relation that extends the partial relation $\succeq_0$ (over social states and over personal lotteries) and obeys the axioms of Completeness, Transitivity, Continuity, and State Monotonicity?

We show that a necessary and sufficient condition that it be possible to deduce the entire preference relation $\succeq$ from the sub-relation $\succeq_0$ is that the $\mathcal{DM}$ finds every social state to be indifferent to some personal state. When this necessary and sufficient condition is *not* satisfied, there will be *many* lotteries in $L(\Omega)$ over which the preference ordering of the $\mathcal{DM}$ $\succeq$ *cannot* be deduced from $\succeq_0$.

Theorem 1 *Assume that the $\mathcal{DM}$'s preference relation $\succeq$ satisfies Completeness, Transitivity, Continuity, and State Monotonicity. In order that $\succeq$ can be deduced from $\succeq_0$ it is necessary and sufficient that the $\mathcal{DM}$ finds every social state $\omega \in \Omega \setminus P$ to be indifferent to some personal state $\tilde{\omega} \in P$.*

Although we defer the proof of Theorem **1** to the Appendix, we note here that the proof is more subtle than might have been expected. By definition, deducing $\succeq$ from $\succeq_0$ means that $\succeq$ is the *unique* preference relation that extends $\succeq_0$ and satisfies the axioms. Thus, to establish that the condition is necessary, we must show that if the condition *fails* then we can construct a preference relation $\succeq^*$ that *differs from* $\succeq$, extends $\succeq_0$, and obeys the axioms. As we shall see, guaranteeing that the candidate $\succeq^*$ obeys State Monotonicity will require some care.[12]

It might be useful to provide some interpretation of the condition that the $\mathcal{DM}$ finds every social state $\omega \in \Omega \setminus P$ to be indifferent to some personal state $\tilde{\omega} \in P$. Suppose social states represent allocations of money to society—of which the $\mathcal{DM}$ is a member—in addition to the current allocation and that the personal states are those social states in which no (additional) money is allocated to others (which is the setting in our experiment). To say that the $\mathcal{DM}$ finds every social state to be indifferent to some personal state means

[12]Our Theorem **1** bears similarity to a special case of the results of (Nishimura et al., 2017). We elaborate on this in the discussion of the related literature in Section 9.

that, for every possible allocation of additional money to society, there is some other allocation that gives *nothing* additional to others and that the $\mathcal{DM}$ finds equally desirable. Put loosely in political terms: no matter what is proposed for society, there is a "bribe" that the $\mathcal{DM}$ could be offered that would leave him/her indifferent between accepting the bribe and implementing the social proposal—and, implicitly, some "bribe" that the $\mathcal{DM}$ would strictly prefer.

4 Testable Implications

This section provides a bridge between the general theory described above and our experiment, which is designed to test the implications of the theory. We first describe the choice domains in the experimental design and their theoretical properties, and then develop a number of theoretical results that are testable on the basis of experimental observations.

It is convenient to isolate the argument for sufficiency in Theorem **1** and extend it to a setting in which we consider only some family of lotteries, rather than all lotteries. To this end, fix a non-empty set Π of probability vectors. For each non-empty subset $\Theta \subset \Omega$, let $L_\Pi(\Theta)$ be the set of lotteries of the form $p_1\theta_1 + \ldots + p_k\theta_k$, where $(p_1, \ldots, p_k) \in \Pi$ and $\theta_1, \ldots, \theta_k \in \Theta$. We identify the lottery $p_1\theta + \ldots + p_k\theta$ with θ itself, so $\Theta \subset L_\Pi(\Theta)$. If Π is the set of all probability vectors then $L_\Pi(\Theta) = L(\Theta)$; in particular, $L_\Pi(P) = L(P)$ and $L_\Pi(\Omega) = L(\Omega)$. Recall that observing $\succeq_0$ means observing the restrictions of $\succeq$ to Ω and to $L(P)$. Thus, the following proposition generalizes the sufficient condition in Theorem **1**.

Proposition 1 *Let Π be a non-empty set of probability vectors and let $\succeq$ be a preference relation on $L_\Pi(\Omega)$ that satisfies Completeness, Transitivity, Continuity, and State Monotonicity. In order that $\succeq$ can be deduced from its restrictions $\succeq_\Omega$ to Ω and $\succeq_{L_\Pi(P)}$ to $L_\Pi(P)$, it is sufficient that the $\mathcal{DM}$ finds* every *social state $\omega \in \Omega \setminus P$ to be indifferent to some personal state $\tilde{\omega} \in P$.*

In the experiment there is a subject *self* (the $\mathcal{DM}$) and an (unknown) *other*. The set of social states Ω consists of *monetary* payout pairs (a, b), where $b \geq 0$ is the payout for *self* and $a \geq 0$ is the payout for *other*. However, the set of lotteries we can present to (human) subjects is limited; in fact we consider only *equiprobable binary* lotteries: $\frac{1}{2}(a, b) + \frac{1}{2}(c, d)$. In the framework of Proposition 2, $\Pi = \{(\frac{1}{2}, \frac{1}{2})\}$. To simplify notation, let $\mathbb{L} = L_\Pi(\Omega)$ be the set of all such equiprobable binary lotteries.[13] Within $\mathbb{L}$ we distinguish three subsets:

$$\begin{aligned} \mathcal{PR} &= \{\tfrac{1}{2}(0,b) + \tfrac{1}{2}(0,d)\} \\ \mathcal{SC} &= \{\tfrac{1}{2}(a,b) + \tfrac{1}{2}(a,b)\} \\ \mathcal{SR} &= \{\tfrac{1}{2}(a,b) + \tfrac{1}{2}(b,a)\} \end{aligned}$$

Notice that whereas $\mathbb{L}$ is a 4-dimensional convex cone, each of the three subsets $\mathcal{PR}$, $\mathcal{SC}$, $\mathcal{SR}$ is a 2-dimensional convex sub-cone and can be easily presented using our graphical experimental interface (details are below).

Using different brackets $\langle x,y\rangle$, (x,y), $[x,y]$ as reminders that we are thinking of the pair x,y as representing, respectively, a personal lottery in $\mathcal{PR}$, a social state in $\mathcal{SC}$, and a social lottery in $\mathcal{SR}$, we can interpret choice in each domain—Personal Risk, Social Choice, Social Risk—as choice in the corresponding subset above by making obvious identifications:

Personal Risk	$\langle x,y\rangle \mapsto \frac{1}{2}(0,x) + \frac{1}{2}(0,y)$
Social Choice	$(x,y) \mapsto \frac{1}{2}(x,y) + \frac{1}{2}(x,y)$
Social Risk	$[x,y] \mapsto \frac{1}{2}(x,y) + \frac{1}{2}(y,x)$.

Thus, in the Personal Risk domain, the objects of choice are equiprobable lotteries in which *self* receives a payoff and *other* receives nothing. In this sense, choices in the Personal Risk domain have consequences for the $\mathcal{DM}$ and only for the $\mathcal{DM}$. In the Social Choice domain the objects of choice are deterministic payout pairs for *self* and *other*; in the Social Risk domain, the objects of choice are equiprobable social lotteries over symmetric pairs of payouts for *self* and for *other*.

Let $\succeq_{\mathbb{L}}$ be a preference relation on $\mathbb{L}$ and write $\succeq_{\mathcal{PR}}$, $\succeq_{\mathcal{SC}}$, $\succeq_{\mathcal{SR}}$ for its restrictions (sub-preference relations) to $\mathcal{PR}$, $\mathcal{SC}$, $\mathcal{SR}$, respectively. The restriction $\succeq_{\mathcal{PR}}$ of $\succeq_{\mathbb{L}}$ to $\mathcal{PR}$ prescribes preferences over personal lotteries and the restriction $\succeq_{\mathcal{SC}}$ of $\succeq_{\mathbb{L}}$ to $\mathcal{SC}$ prescribes preferences over social states. Thus to observe $\succeq_0$ in this setting is *exactly* to observe both $\succeq_{\mathcal{PR}}$ and $\succeq_{\mathcal{SC}}$ so Proposition 2 provides a sufficient condition that $\succeq_{\mathbb{L}}$ can be deduced from $\succeq_{\mathcal{PR}}$and $\succeq_{\mathcal{SC}}$. In particular, we can deduce the restriction $\succeq_{\mathcal{SR}}$ to $\mathcal{SR}$ from $\succeq_{\mathcal{PR}}$ and $\succeq_{\mathcal{SC}}$. That is, if we observe $\succeq_{\mathcal{PR}}$ and $\succeq_{\mathcal{SC}}$, and every social state is indifferent to some personal state (a condition that is determined completely by $\succeq_{\mathcal{SC}}$), then we can deduce $\succeq_{\mathcal{SR}}$.

But to test this implication, it is not enough to know just the mere *fact* that every social state is indifferent to some personal state. For each social state we need to know a *particular* personal state to which that social state is indifferent. For some of our subjects, this would require making additional assumptions about the form, parametric or otherwise, of the

[13] Fudenberg and Levine (2012) also study preference relations on the set $\mathbb{L}$ of equiprobable binary lotteries but their purpose is quite different: they are primarily interested in social fairness. They show that familiar theories of social fairness—Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002), Andreoni and Miller (2002), and Cox et al. (2008)—that are defined over riskless outcomes cannot be extended to lotteries—even equiprobable binary lotteries—without violating either the Independence Axiom or suggested notions of fairness over risky outcomes.

underlying preferences. However, for the two classes of subjects—*selfish* and *impartial*—defined below, we can construct a formal nonparametric test.

Definition 1 *We say that preferences $\succeq_{\mathcal{SC}}$ in the* Social Choice *domain are* selfish *if* $(x, y) \sim_{\mathcal{SC}} (0, y)$ *and* symmetric *if* $(x, y) \sim_{\mathcal{SC}} (y, x)$ *for all* (x, y). *We say that the* $\mathcal{DM}$ *is* selfish *when the preferences* $\succeq_{\mathcal{SC}}$ *are* selfish *and* impartial *when the preferences* $\succeq_{\mathcal{SC}}$ *are* symmetric.

For a selfish $\mathcal{DM}$, we test whether choice behavior in the Personal Risk domain *coincides* with choice behavior in the Social Risk domain; for an impartial $\mathcal{DM}$, we test whether that choice behavior in the Social Choice domain *coincides* with choice behavior in the Social Risk domain (so an impartial $\mathcal{DM}$ is immune to social risk).

With an obvious abuse of language, we shall say that *preferences $\succeq_{\mathcal{PR}}$ in the* Personal Risk *domain coincide with preferences $\succeq_{\mathcal{SR}}$ in the* Social Risk *domain* provided that for all x, y, x', y' it is the case that

$$\langle x, y\rangle \succeq_{\mathcal{PR}} \langle x', y'\rangle \iff [x, y] \succeq_{\mathcal{SR}} [x', y'].$$

Similarly, we say that *preferences $\succeq_{\mathcal{SC}}$ in the* Social Choice *domain coincide with preferences $\succeq_{\mathcal{SR}}$ in the* Social Risk *domain* provided that for all x, y, x', y' it is the case that

$$(x, y) \succeq_{\mathcal{PR}} (x', y') \iff [x, y] \succeq_{\mathcal{SR}} [x', y'].$$

In our experiments, we present subjects with a sequence of consumer decision problems: selection of a bundle of commodities from a standard budget line. We write $\mathbb{B}$ for the set of all budget lines with a typical budget line

$$\mathcal{B} = \{(x, y) \in \Omega : p_x x + p_y y = w\}$$

where $p_x, p_y > 0$ and w is the $\mathcal{DM}$'s budget. The $\mathcal{DM}$ can choose any allocation that satisfies the budget constraint which represents an allocation between accounts x, y (corresponding to the usual horizontal and vertical axes). The actual payoffs of a particular choice in a specific domain are determined by the allocation to the x and y accounts, according to the particular domain—Personal Risk, Social Choice, Social Risk. With these preliminaries in hand, we can state the theoretical predictions for selfish subjects and impartial subjects in the experimental setting.

Proposition 2 *If preferences $\succeq_{\mathcal{SC}}$ in the* Social Choice *domain are* selfish *then preferences $\succeq_{\mathcal{PR}}$ in the* Personal Risk *domain coincide with preferences $\succeq_{\mathcal{SR}}$ in the* Social Risk *domain. In particular: if preferences $\succeq_{\mathcal{SC}}$ in the* Social Choice *domain are selfish then choice behavior in the* Personal Risk *domain and choice behavior in the* Social Risk *domain coincide. So if $\mathcal{B}$ is a budget line then $\langle x, y\rangle \in \arg\max_{\mathcal{B}}(\succeq_{\mathcal{PR}})$ in the* Personal Risk *domain if and only if $[x, y] \in \arg\max_{\mathcal{B}}(\succeq_{\mathcal{SR}})$ in the* Social Risk *domain.*

Proposition 3 *If preferences* $\succeq_{\mathcal{SC}}$ *in the* Social Choice *domain are* symmetric *then preferences* $\succeq_{\mathcal{SC}}$ *in the* Social Choice *domain coincide with preferences* $\succeq_{\mathcal{SR}}$ *in the* Social Risk *domain. In particular: if preferences* $\succeq_{\mathcal{SC}}$ *in the* Social Choice *domain are symmetric then choice behavior in the* Social Choice *domain and choice behavior in the* Social Risk *domain coincide—so if* $\mathcal{B}$ *is a budget line then* $(x, y) \in \arg\max_{\mathcal{B}}(\succeq_{\mathcal{SC}})$ *in the* Social Choice *domain if and only if* $[x, y] \in \arg\max_{\mathcal{B}}(\succeq_{\mathcal{SR}})$ *in the* Social Risk *domain.*

To summarize: selfish subjects find any (x, y) to be indifferent to $(0, y)$—we say that $(0, y)$ is the *selfish equivalent* of (x, y)—and impartial subjects find any (x, y) to be indifferent to (y, x). Thus, Proposition 2 asserts that, for selfish subjects, preferences $\succeq_{\mathcal{SR}}$ in the Social Risk domain coincide with preferences $\succeq_{\mathcal{PR}}$ in the Personal Risk domain. Proposition 3 states that for impartial subjects preferences $\succeq_{\mathcal{SR}}$ in the Social Risk domain are immune to risk so these preferences coincide with the preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain. We note that our theory does not rule out that preferences in the Personal Risk and Social Risk domains can coincide $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ for non-selfish $\mathcal{DM}$ and preferences in the Social Choice and Social Risk domains can coincide $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$ for the non-impartial $\mathcal{DM}$, nor that preferences in other domains can coincide for selfish and/or impartial $\mathcal{DM}$.[14]

5 The Experiment

We next describe our experiment, which is designed to test the theory described above. We conducted the experiment at the University of Bergen and NHH Norwegian School of Economics. The 276 subjects in the experiment were recruited from undergraduate classes in these institutions. Further information about the subject pool and full experimental instructions, including the computer program dialog windows, are available in the Online Appendix.

In our experiment, subjects choose a bundle from a budget line; the subjects can choose any bundle that satisfies this constraint. A choice represents an allocation between accounts x, y (corresponding to the usual horizontal and vertical axes). These budget lines

[14]For example, assume that the preferences in the Social Choice domain $\succeq_{\mathcal{SC}}$ can be represented by the utility function proposed by Charness and Rabin (2002):

$$u(x, y) = (\rho r + \sigma s)y + (1 - \rho r - \sigma s)x,$$

where, as in our notation, x is the payout for *other* and y is the payout for *self*, $s = 1$ ($r = 1$) if $x > y$ ($x < y$) and zero otherwise. The parameters ρ and σ allow for a range of different social preferences—proportionally increasing ρ and σ indicates a decrease in self-interestedness whereas increasing the ratio ρ/σ indicates an increase in concerns for increasing aggregate payoffs rather than reducing differences in payoffs. The $\mathcal{DM}$ is selfish when $\sigma = \rho = 0$ and impartial when $\sigma = \rho = 1$. If we assume that $\succeq_{\mathbb{L}}$ (the induced preference relation on $\mathbb{L}$) admits an Expected Utility representation then $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ also for non-selfish $\mathcal{DM}$ and $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}=\succeq_{\mathcal{SC}}$ for impartial $\mathcal{DM}$. We thank a referee for pointing this out.

are presented using the graphical interface introduced by Choi et al. (2007b).[15,16] Subjects make choices by using the computer mouse to move the pointer on the computer screen to the desired point, and were restricted to allocations on the budget constraint.[17]

The actual payoffs of a particular choice in a particular experimental domain are determined by the allocation to the x and y accounts, according to the particular domain. In the experiment we consider three domains, corresponding to the domains discussed above:

- In the PERSONAL RISK domain *self* (the subject) receives the tokens allocated to one of the accounts x or y, determined at random with equal probability; the tokens allocated to the other account are lost. This domain involves only risk to *self*—*other* receives nothing—and is identical to the (symmetric) risk experiment of Choi et al. (2007a).

- In the SOCIAL CHOICE domain, *self* receives the tokens allocated to y account, while *other* (an anonymous other subject, chosen at random from the group of other subjects in the experiment) receives the tokens allocated to the x account. This domain involves only selfishness and altruism, and is identical to the (linear) two-person dictator experiment of Fisman et al. (2007).

- In the SOCIAL RISK domain, *self* receives the tokens allocated to one of the accounts x or y, determined at random with equal probability; *other* receives the tokens allocated to the other account. This domain is new: it involves risky social choices (whose consequences are not for *self* alone).

[15]Ahn et al. (2014) extended the work of Choi et al. (2007a) on risk to settings with ambiguity. Building on the experimental methodology and utilizing the CentERpanel (a nationally representative panel of households in the Netherlands), Choi et al. (2014) relate findings on individual-level behaviors from the experimental data with economic information and socio-demographic information on individuals. The datasets of Choi et al. (2007a, 2014) have also been analyzed by Halevy et al. (2018), Polisson et al. (2020), de Clippel and Rozen (2023), and Echenique et al. (2023), among others. Fisman et al. (2015b,a, 2017, 2023), and Li et al. (2017, 2022) build on the work of Fisman et al. (2007) to study social preferences with different samples, including the American Life Panel (ALP) (a nationally representative U.S.-based sample). Because all experimental designs share the same graphical interface, we are building on the expertise we have acquired in previous work.

[16]Of course it is possible that presenting choice problems graphically biases choice behavior in some particular way, but there is no evidence that this is the case. For instance: behavior in the SOCIAL domain elicited graphically (Fisman et al., 2007) is quite consistent with behavior elicited by other means (Camerer, 2003), and behavior in the PERSONAL RISK domain elicited graphically (Choi et al., 2007a) is quite consistent with behavior elicited by other means (Holt and Laury, 2002).

[17]In the two-person dictator experiment of Fisman et al. (2007), choices were not restricted to lie on the budget line. They report that most subjects had no violations of budget balancedness using a narrow confidence interval (those who did violate budget balancedness also had many GARP violations even among the subset of their choices that were on the budget constraint). All future experiments thus restricted choices to allocations on the budget constraint, which simplified the decision problem and made the computer program easier to use.

Each experimental subject faced 50 independent decisions in each of the three domains. For each subject, the computer selected 50 budget lines randomly from the set of lines that intersect at least one axis at or above the 50 token level and intersect both axes at or below the 100 token level. Each subject faced exactly the same 50 budget lines in each domain, but the order of presentation was randomized between domains. The budget lines selected for each subject in his/her decision problems were independent of each other and of the budget lines selected for other subjects in their decision problems. In the Personal Risk and Social Risk domains, subjects were not informed of the account that was actually selected until the end of the experiment. This procedure was repeated until all 50 rounds were completed.[18]

The experimental subjects first faced the Social Risk domain (because it is the centerpiece of the analysis). The order of the other experimental domains—Personal Risk and Social Choice —was counterbalanced across sessions to balance out domain order effects.[19] At the beginning of the experiment subjects received only general instructions on the experimental procedures and the use of the computer interface. At the beginning of each domain, subjects received specific instructions for that domain but not for subsequent domains. Each part of the experiment ended after all subjects had made all their decisions.

At the end of the experiment, the computer randomly selected one of the 50 decision rounds from each of the three domains of the experiment to carry out for payoffs. The round selected from each domain depended solely on chance. In the Social Choice and Social Risk domains, each subject then received the tokens that he/she allocated to *self* in the round and the subject with whom he/she was matched received the tokens that she allocated to *other*. The computer program ensured that no two subjects were ever paired as both *self*-*other* and *other*-*self*.[20] Payoffs were calculated in terms of tokens and then converted into money.[21]

[18]The x- and y-axes were scaled from 0 to 100 tokens. The resolution compatibility of the budget lines was 0.2 tokens, and the appearance and behavior of the pointer were set to the Windows mouse default. At the beginning of each decision round, the subject was presented with a budget line, with the pointer positioned randomly on the line. At the end of each decision round, the experimental program dialog window went blank, after which the entire setup reappeared for the next decision round.

[19]We also had an Observer treatment where each subject faced the same menu of 50 budget lines representing monetary payoffs for two (anonymous) *others*, but that treatment does not provide testable implications of our theory so we make no use of it here. The order effect analysis provided in the Online Appendix shows that subjects' behaviors remain consistent regardless of the sequence in which the experimental treatments—including the Observer—were presented.

[20]As is customary in social preference experiments (Andreoni and Miller, 2002; Fisman et al., 2007, among others) each subject received two groups of tokens, one based on his/her own decision to allocate tokens and another based on the decision of another random subject to allocate tokens. A concern with this payout method is that it may create a sense of reciprocity among subjects. But in both Andreoni and Miller (2002) and Fisman et al. (2007) the fraction given to *other* is about 20%, similar to the average reported by Camerer (2003) in a summary of dictator games. The computer arranges pairings between subjects so that if subject i could receive tokens from subject j then subject j could not receive tokens from subject i.

[21]Each token was worth 1.2 Norwegian Krone (NOK) (approximately 0.2 USD). A 100 NOK participation fee and subsequent earnings, which averaged about 270 NOK, were paid in private at the end of the session.

6 Data Description

We next provide an overview of the basic features of the experimental data. The experiments provide us with a very rich data set. For each subject we observe a choice from each of 50 budget lines in each of the experimental domains—Personal Risk, Social Choice, Social Risk—and this yields a rich data set that is well-suited to analysis at the level of the individual subject without the need to pool data or assume that preferences are identical across subjects. Most importantly, the changes in relative prices are such that budget lines cross frequently. This means that our data lead to high power tests of revealed preference conditions (Choi et al., 2007b).

6.1 Aggregate Behavior

In this section, we provide an overview of some important features of the experimental data, which we summarize by reporting the distribution of allocations in a number of ways. The black histogram in Figure 3 below depicts the distribution (across individuals) in the Social Choice domain of the average across all choices of the number of tokens kept by *self* as a fraction of the total of tokens allocated to *self* and *other*; that is, the average across all choices of the fraction $y/(x+y)$. On the horizontal axis we show bins of the average $y/(x+y)$; on the vertical axis we show the fraction of subjects whose average is in each bin. As might be expected, there were very few subjects whose averages are much below the midpoint of 0.5; of our 276 subjects, only six (2.2%) kept on average fewer than 0.45 of the tokens and of these only two kept fewer than 0.4.

If we classify a subject as *selfish* if it allocates 95% of tokens to *self* in the Social Choice domain (that is, the average satisfies $y/(x+y) > 0.95$) and as *impartial* if it allocates 45-55% of tokens to self (that is, the average satisfies $0.45 < y/(x+y) < 0.55$) then, of our 276 subjects, 103 (37.3%) are classified as selfish and 19 (6.9%) are classified as impartial. (In our formal tests of the theory, we offer alternative classifications of subjects as selfish or impartial.) Among the 19 impartial subjects, two subjects *always* allocated all their tokens to *self* when $p_y < p_x$ and to *other* when $p_y > p_x$, which is consistent with utilitarian preferences (with respect to money); two subjects *always* made approximately equal allocations regardless of prices (with an average relative difference, $|x-y|/(x+y)$, of less than 0.05), which is consistent with Rawlsian preferences (with respect to money). As the histogram in Figure 3 shows, there is a great deal of heterogeneity among the subjects who are neither selfish nor impartial.

Because the Personal Risk and Social Risk domains are symmetric (the two accounts x and y were equally likely) and budget lines are drawn from a symmetric distribution, reporting the distribution of the average $y/(x+y)$ would not be very informative. Instead, the dark and light gray histograms in Figure 3 depict the distributions in the Personal Risk and Social Risk domains of the fraction of tokens allocated to the *cheaper* account (that is, to x when $p_x < p_y$ and to y otherwise). The distributions are quite similar: both have a mode near the midpoint of 0.5, fall off sharply above the midpoint, and have no observations below the midpoint of 0.5. Recall that because the two accounts are equally likely, any decision to allocate *fewer* tokens to the *cheaper* account would not respect FOSD.

Of our 276 subjects, 41 subjects (14.9%) allocated more than 0.95 of the the (available) tokens to the cheaper account in the Personal Risk domain; this is consistent with risk neutrality. 30 subjects (10.9%) allocated more than than 0.95 of the tokens to the cheaper account in the the Social Risk domain; this is consistent with utilitarianism (in money). Only 9 subjects (3.3%) allocated less than 0.55 of the tokens to the cheaper account in the Personal Risk domain, which is consistent with infinite risk aversion. And only 9 subjects (3.3%) allocated less than 0.55 of the tokens to the cheaper account in the Social Risk domain, which is consistent with Rawlsianism (in money). Among the remaining subjects, there is considerably heterogeneity in choice behavior in both the Personal Risk and Social Risk domains.

[Figure 3 here]

6.2 Individual Behavior

The aggregated data above tells us little about the choice behavior of individual subjects. To get some idea of the wide range of behavior observed, Figure 4 displays scatterplots of choices of four subjects. For ease of exposition, we have chosen subjects who we classified as selfish on the basis of their choices in the Social Choice domain. In these scatterplots, each entry shows the subject's relative demand $y/(x+y)$ at a given log-price ratio $\ln(p_y/p_x)$ in the Personal Risk and Social Risk domains. (We show all 50 choices for each subject in each of these domains.) We chose these particular subjects because their behavior corresponds to one of several prototypical preference relations and because their behavior illustrates both the striking regularity *within* subjects and the heterogeneity *across* subjects that is characteristic of all our data. These scatterplots also demonstrate the sensitivity of decisions to changes in relative prices in terms of token shares in all domains. The scatterplots for all subjects (in all domains) are available upon request.

[Figure 4 here]

Because all four of these subjects are selfish in the Social Choice domain, the prediction of the theory is that their choice behavior in the Personal Risk domain should coincide with their choice behavior in the Social Risk domain. This was true for three of the subjects but not the fourth. ID 511 (see Figure 4A) allocated all the tokens to the cheaper account in the Personal Risk and Social Risk domains; this behavior is consistent with risk neutrality in the Personal Risk domain and utilitarianism in the Social Risk domain. ID 635 (see Figure 4B) chose nearly equal expenditures ($p_x x = p_y y$) in the

PERSONAL RISK and SOCIAL RISK domains; this behavior is consistent with maximizing the utility function $\log x + \log y$ in both domains.[22]

ID 317 (see Figure 4C) allocated all the tokens to the cheaper account for extreme price ratios but chose equal allocations for intermediate price ratios. In the PERSONAL RISK domain, this subject seems to be 'switching' between risk neutrality and infinite risk aversion. Each of these behaviors is consistent with Expected Utility, but not with the same underlying felicity function. This subject's choices are suggestive of *disappointment aversion* (Dekel, 1986; Gul, 1991) where the safe allocation $x = y$ is the reference point.[23] Interestingly, this subject displays the same choice behavior in the SOCIAL RISK domain, as the theory predicts. The fourth subject, ID 645 (see Figure 4D), almost always allocated all the tokens to the cheaper account in the PERSONAL RISK domain but not in the SOCIAL RISK domain.

As noted, all four of the subjects in Figure 4 are selfish; the first three display the same choice behaviors in the PERSONAL RISK and SOCIAL RISK domains, as the theory predicts; the fourth subject does not. Of course, these are special cases for which the regularities in the data are very clear. Choice behavior is much less clear for many other subjects, and there is no obvious taxonomy that allows us to classify all subjects unambiguously. Furthermore, an inspection of scatterplots cannot provide an adequate test of the theory for most subjects. This is the purpose of our individual-level revealed preference tests described below.

6.3 Testing Rationality

Because subjects' consistency (or lack of it) *within* a domain must be taken into account when testing for consistency *across* domains, we begin by measuring the extent to which subjects' behavior in each of the three domains is consistent with utility maximization. Afriat's (1967) Theorem tells us that a finite number of individual choices can be rationalized by a well-behaved utility function if and only if the data satisfies GARP.[24] Because our subjects make choices in a wide range of budget lines, our data provide a strong test of utility maximization.

Let $\{(\mathbf{p}^i, \mathbf{x}^i)\}_{i=1}^{50}$ be the data generated by some individual's choices, where $\mathbf{p}^i$ denotes the i-th observation of the price vector and $\mathbf{x}^i$ denotes the associated allocation. An

[22]No selfish subject made nearly equal allocations in the PERSONAL RISK and SOCIAL RISK domains; this behavior would be consistent with infinite risk aversion in the PERSONAL RISK domain and Rawlsianism in the SOCIAL RISK domain.

[23]The utility function in Gul (1991) takes the form $\min\{\alpha u(x) + u(y), u(x) + \alpha u(y)\}$, where $\alpha \geq 1$ is a parameter measuring disappointment aversion and the safe allocation $x = y$ is taken to be the reference point. If $\alpha > 1$ there is a kink at the 45-degree line, which corresponds to an allocation with a certain payoff. Expected Utility is the special case when $\alpha = 1$. See Choi et al. (2007a) for more information on this representation.

[24]See Varian (1982, 1983) for more details. For excellent overviews of the literature, see Chambers and Echenique (2016) and the papers by Afriat (2012), Diewert (2012), Varian (2012) and Vermeulen (2012) published in a special volume of the *Economic Journal* on the foundations of Revealed Preference.

allocation $\mathbf{x}^i$ is *directly revealed preferred* to $\mathbf{x}^j$ denoted $\mathbf{x}^i R^D \mathbf{x}^j$ if $\mathbf{p}^i\mathbf{x}^i \geq \mathbf{p}^i\mathbf{x}^j$ and *strictly directly revealed preferred* if the inequality is strict. The relation *indirectly revealed preferred* denoted $\mathbf{x}^i R \mathbf{x}^j$ is the transitive closure of the directly revealed preferred relation. GARP requires that if $\mathbf{x}^i$ is indirectly revealed preferred to $\mathbf{x}^j$, then $\mathbf{x}^j$ is not strictly directly revealed preferred to $\mathbf{x}^i$.

We assess how well individual choice behavior complies with GARP by using Afriat's (1972) Critical Cost Efficiency Index (CCEI), which measures the fraction by which each budget constraint must be tightened in order to remove all violations of GARP. Formally: for $0 \leq e \leq 1$, define the direct revealed preference relation $R^D(e)$ as

$$\mathbf{x}^i R^D(e)\mathbf{x}^j \Longleftrightarrow e\mathbf{p}^i\mathbf{x}^i \geq \mathbf{p}^i\mathbf{x}^j,$$

and define $R(e)$ to be the transitive closure of $R^D(e)$. The CCEI is the largest value of e such that the relation $R(e)$ satisfies GARP. By definition, the CCEI is between 0 and 1; CCEI closer to 1 means the data are closer to perfect consistency with GARP and hence to perfect consistency with utility maximization.

Mean CCEI's across all subjects are 0.959, 0.952, and 0.902 in the Personal Risk, Social Choice and Social Risk domains, respectively. Figure 5 depicts the distributions of CCEI scores in our three domains. The horizontal axis presents bins of CCEI ranges; the vertical axis indicates the percent of subjects whose CCEI is in each bin. The fact that for most subjects, choices are sufficiently consistent to be considered utility-generated in all three domains is a striking result in its own right (more below). Nevertheless, the distribution of CCEI scores is generally further to the left for the Social Risk domain. This might be expected, because the Social Risk domain seems more complicated and less familiar than the other domains. Of our 276 subjects, the CCEI scores of 248 (89.9%) and 237 (85.9%) subjects were above 0.90 in the Personal Risk and Social Choice domains, respectively, while only 193 (69.9%) were as high in the Social Risk domain.[25]

[Figure 5 here]

We interpret the CCEI scores as confirmation that subject choices are generally consistent with utility maximization but there is no natural threshold for determining whether subjects are close enough to satisfying GARP. To provide additional evidence, we follow Bronars (1987), which builds on Becker (1962), and compare the behavior of our actual subjects to the behavior of simulated subjects who randomize uniformly on each budget line. Mean CCEI's for 100,000 simulated subjects are only 0.585. Figure 6 compares the distributions of the minimum and maximum CCEI scores in the three domains for the actual subjects to the distribution of the CCEI scores generated by the simulated subjects. This provides a clear graphical illustration of the extent to which subjects did worse than

[25]For comparison, Cappelen et al. (2023) compare the consistency of the choices of students in the US and Tanzania in the Personal Risk domain. If we follow the threshold of 0.9 for the CCEI, the corresponding percentages are 85.7% and 52.3% for the US and the Tanzania subjects, respectively.

choosing consistently and the extent to which they did better than choosing randomly. Of our 276 subjects, 160 (58.0%) subjects have a minimum CCEI score above 0.90, while only a very few simulated subjects have CCEI's that high.

[Figure 6 here]

The Bronars (1987) test rules out the possibility that consistency is the accidental result of random behavior, but it cannot tell whether utility maximization is the correct model. To this end, Choi et al. (2007b) and Fisman et al. (2007) propose generating a sample of hypothetical subjects who maximize a utility function with an idiosyncratic preference shock that has a logistic distribution.[26] Their analysis provides a clear benchmark of the extent to which subjects do worse than choosing consistently and the extent to which they do better than different levels of bounded rationality, and demonstrates that if utility maximization is not in fact the correct model, then our experiment is powerful enough to detect it. We refer the interested reader to Choi et al. (2007b, 2014) for more details on the use of GARP to test for consistency and a discussion of various alternative measures that have been proposed for this purpose by Varian (1990, 1991), Echenique et al. (2011) and Houtman and Maks (1985).[27] The subjects' CCEI scores, and the alternative consistency scores, in the three domains are available from the authors upon request. In practice, all indices yield similar conclusions.

7 Testing the Theory

As explained in Section 4, to test the predictions of our theory for a given subject, we need not only to *know* that the subject finds each social state to be indifferent to a personal state, but also to *identify* which social states are found indifferent to which personal state. For some of our subjects, this would require postulating a parametric form for the underlying utility function. However, as we have shown, for *selfish* and for *impartial* subjects the predictions of our theory *are* testable:

- If the subject's preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain are selfish then preferences in the Personal Risk and in the Social Risk domains coincide: $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ (Proposition 2).

[26]Specifically, the hypothetical subjects implement (with error) the power utility function (commonly employed in the empirical analysis) in the Personal Risk domain and a constant elasticity of substitution (CES) utility function for giving (also commonly employed in the empirical analysis) in the Social Choice domain.

[27]Varian (1990, 1991) refined Afriat's CCEI to provide a score that reflects the minimum adjustment required to eliminate the violations of GARP associated with each budget constraint. The score of Echenique et al. (2011) is based on the idea that an individual who violates GARP can be exploited as a "money pump." The discrepancies between the CCEI and the Varian (1990, 1991) score and the money pump score are discussed in Echenique et al. (2011). Houtman and Maks (1985) finds the largest subset of choices that is consistent with GARP.

- If preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain are symmetric then preferences in the Social Choice domain and preferences in the Social Risk domain coincide: $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$ (Proposition 3).

7.1 Subject Classification

We have defined a subject to be *selfish* if its preferences in the Social Choice domain are selfish—obey $(x, y) \sim_{\mathcal{SC}} (0, y)$ for all (x, y). Such a subject would allocate all tokens to *self* in the Social Choice domain so $y/(x + y) = 1$. To allow for errors, we set lower thresholds: 0.99, 0.975, 0.95, and 0.90. Using these thresholds, of our 276 subjects, 68 (24.6%), 86 (31.2%), 103 (37.3%), and 129 (46.7%) of subjects, respectively, are classified as selfish.

Similarly, we have defined a subject to be *impartial* if its preferences in the Social Choice domain are symmetric—obey $(x, y) \sim_{\mathcal{SC}} (y, x)$ for all (x, y). One way to classify a subject as impartial is by requiring that the average of the tokens kept as a fraction of the sum of the tokens kept and given $y/(x + y)$ is between 0.45 and 0.55. By this criterion, 19 (6.9%) are classified as impartial. An alternative approach to classify a subject as impartial builds on the revealed preference techniques used above. If a subject has a complete and transitive preference ordering $\succeq_{\mathcal{SC}}$ in the Social Choice domain then the choice data in that domain should satisfy GARP. If the subject is also impartial—that is, preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain are also symmetric—then the union of the data and the *mirror-image* data (and *a fortiori*, any 50-element subset of this union) should also satisfy GARP.[28] By definition, the CCEI score for the combined data set can be no bigger than the CCEI score for the actual data. Relying on Nishimura et al. (2017), Polisson et al. (2020) provide an easy-to-implement (necessary and sufficient) test of whether preferences are also symmetric.

Because, as we have already noted, for many subjects, choices do not satisfy GARP *exactly*, we draw at random 10,000 50-element subsets from the union of the data and the mirror-image data, where each draw is made independently and with equal probability from the data or the mirror-image data. The actual data set from the Social Choice domain is obviously a particular realization of the permuted data sets. If preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain are symmetric, we should expect the CCEI of scores of the permuted data sets to be at least very close to the actual CCEI score; if these scores are substantially below then we should reject the null that preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain are symmetric.[29] By this criterion, of our 276 subjects, we classify

[28] The data generated by a subject's choices are $\left\{\left(\bar{x}^i, \bar{y}^i, x^i, y^i\right)\right\}_{i=1}^{50}$, where $(\bar{x}^i, \bar{y}^i)$ are the endpoints of the budget line. Thus, the i-th budget line is given by $x^i/\bar{x}^i + y^i/\bar{y}^i = 1$ and the price ratio $p_x^i/p_y^i = \bar{y}^i/\bar{x}^i$. The mirror-image data are obtained by reversing the prices and the associated allocation for each observation $\left\{\left(\bar{y}^i, \bar{x}^i, y^i, x^i\right)\right\}_{i=1}^{50}$.

[29] The permutation tests we use for this purpose are similar to those we use to test Propositions 2 and 3; the discussion in the following subsection explains how we do this, so we skip the technical details here

30 (10.9%), 33 (12.0%), and 37 (13.4%) subjects as impartial using 10%, 5%, and 1% significance levels, respectively. Of the 19 subjects classified as impartial by keeping an intermediate fraction on average ()$0.45 < y/(x+y) < 0.55$), 16 (84.2%) are also classified as impartial by the nonparametric revealed preference test at all significance levels.[30]

Finally we note that among the selfish subjects there is substantial heterogeneity in choice behaviors in the Personal Risk and in the Social Risk domains, and among the impartial subjects there is substantial heterogeneity in choice behaviors in the Social Choice and in the Social Risk domains; this facilitates a serious test of the implications of the theory for these subjects (more below).

7.2 A Nonparametric Test

To test whether preferences, and hence choice behavior, in two domains coincide, an obvious approach would be to compare the two choices (one in each domain) from each budget line. However, such an obvious approach will not do, for several reasons. The first is that while different choices might arise from different preferences, they might also arise from indifference; there is no reason to believe optimal choices are unique. The second is that many choices involve errors, so—even if we were to assume that optimal choices are unique—different choices might arise from different realizations of errors. And, once we admit the possibility of errors, it is not clear how far apart choices should be to be regarded as different.

An alternative approach would be to impose parametric forms for the underlying utility functions in the different domains, derive the associated demand functions, fit these to the data, and test to see if they conform to the special restrictions imposed by the theory. The inherent shortcomings of this approach are precisely that it *is* parametric: The utility functions postulated must be good approximations of the "true" underlying preferences—a hypothesis that is not directly testable—and the conclusions will be sensitive to the functional forms, the estimation technique, and the manner in which the error term is introduced.

Instead we create an individual-level nonparametric *permutation* test (Good, 2005). This approach builds on the revealed preference techniques used above to test the consistency of choice behavior *within* each domain to test for consistency of choice behaviors *across* domains. It is *nonparametric*, making no assumptions about the form of the subject's underlying utility functions in the three domains—Personal Risk, Social Choice

and refer the interested reader to the Online Appendix.

[30]For comparison, of our 276 subjects, we reject that preferences in the Personal Risk domain are symmetric for only 4 (1.8%), 10 (3.6%), and 14 (5.1%) using 1%, 5%, and 10% significance levels, respectively. The corresponding numbers in the in the Social Risk domain are also very small—8 (2.9%), 17 (6.2%), and 26 (9.4%). We note that in the case where the states are equally likely (as in our Personal Risk and Social Risk domains), requiring preferences to respect FOSD is equivalent to requiring them to be symmetric.

and Social Risk—and allowing for the reality that subjects' behavior is not *perfectly* consistent with utility maximization.

The basis of our test is the following observation: If a subject has a complete and transitive preference ordering $\succeq_{\mathcal{I}}$ in some domain $\mathcal{I}$ then the set of choices in that domain should satisfy GARP. If preferences $\succeq_{\mathcal{I}}$ and $\succeq_{\mathcal{J}}$ in the domains $\mathcal{I}$ and $\mathcal{J}$ (respectively) are the same—that is, $\succeq_{\mathcal{I}}=\succeq_{\mathcal{J}}$—then the union of the sets of choices in these two domains (and *a fortiori*, any 50-element subset of this union) should also satisfy GARP. However, our actual test cannot be so simple because, as we have already noted, for many subjects, choices in a given domain do not satisfy GARP *exactly*, so we should certainly not expect that choices across two domains should satisfy GARP *exactly*. Instead, we view the observed choice from a given budget line $\mathcal{B} \in \mathbb{B}$ in domain $\mathcal{I}$ as a random draw from some distribution function $F_{\mathcal{I}}^{\mathcal{B}}$ over all allocations that satisfy the budget constraint. If preferences are the same in the two domains $\mathcal{I}$ and $\mathcal{J}$, then choices in the two domains should be independent draws from the *same* distribution function—that is, $F_{\mathcal{I}}^{\mathcal{B}} = F_{\mathcal{J}}^{\mathcal{B}}$. This is the null hypothesis we test.

Formally, let $\{(\mathbf{p}^i, \mathbf{x}_{\mathcal{I}}^i)\}_{i=1}^{50}$ and $\{(\mathbf{p}^i, \mathbf{x}_{\mathcal{J}}^i)\}_{i=1}^{50}$ be the data generated by some individual's choices in the two domains $\mathcal{I}$ and $\mathcal{J}$, where $\mathbf{p}^i$ denotes the i-th observation of the price vector and $\mathbf{x}_{\mathcal{I}}^i$ and $\mathbf{x}_{\mathcal{J}}^i$ denote the associated choices in domains $\mathcal{I}$ and $\mathcal{J}$, respectively. There are $\binom{100}{50}$ possible distinct 50-element subsets of this union; each such subset is formed by drawing the choice from domain $\mathcal{I}$ or $\mathcal{J}$ for each of the 50 budget lines $\mathbf{p}^i$. (Recall that subjects see the *same* 50 budget lines in each domain.) Clearly we cannot examine all $\binom{100}{50}$ possible subsets; instead we draw 10,000 subsets at random, where each draw is made independently and with equal probability from the choice in $\mathcal{I}$ or from the choice $\mathcal{J}$.

Note that the actual data sets from domains $\mathcal{I}$ and $\mathcal{J}$ are simply the particular realizations in which each choice happened to be drawn from the same domain. Similarly, the actual CCEI scores in the two domains $\mathcal{I}$ and $\mathcal{J}$, denoted by $e_{\mathcal{I}}$ and $e_{\mathcal{J}}$ respectively, are simply realizations from the distribution of CCEI scores calculated for each of the permuted data sets $\{(\mathbf{p}^i, \mathbf{x}^i)\}_{i=1}^{50}$ we draw. If $e_{\mathcal{I}}, e_{\mathcal{J}} = 1$ (so actual choices within each domain $\mathcal{I}$ and $\mathcal{J}$ are perfectly consistent), we should expect the CCEI of scores of the permuted data sets to be equal to 1—or at least very close; if these scores are substantially below 1 then we should reject the null that preferences in two domains $\mathcal{I}$ and $\mathcal{J}$ coincide—that is, $\succeq_{\mathcal{I}}=\succeq_{\mathcal{J}}$.[31]

To obtain a distribution function F for the test statistic under the null hypothesis, for each subject we randomly draw 10,000 data sets $\{(\mathbf{p}^i, \mathbf{x}^i)\}_{i=1}^{50}$, calculate the CCEI score for each of these data sets, and compare the distribution of CCEI scores to the actual CCEI

[31] In the permutation test, we focus on the CCEI, which offers a straightforward interpretation and is the most commonly used index in the revealed preference literature. Performing the test using the other indices we have mentioned—Varian (1990, 1991), Echenique et al. (2011) and Houtman and Maks (1985)—for each subject for each of our 10,000 permuted datasets is computationally intensive, especially if, roughly speaking, there were a large number of GARP violations.

scores $e_{\mathcal{I}}$ and $e_{\mathcal{J}}$. Set $e^- = \min\{e_{\mathcal{I}}, e_{\mathcal{J}}\}$ and $e^+ = \max\{e_{\mathcal{I}}, e_{\mathcal{J}}\}$ and let p^- and p^+ be the corresponding p-values. Under the null, we approximate

$$p^- = (1 - \widehat{F}(t - \epsilon))^2 \text{ and } p^+ = 1 - (\widehat{F}(t - \epsilon))^2,$$

where $\widehat{F}$ is the estimate of the permutation distribution function F and $\epsilon > 0$ is small (introduced to account for discrete bunching in the permuted CCEI scores illustrated in Figure 7 below). To counteract the problem of multiple comparisons, in addition to p^- and p^+ we also use the Bonferroni correction,

$$\min\{2\min\{p^-, p^+\}, 1\},$$

as our p-value. This is known to be conservative but difficult to improve on without imposing further structure (Bland and Altman, 1995).

To illustrate our test, we consider the four selfish subjects identified in Figure 7. Because the preferences $\succeq_{\mathcal{SC}}$ of these subjects in the Social Choice domain are selfish then their preferences in the Personal Risk and in the Social Risk domains should coincide $\succeq_{\mathcal{PR}} = \succeq_{\mathcal{SR}}$ (Proposition 2). Each panel of Figure 7 presents a histogram of the permuted CCEI scores and the two actual CCEI scores from the Personal Risk and Social Risk domains for one of these selfish subjects. For ID 505 (see Figure 7A) $e^- = 1$ and all the permuted CCEI scores are also equal to 1 so we do not reject the null that this subject's preferences in Personal Risk and Social Risk domains coincide: $\succeq_{\mathcal{PR}} = \succeq_{\mathcal{SR}}$. However, for ID 729 (see Figure 7B), $e^- < e^+ < 1$ and all permuted CCEI scores are below e^- so we do reject the null that $\succeq_{\mathcal{PR}} = \succeq_{\mathcal{SR}}$.

For the other two subjects identified in Figure 7 it is more difficult to draw clear conclusions because many of the permuted CCEI scores for these subjects lie (weakly) between e^- and e^+. For ID 514 (see Figure 7C), e^+ is sufficiently far into the (right) tail of the permuted CCEI scores that the Bonferroni correction allows us to reject the null that $\succeq_{\mathcal{PR}} = \succeq_{\mathcal{SR}}$. For ID 502 (see Figure 7D), e^- and e^+ are not extreme with respect to the distribution of the permuted CCEI scores, and we cannot reject the null $\succeq_{\mathcal{PR}} = \succeq_{\mathcal{SR}}$. (Diagrams for all subjects are available upon request.)

[Figure 7 here]

Table 1 provides population-level summaries of the individual-level test results. Table 1A tabulates the number of subjects we classify as selfish/non-selfish and Table 1B tabulates the number of subjects we classify as impartial/non-impartial, based on their choices in the Social Choice domain, as discussed above. We present the results—using the Bonferroni correction—for the conventional significance levels 1% (left panels), 5% (middle panels) and 10% (right panels). In each cell of Table 1, we tabulate the percent of subjects for whom we can reject the null that preferences coincide. In Table 1A (resp. Table 1B), the top entry at each cell is for the selfish (resp. impartial) subjects and the bottom entry is for the non-selfish (resp. non-impartial) subjects.

In column (1) of each panel, we test the null that preferences in the Personal Risk and Social Risk domains coincide ($\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$); in column (2) we test the null that preferences in the Social Choice and Social Risk domains coincide ($\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$); in column (3) we test the null that preferences in the Personal Risk and Social Choice domains coincide ($\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SC}}$). In column (4) of each panel, we test the null that preferences in *all three domains*—Social Choice, Personal Risk and Social Risk—coincide ($\succeq_{\mathcal{SC}}=\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$).[32] We thus test the theoretical predictions above—that the preferences coincide in Personal Risk and Social Risk $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ for the selfish subjects (Proposition 2) and in Social Choice and Social Risk $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$ for the impartial subjects (Proposition 3). The tests of the theoretical predictions reported in column (1) in Table 1A and column (2) in Table 1B are presented in bold type. In addition to testing our theoretical predictions, the tests that preferences in other domains coincide might reveal other linkages between preferences, about which our existing theory has little to say.

[Table 1 here]

Selfish (Table 1A) At the 1% percent significance level, we reject the null that preferences coincide in the Personal Risk and Social Risk domains (Proposition 2) for only 5.9%-9.3% of the selfish subjects across all classifications (I)-(IV). The corresponding rejection rates for non-selfish subjects are also low, ranging between 14.7% and 16.3%, but substantially higher than the rejection rates for the selfish subjects (see column (1), left panel). At the 5% and 10% percent significance levels, we reject the null that preferences coincide in the Personal Risk and Social Risk domains for 20.4%-25.6% of the selfish subjects. The rejection rates for the non-selfish are quite similar, ranging from 22.4% to 31.8% (see column (1), middle and right panels). As we pointed out above, our theory does not rule out that preferences coincide in the Personal Risk and Social Risk domains also for non-selfish subjects. We reject the null that preferences coincide in any other two domains—as well as in all three domains—for nearly all selfish subjects and the majority of non-selfish subjects at all significance levels (see columns (2)-(4)).

Impartial (Table 1B) At all significance levels, we reject the null that preferences coincide in the Social Choice and Social Risk domains (Proposition 3) for 15.8%-35.1% of the impartial subjects across all classifications (I)-(IV). The corresponding rejection rates for the non-impartial subjects are substantially higher, ranging between 86.0% to 95.0%. The rejection rates of both

[32] Testing that the preferences in all three domains coincide is an obvious extension of testing that preferences in two domains coincide, so we skip the technical details to economize on space and refer the reader to the Online Appendix.

impartial and non-impartial subjects are very similar across significance levels (see column (2)). However, among the impartial subjects we also reject the null that preferences coincide in the Personal Risk and Social Choice domains—as well as in all three domains—at even lower rates at the 1% and 5% significance levels (see columns (3) and (4), left and middle panels). As we pointed out above, our theory does not rule out that preferences also coincide in the Personal Risk and Social Risk domains for impartial subjects. Perhaps as expected, for the impartial subjects, we reject the null that preferences coincide in the Personal Risk and Social Risk domains at similar rates as for the non-selfish subjects (see column (1)).[33]

Our conclusion from this theoretical/empirical exercise is that for the large majority of selfish and impartial subjects, the theoretical predictions are well supported by the experimental data.

7.3 Power Analysis

Finally, we generate a benchmark with which we can compare our finding that, for most selfish and impartial subjects, the preferences coincide, exactly as Propositions 2 and 3 predict. We focus on Proposition 2 and add noise to the actual choices of the subjects who we classify as selfish according to classification III (Table 1A). Specifically, we assign a probability μ of replacement, and for each choice in each of the 10,000 randomly drawn data sets from the Personal Risk and the Social Risk domains, with probability μ we replace the actual choice with a choice drawn randomly and independently from the uniform distribution over all allocations on the budget line that allocate more tokens to the cheaper account.[34]

We then calculate the CCEI score for each of these data sets and retest, using the Bonferroni correction, to see whether $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$, as the theory predicts for selfish subjects. Rejecting the null that $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ when we replace only a small fraction of actual choices with random choices will demonstrate that the experiment is sufficiently powerful to detect if our theory is not in fact the correct model. The results are presented in Table 2 below.

Table 2A reports the fraction of selfish subjects for whom we reject the null that $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ at the 1% (top panel) 5% (middle panel) and 10% (bottom panel) levels when

[33]We replicate the tests reported in Table 1B for symmetric preferences—using the union of the data and the mirror-image data—and obtain quantitatively very similar results. We thus conclude that when preferences coincide, they are the same *symmetric* preferences for most subjects. To economize on space, the results are relegated to the Online Appendix.

[34]In the Personal Risk and Social Risk treatments, any decision to allocate fewer tokens to the cheaper account is a violation of respect for FOSD because there are other feasible allocations that yield unambiguously higher monetary payoffs. Overall, the choices made by subjects in our experiment show very low rates of FOSD violations, so we restrict the random choices to allocations that allocate more tokens to the cheaper account (positions on the longer side of the budget line relative to the 45-degree line).

we replace the actual choices with random choices with probabilities $\mu = 0, 0.05.0.1, 0.15, 0.2$. (Keep in mind that $\mu = 0$ means we are using actual choices.) The results show there is a much higher probability that we reject the null even when only a few individual choices are replaced with random choices. Table 2B reports the results of a similar test of Proposition 3—the fraction of subjects who we classify as impartial according to classification VII (Table 1B) for whom we reject the null that $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$—also yields the same conclusion. We obtain consistent results with the other classifications of subjects as selfish and impartial.

[Table 2 here]

8 Observing More

In our main theoretical result—and in the experiment—we assume that we can observe the restrictions $\succeq_0$ of $\succeq$ to Ω and to $L(P)$. That is, we observe the $\mathcal{DM}$'s comparisons between social states and the $\mathcal{DM}$'s comparisons between personal lotteries— but not the $\mathcal{DM}$'s comparisons between social states and personal lotteries. For completeness, we now discuss the setting in which we can observe the restriction $\succeq_1$ of $\succeq$ to $\Omega \cup L(P)$. That is, in addition to observing the $\mathcal{DM}$'s comparisons between social states and between personal lotteries, we *also* observe the $\mathcal{DM}$'s between social states and personal lotteries.

Perhaps surprisingly, Theorem 2 below shows that, in the absence of additional strong assumptions about the $\mathcal{DM}$'s preferences, observing comparisons between social states and personal lotteries is no help at all: exactly as with $\succeq_0$, in order to deduce the complete preference relation $\succeq$ from the incomplete relation $\succeq_1$ it is necessary and sufficient that the $\mathcal{DM}$ finds every social state to be indifferent to some personal state.

Theorem 2 *Assume that the $\mathcal{DM}$'s preference relation $\succeq$ satisfies Completeness, Transitivity, Continuity, and State Monotonicity. In order that $\succeq$ can be deduced from $\succeq_1$ it is necessary and sufficient that the $\mathcal{DM}$ finds every social state $\omega \in \Omega \setminus P$ to be indifferent to some personal state $\tilde{\omega} \in P$.*

However, Theorem 3 below shows that if we are willing assume that the $\mathcal{DM}$'s preferences $\succeq$ obey the Independence Axiom, and hence have an Expected Utility representation, then observing comparisons between social states and personal lotteries makes a big difference: in order to deduce the complete preference relation $\succeq$ from the incomplete relation $\succeq_1$ it is necessary and sufficient only that the $\mathcal{DM}$ finds every social state to be indifferent to some personal *lottery*. (In light of the Continuity Axiom, the $\mathcal{DM}$ will find every social state to be indifferent to some personal *lottery* exactly when the $\mathcal{DM}$ finds every social state to be ranked (weakly) between two personal states; equivalently, when the best and worst social states are viewed as indifferent to personal states.)

Theorem 3 *Assume that the $\mathcal{DM}$'s preference relation $\succeq$ satisfies Completeness, Transitivity, Continuity, and Independence (and hence admits an Expected Utility representation).*

In order that $\succeq$ can be deduced from $\succeq_1$ it is necessary and sufficient that the $\mathcal{DM}$ finds every social state $\omega \in \Omega \setminus P$ to be indifferent to some personal lottery $\sum_i p_i\omega_i \in L(P)$.

9 Related Literature

From a purely technical point of view, our paper poses a problem in decision theory: under what circumstances is a preference relation over some set of lotteries completely determined by its restriction to a subset of lotteries? Grant et al. (1992), which is closest to the present work, pose the problem in the context of lotteries whose outcomes are commodity bundles and lotteries whose outcomes are monetary payoffs. Given fixed prices for commodities, they seek conditions guaranteeing that preferences over lotteries whose outcomes are commodity bundles are completely determined by the restrictions of those preferences to lotteries whose outcomes are monetary payoffs; the sufficient condition they identify is one we call State Monotonicity (and they call Degenerate Independence).

But because our intent is different from Grant et al. (1992), we pose different questions and face quite different issues. In particular, although prices play a crucial role for Grant et al. (1992) (prices mediate between monetary outcomes and consumption bundles), prices play no role at all in our setting. More subtly, the central issue in our setting is whether all choices in a larger set (social choices) have equivalents (are viewed as indifferent to) choices in a smaller set (personal choices). In Grant et al. (1992) it is *assumed* that all choices in the larger set have equivalents in the smaller set; the central issue is whether this condition is strong enough to determine preferences over lotteries.

Our Theorem 1 bears a family resemblance to a special case of the results of (Nishimura et al., 2017). They consider a universal set X of alternatives, a pre-order $\trianglerighteq$ on X, a set $\mathcal{A}$ of non-empty subsets of X, and a choice correspondence $c : \mathcal{A} \rightrightarrows X$. Given these data, their Theorem 1 provides a necessary and sufficient condition—a generalization of Afriat's (1967) cyclic consistency—that there exist a preference relation $\succeq$ on X that rationalizes the given choice correspondence and respects the pre-order $\trianglerighteq$. In particular, if X is a set of lotteries, a necessary and sufficient condition on the choice correspondence $c : \mathcal{A} \rightrightarrows X$ in order that it be rationalizable by a preference relation that respects FOSD is that the pre-order $\trianglerighteq$ respects FOSD (that is, lotteries $x, y \in X$ satisfy $x \trianglerighteq y$ exactly when x first-order stochastically dominates y).

Like Theorem 1 in (Nishimura et al., 2017), our Theorem 1 addresses preference relations that respect FOSD. However, where their Theorem 1 provides necessary and sufficient conditions for the *existence* of such a preference relation subject to the constraint that it be consistent with a prescribed family of *choices*, our Theorem 1 provides necessary and sufficient conditions for the *uniqueness* of a *given* preference relation subject to the constraint that it be consistent with a prescribed family of *preference comparisons*. In our context, Theorem 1 of (Nishimura et al., 2017) is trivial, because we are *given* a preference relation on the set $X = L(\Omega)$ of alternatives.[35]

Finally, we will not attempt to review the large and growing literature that examines—theoretically and/or experimentally—social preferences in the presence of risk.[36] These papers seek to disentangle concerns for *ex-ante* and *ex-post* fairness—the fairness of opportunities, rules, or processes *versus* the fairness of realized outcomes. The empirical findings in this literature have been mixed: Some of the experimental papers provide evidence of both ex-ante and ex-post fairness(Brock et al., 2013; Cappelen et al., 2013), while others argue that concerns for ex-ante fairness can largely account for the experimental data (Krawczyk and Le Lec, 2016). A recent paper that is particularly relevant to our study is Feldman and López Vargas (2024). They also present subjects with a sequence of choices from budget lines in different domains where each choice has consequences for *self* and for an *other*. Two of their domains are similar to our Social Choice domain and Social Risk domain. In another domain, both *self* and *other* receive the number of tokens allocated to one of the accounts x or y, determined at random with equal probability. They propose and estimate a *parametric* model that accommodates both ex-ante and ex-post fairness considerations, so the overall goal of their study is very different from ours. Feldman and López Vargas (2024) also provide an excellent review of this literature that the reader may wish to consult.

10 Concluding Remarks

It is often said that private choices should remain private. As Paul Krugman has written "... I'm talking about professional mistakes. The other kinds of mistakes ... are none of your business." This point of view seems reasonable when applied to Krugman, who is not a candidate for a public office. But, to the extent that choices are not mistakes but rather are the consequences of attitudes toward risk and attitudes toward personal risk are indicative of attitudes toward social risk, then this point of view would seem mistaken when applied to candidates for public office. Accordingly, we *should* care about the personal choices of those individuals who might be in a position to make choices that have consequences for others—at least to the extent that those choices involve risk.[37]

[35]We note that the argument given by (Nishimura et al., 2017) et al is non-constructive—it employs Szilprajn's extension theorem, which relies on the axiom of choice; moreover, there is no discussion of uniqueness or lack of it. By contrast, our argument establishing that uniqueness fails when our conditions are not satisfied is entirely constructive.

[36]See, (Karni and Safra, 2002, 2008; Bolton et al., 2005; Trautmann, 2009; Schildberg-Hörisch, 2010; Krawczyk and Le Lec, 2008, 2016; Fudenberg and Levine, 2012; Brock et al., 2013, 2016; Saito, 2013; López-Vargas, 2015; Feldman and López Vargas, 2024), among others.

[37]Although Krugman might be regarded as something of a celebrity among economists, he is not well-known among voters as a whole, and his private choices are not readily observable by potential voters. But there are many celebrities—Arnold Schwarzenegger, Jesse Ventura, and Volodymyr Zelenskyy, to name just a few—who have become (successful) candidates for public office, and whose private choices were observable by voters, and presumably played a role in their electoral successes.

In political science, a substantial literature argues that personal character is an important predictor of Presidential conduct. The argument is made most famously and forcefully in a classic book *The Presidential Character* (1972), by James Barber, who writes: "Character is the force, the motive power, around which the person gathers his view of the world, and from which his style receives its impetus. The issues will change; the character of the president will not." Barber argues in particular that candidate's character provides "a realistic estimate of what will endure into a man's White House years." A number of Presidents—real and fictional—and Presidential aides agree:

> With all the power that a President has, the most important thing to bear in mind is this: You must not give power to a man unless, above everything else, he has character. Character is the most important qualification the President of the United States can have. – Richard Nixon

> For the past several months ... [my opponent] ... has suggested that being President of this country was, to some extent, about character ... I have been President for three years and two days and I can tell you without hesitation that being President of this country is entirely about character. – Andrew Shepard[38]

> In a president, character is everything. A president does not have to be brilliant... He does not have to be clever; you can hire clever... You can hire pragmatic, and you can buy and bring in policy wonks. But you cannot buy courage and decency, you cannot rent a strong moral sense. A president must bring those things with him... He needs to have, in that much maligned word, but a good one nonetheless, a "vision" of the future he wishes to create. But a vision is worth little if a president does not have the character—the courage and heart—to see it through. – Peggy Noonan[39]

Neither Barber nor Noonan nor the quoted Presidents define character but Barber and others argue (explicitly or implicitly) that character is revealed by personal choices—and early in life. As Barber puts it "the personal past foreshadows the presidential future." Such an argument would seem coherent—and of use to voters—only if the candidate's attitude with respect to social policy—and in particular toward social risk—after achieving office can be deduced from the candidate's attitude with respect to personal choices—and in particular toward personal risk—before achieving office. Such a deduction would seem

[38]Fictional President, portrayed by Michael Douglas in the film *The American President* (1995), responding to a political opponent's attack on his character in a surprise appearance in the White House press room.

[39]Political writer and columnist for *The Wall Street Journal* and former speechwriter and Special Assistant to Ronald Reagan (who believed that "you can tell a lot about a fella's character" by his way of eating jelly beans).

require the existence of a strong *linkage* between the candidate's attitudes toward personal risk and social risk. Barber argues that such a strong link exists. At the other extreme, it is sometimes argued that the constraints imposed by the institutional aspects of the Presidency completely outweigh any possible influence of personal character. Nelson (2018) provides excellent discussions of the arguments.[40,41]

However, it can be dangerously easy to err and infer too much from observations that are too imperfect. During (and after) the 1992 presidential campaign, stories were widely told about Bill Clinton's personal choices, which—entirely aside from its moral content—were surely quite risky, and many pundits—and no doubt many voters—used these stories as the basis for predictions about his choices in the public domain. Such predictions did not stop with Clinton's election; a 1994 article in Newsweek, for instance, concluded that "...it may well be that this is one case where personal behavior does give an indication of how a politician will perform in the arena." History is yet to write its judgement of that prediction, but voters have already done so: Clinton left office with the highest approval rating of any President in recent history.

Appendix: Proofs

Proof of Lemma It is evident that if $\succeq$ respects FOSD then $\succeq$ obeys State Monotonicity, so we only need to prove the converse.

Assume, therefore, that $\succeq$ satisfies State Monotonicity. By assumption, Ω is finite. Hence the indifference relation partitions Ω into a finite number of equivalence classes $\Omega_1, \ldots, \Omega_J$. For each j, choose and fix a representative $\theta_j \in \Omega_j$; renumbering, if necessary, we may assume that $\theta_1 \succ \theta_2 \succ \ldots \succ \theta_J$. Set $\Theta = \{\theta_j\}$.

Every $\omega \in \Omega$ belongs to some equivalence class Ω_j and hence is indifferent to some unique θ_j. Hence every lottery $W \in L(\Omega)$ is indifferent to some unique lottery $\widehat{W} \in L(\Theta)$. It follows from Transitivity that for $W, W' \in L(\Omega)$,

$$W \geq_{\text{FOSD}} W \Longleftrightarrow \widehat{W} \geq_{\text{FOSD}} \widehat{W'} \text{ and } W \succeq W' \Longleftrightarrow \widehat{W} \succeq \widehat{W'}.$$

[40] Again, some voters might care about a candidate's personal choices on purely moral grounds, independent of the implications for choices the candidate might make or policies the candidate might follow when he or she actually assumes office—but that is *not* the argument being made by Barber and others. As Jonathan Yardley concluded "in Washington, and wherever else two or more politicians may gather, he who does not get caught has 'character' and he who gets caught has none."

[41] Other aspects of behavior might also matter, especially if they are viewed as signaling "strength" or "weakness". The reader may recall that Edmund Muskie was widely regarded as the leading candidate for the Democratic Presidential nomination in 1972 until, at a press conference during the New Hampshire primary, Muskie gave an emotional response to attacks on his wife. (Many press accounts of this incident even reported that Muskie cried.) This emotional incident was widely viewed as revealing "weakness" and appears to have fatally damaged Muskie's candidacy. It would be interesting to carry out an analysis linking "weakness in the private domain" with "weakness in the public domain"–but this does not seem easy; certainly we do not know how to do it.

Hence to show that $\succeq$ respects FOSD for lotteries in $L(\Omega)$ it will suffice to show that $\succeq$ respects FOSD for lotteries in $L(\Theta)$.

For each $j = 1, \ldots, J$ and $X \in L(\Theta)$, write $q_j(X)$ for the probability weight that X puts on θ_j. Thus,

$$X = \sum_{j=1}^{J} q_j(X)\theta_j.$$

Now consider two lotteries $X, X' \in L(\Theta)$ for which $X \geq_{\text{FOSD}} X'$. In the notation we have just introduced, FOSD means that for every $k \leq J$:

$$\sum_{j=1}^{k} q_j(X) \geq \sum_{j=1}^{k} q_j(X'). \tag{1}$$

To show that $X \succeq X'$ we first construct a chain of intermediate lotteries $X_0, X_1, \ldots, X_{J-1}$ in $L(\Theta)$ with the following properties:

(a) $X_0 = X$
(b) For $1 \leq j \leq k \leq J-1$: $q_j(X_k) = q_j(X'_k)$
(c) For $1 \leq k+1 < j \leq J$: $q_j(X_k) = q_j(X)$
(d) $q_k(X_k) + q_{k+1}(X_k) = q_k(X_{k+1}) + q_{k+1}(X_{k+1})$
(e) $X_{J-1} = X'$

As illustrated in the table below, such a chain gradually shifts probability mass from more-preferred states X to less-preferred states X', one step at a time. Note that, as is the case in the example, the intermediate lotteries might not all be different.

Table A1:

	X	X_0	X_1	X_2	X_3	X_4	X_5	X'	
q_1	.30	.30	.25	.25	.25	.25	.25	.25	
q_2	.20	.20	.25	.25	.25	.25	.25	.25	
q_3	.30	.30	.30	.30	.00	.00	.00	.00	
q_4	.10	.10	.10	.10	.40	.40	.40	.40	
q_5	.10	.10	.10	.10	.10	.10	.00	.00	
q_6	.00	.00	.00	.00	.00	.00	.10	.10	

The construction proceeds by induction. Set $X_0 = X$. Assume $X_0, \ldots, X_{k-1}$ have been constructed. It is evident that $X_{j-1} \geq_{\text{FOSD}} X_j$ for $j \leq k-1$, and equation (1) guarantees that $q_k(X_k) - q_k(X') \geq 0$. Hence we may define X_k as follows

(i) For $j < k$: $q_j(X_k) = q_j(X_{k-1})$
(ii) $q_k(X_k) = q_k(X')$
(iii) $q_{k+1}(X_k) = q_{k+1}(X_{k-1}) + [q_k(X_k) - q_k(X')]$
(iv) For $j > X_{k+1}$: $q_j(X_k) = q_j(X_{k-1})$

We assert that $X_{k-1} \succeq X_k$ for each k. To see this, note first that $q_j(X_{k-1}) = q_j(X_k)$ for all $j \neq k, k+1$. Hence if we write $\beta_j = q_j(X_{k-1}) = q_j(X_k)$ for $j \neq k, k+1$ our construction implies that we can write:

$$\begin{aligned} X_{k-1} &= \sum_{j \neq k,k+1} \beta_j \theta_j + q_k(X_{k-1})\theta_k + q_{k+1}(X_{k-1})\theta_{k+1} \\ &= \sum_{j \neq k,k+1} \beta_j \theta_j + q_k(X')\theta_k + \left[q_k(X_{k-1}) - q_k(X')\right]\theta_k + q_{k+1}(X_{k-1})\theta_{k+1}. \\ X_k &= \sum_{j \neq k,k+1} \beta_j \theta_j + q_k(X_k)\theta_k + q_{k+1}(X_k)\theta_{k+1} \\ &= \sum_{j \neq k,k+1} \beta_j \theta_j + q_k(X')\theta_k + \left[q_k(X_{k-1}) - q_k(X')\right]\theta_{k+1} + q_{k+1}(X_{k-1})\theta_{k+1}. \end{aligned}$$

We have numbered so that $\theta_k \succ \theta_{k+1}$ so State Monotonicity guarantees that $X_{k-1} \succeq X_k$, as asserted. Transitivity now implies that $X \succeq X'$, so the proof is complete. ■

Proof of Theorem 1: Sufficiency To see that this condition is sufficient, assume that every social state ω admits a personal state equivalent. We must show that $\succeq$ is the unique preference relation that satisfies the axioms and agrees with $\succeq_0$ on $L(P)$ and on Ω. To this end, suppose $\succeq^*$ is some preference relation that satisfies the axioms and agrees with $\succeq_0$ on $L(P)$ and on Ω. We must show that $\succeq^* = \succeq$; that is, for every pair of lotteries $\sum p_i\omega_i$ and $\sum q_i\omega_i$ we have the equivalence

$$\sum p_i\omega_i \succeq \sum q_i\omega_i \iff \sum p_i\omega_i \succeq^* \sum q_i\omega_i$$

By assumption, each social state ω_i admits a personal state equivalent $\tilde{\omega}_i$; that is, $\omega_i \sim_0 \tilde{\omega}_i$. Because $\succeq$ and $\succeq^*$ agree with $\succeq_0$ on Ω, it follows that $\omega_i \sim \tilde{\omega}_i$ and $\omega_i \sim^* \tilde{\omega}_i$. State Monotonicity of $\succeq$ implies that $\sum p_i\omega_i \sim \sum p_i\tilde{\omega}_i$ and $\sum q_i\omega_i \sim \sum q_i\tilde{\omega}_i$; State Monotonicity of $\succeq^*$ implies that $\sum p_i\omega_i \sim^* \sum p_i\tilde{\omega}_i$ and $\sum q_i\omega_i \sim^* \sum q_i\tilde{\omega}_i$. Transitivity provides a chain of equivalences:

$$\begin{array}{ccc} \sum p_i\omega_i & \succeq & \sum q_i\omega_i \\ & \Updownarrow & \\ \sum p_i\tilde{\omega}_i & \succeq & \sum q_i\tilde{\omega}_i \\ & \Updownarrow & \\ \sum p_i\tilde{\omega}_i & \succeq_0 & \sum q_i\tilde{\omega}_i \\ & \Updownarrow & \\ \sum p_i\tilde{\omega}_i & \succeq^* & \sum q_i\tilde{\omega}_i \\ & \Updownarrow & \\ \sum p_i\omega_i & \succeq^* & \sum q_i\omega_i \end{array}$$

It follows that $\sum p_i\omega_i \succeq \sum q_i\omega_i$ if and only if $\sum p_i\omega_i \succeq^* \sum q_i\omega_i$. Because the lotteries $\sum p_i\omega_i, \sum q_i\omega_i$ were arbitrary, we conclude that $\succeq = \succeq^*$, as asserted.

Necessity To see that this condition is necessary, assume that there is some social state X that the $\mathcal{DM}$ does *not* find indifferent to any personal state; we construct a preference relation that agrees with $\succeq$ on $L(P)$ and on Ω but not on all of $L(\Omega)$. It will require some care to ensure that the preference relation we construct obeys State Monotonicity.

For later use, we begin by choosing a utility representation for $\succeq$. To do so, note that, because $\succeq$ is continuous (by assumption) and $L(\Omega)$ can be identified with a finite-dimensional simplex, which is a separable metric space, we can use Debreu's (1954) representation theorem to find a continuous utility function $u : L(\Omega) \to \mathbb{R}$ that represents $\succeq$, that is

$$\text{for all } \Gamma, \Gamma' \in L(\Omega) : \Gamma \succeq \Gamma' \Leftrightarrow u(\Gamma) \geq u(\Gamma').$$

To make the remainder of the proof easier to follow, suppose for the moment that $\Omega = \{A, X, B\}$, $P = \{A, B\}$ and that X is a social state that is not indifferent to either A or B. As our earlier discussion of State Monotonicity suggests, it is easy to construct a utility function U on $L(\Omega)$ with the desired properties in this setting—but it is less easy to do so in a way that generalizes to the general setting with more than three states. Assume without loss that $A \succ B$. We distinguish three cases: (i) $A \succ X \succ B$, (ii) $X \succ A \succ B$, and (iii) $A \succ B \succ X$.

- **Case (i) $\mathbf{A \succ X \succ B}$:** Because this is the leading (and most interesting) case, we will actually prove a bit more than is needed. Note that this is the setting illustrated in Panel A of Figure 2. Choose and fix lotteries $D_1 = a_1A + x_1X + b_1B$ and $D_2 = a_2A + x_2X + b_2B$ with the property that $a_1 + x_1 > a_2 + x_2$ and $x_1 + b_1 < x_2 + b_2$. (Note that neither of D_1, D_2 first-order stochastically dominates the other.) We define two preference relations that satisfy our axioms but provide opposite rankings of D_1, D_2, so that at least one of them differs from the given preference relation $\succeq$.

  To this end, we first define two auxiliary functions f, g:

$$\begin{aligned} f(aA + xX + bB) &= u(aA + xA + bB) \\ g(aA + xX + bB) &= u(aA + xB + bB) \end{aligned}$$

  for every lottery $aA + xX + bB \in L(\Omega)$. Note that f, g both agree with u on $\Omega = \{A, X, B\}$ and on $L(P)$ (which is the hypotenuse AB). Because u is continuous, both f and g are continuous. Evidently, f is constant on vertical lines and strictly increasing from right to left along horizontal lines, while g is strictly increasing upward on vertical lines and constant on horizontal lines. Hence for every $\lambda \in (0, 1)$ the convex combination $\lambda f + (1 - \lambda)g$ is strictly increasing upward on vertical lines *and* strictly increasing from right to left on horizontal lines.

Taking all these things together, we conclude that, for every $\lambda \in (0,1)$ the continuous utility function

$$u_\lambda(aA + xX + bB) = \lambda f(aA + xX + bB) + (1-\lambda) g(aA + xX + bB)$$

agrees with u on Ω and on $L(P)$ and is strictly increasing upward on vertical lines and strictly increasing from right to left on horizontal lines. Hence the preference relation $\succeq_\lambda$ induced by u_λ agrees with $\succeq$ on Ω and on $L(P)$ and obeys Completeness, Continuity, Transitivity, and State Monotonicity.

Note that the definitions of u_1, u_0 imply that

$$\begin{aligned} u_1(D_1) &= u(a_1 A + x_1 A + b_1 B) > u(a_2 A + x_2 A + b_2 B) = u_1(D_2) \\ u_0(D_1) &= u(a_1 A + x_1 B + b_1 B) < u(a_2 A + x_2 B + b_2 B) = u_0(D_2). \end{aligned}$$

If we choose λ_1 sufficiently close to 1 and λ_2 sufficiently close to 0 we will obtain

$$u_{\lambda_1}(D_1) > u_{\lambda_1}(D_2) \text{ and } u_{\lambda_2}(D_1) < u_{\lambda_2}(D_2).$$

Hence the preference relations $\succeq_{\lambda_1}, \succeq_{\lambda_2}$ both satisfy our axioms but provide opposite rankings of D_1, D_2. In particular, at least one of $\succeq_{\lambda_1}, \succeq_{\lambda_2}$ must differ from $\succeq$.[42]

- **Case (ii) $\mathbf{X \succ A \succ B}$:** This is easier than Case (i), making use of the utility representation u constructed above. Continuity implies that there is some $\nu \in (0,1)$ for which $A \sim \nu X + (1-\nu)B$. Choose $\lambda > 0$ for which $u(\nu A + (1-\nu)B) + \lambda\nu \neq u(A)$ and set
$$U(xX + aA + bB) = u(xA + aA + bB) + \lambda x.$$
By construction, U agrees with u on $L(P)$ and $U(X) = u(A) + \lambda x > u(A) = U(A)$ so the preference relation $\succeq_U$ represented by U is indeed an extension of $\succeq$. It is easily checked that $\succeq_U$ satisfies all the required axioms; because $U(\nu X + (1-\nu)B) = u(\nu A + (1-\nu)B) + \lambda\nu \neq u(A)$ we conclude that $\succeq_U \neq \succeq$; this completes the construction in Case (ii).

- **Case (iii) $\mathbf{A \succ B \succ X}$:** The argument is almost the same as in Case (ii): we simply interchange the roles of A, B and change the sign of the linear term. Continuity guarantees that there is some $\eta \in (0,1)$ for which $B \sim \eta A + (1-\eta)X$. Choose $\lambda > 0$ for which $u(\eta A + (1-\eta)B) - \lambda\eta \neq u(B)$ and set
$$U(aA + bB + xX) = u(aA + bB + xB) - \lambda x.$$

[42]In fact, it can be shown that if $\lambda \neq \lambda'$ then $\succeq_\lambda \neq \succeq_{\lambda'}$, so we obtain a continuum of distinct preference relations that are Complete, Continuous and Transitive, obey State Monotonicity, and agree with $\succeq$ on Ω and on $L(P)$. We leave the proof to the interested reader.

By construction, U agrees with u on $L(P)$ and $U(X) = u(B) - \lambda x < u(B) = U(B)$ so the preference relation $\succeq_U$ represented by U is indeed an extension of $\succeq$. It is easily checked that $\succeq_U$ satisfies all the required axioms; because $U(\eta A + (1-\eta)X) = u(\eta A + (1-\eta)B) - \lambda\eta \neq u(B)$ we conclude that $\succeq_U \neq \succeq$; this completes the construction in Case (iii).

We now turn to the general setting. Here we must take account of the possible presence of additional personal and social states and of the possible differences in the ranking of the distinguished social state with respect to the additional personal states, but the main idea remains the same. Assume that there is some social state X that the $\mathcal{DM}$ does *not* find indifferent to any personal state. Write $\mathcal{A}$ for the set of states that are strictly preferred to X according to $\succeq$, $\mathcal{B}$ for the set of states that are strictly dis-preferred to X, and $\mathcal{X}$ for the set of states that are indifferent to X. Because X is not indifferent to any personal state, no member of $\mathcal{X}$ is indifferent to any personal state; moreover, at least one of $\mathcal{A}, \mathcal{B}$ is not empty.

If $\mathcal{A} \neq \emptyset$, let A be any $\succeq$-minimal element of $\mathcal{A}$; if $\mathcal{B} \neq \emptyset$ let B be any $\succeq$-maximal element of $\mathcal{B}$. (Such minimal and maximal elements exist because Ω is finite.) For each lottery $\Gamma = \sum p_i \omega_i \in L(\Omega)$ write

$$\Gamma_{\mathcal{A}} = \sum_{\omega_i \in \mathcal{A}} p_i \omega_i \ ; \ \ \Gamma_{\mathcal{B}} = \sum_{\omega_i \in \mathcal{B}} p_i \omega_i \ ; \ \ \Gamma_{\mathcal{X}} = \sum_{\omega_i \in \mathcal{X}} p_i \omega_i \ ; \ x(\Gamma) = \sum_{\omega_i \in \mathcal{X}} p_i$$

Evidently, $\Gamma = \Gamma_{\mathcal{A}} + \Gamma_{\mathcal{B}} + \Gamma_{\mathcal{X}}$. (If $\mathcal{A}$ or $\mathcal{B}$ is empty, then the corresponding sum is 0.)

We now distinguish three cases that are parallel to the three cases considered above, and carry out constructions parallel to those above.

- **Case (i) $\mathcal{A} \neq \emptyset$ and $\mathcal{B} \neq \emptyset$:** As before, define auxiliary functions $f, g : L(\Omega) \to \mathbb{R}$

  $$\begin{aligned} f(\Gamma) &= u(\Gamma_{\mathcal{A}} + x(\Gamma)A + \Gamma_{\mathcal{B}}) \\ g(\Gamma) &= u(\Gamma_{\mathcal{A}} + x(\Gamma)B + \Gamma_{\mathcal{B}}) \end{aligned}$$

  For each $\lambda \in (0,1)$, define $u; L(\Omega) \to [0,1]$ by

  $$u_\lambda(\Gamma) = \lambda f(\Gamma) + (1-\lambda) g(\Gamma)$$

  and let $\succeq_\lambda$ be the preference relation on $L(\Omega)$ induced by u_λ. Just as before, we see that, for every λ, the preference relation $\succeq_\lambda$ is Complete, Continuous and Transitive, that it obeys State Monotonicity, and that it agrees with $\succeq$ on Ω and on $L(P)$. And, just as before, we see that there is some λ^* for which $\succeq_{\lambda^*} \neq \succeq$. This completes the construction in Case (i).

- **Case (ii) $\mathcal{A} = \emptyset$ and $\mathcal{B} \neq \emptyset$:** Because there are at least two inequivalent personal states, we can choose $B' \in \mathcal{B}$ with $B \succ B'$. Continuity guarantees there is some

$\nu \in (0,1)$ for which $B \sim \nu X + (1-\nu)B'$. Choose $\lambda > 0$ so that $u(\nu B + (1-\nu)B') - \lambda\nu \neq u(B)$ and set

$$U(\Gamma) = u(x(\Gamma)B + \Gamma_{\mathcal{B}}) - \lambda x(\Gamma)$$

It is easily checked that the preference relation $\succeq_U$ induced by U satisfies all the desired axioms, and that, because $U(\nu X + (1-\nu)B') = u(\nu B + (1-\nu)B') \neq u(B)$, we conclude that $\succeq_U \neq \succeq$. This completes the construction in Case (ii).

- **Case (iii) $\mathcal{A} \neq \emptyset$ and $\mathcal{B} = \emptyset$:** Because there are at least two inequivalent personal states, we can choose $A' \in \mathcal{A}$ for which $A' \succ A$. Continuity guarantees there is some $\eta \in (0,1)$ for which $A \sim \eta A' + (1-\eta)X$. Choose $\lambda > 0$ for which $u(\eta A' + (1-\eta)A) + \lambda\eta \neq u(A)$ and set

$$U(\Gamma) = u(\Gamma_{\mathcal{A}} + x(\Gamma)A) + \lambda x(\Gamma)$$

It is easily checked that the preference relation $\succeq_U$ induced by U satisfies all the desired axioms, and that because $U(\eta A' + (1-\eta)X) = u(\eta A' + (1-\eta)A) + \lambda\eta \neq u(A)$, we conclude that $\succeq_U \neq \succeq$. This completes the construction in Case (iii).

In each case we have constructed a preference relation that extends $\succeq_0$, satisfies all of our axioms and differs from $\succeq$; therefore the proof is complete. ■

Proof of Proposition 1 Let $\succeq^*$ be any preference relation on $L_\Pi(\Omega)$ that satisfies Completeness, Transitivity, Continuity, and State Monotonicity, and for which the restriction of $\succeq^*$ to Ω agrees with $\succeq_\Omega$ and the restriction of $\succeq^*$ to $L_\Pi(P)$ agrees with $\succeq_{L_\Pi(P)}$. We must show that $\succeq^*$ agrees with $\succeq$.

By assumption, for every social state there is some (not necessarily unique) personal state $\tilde{\omega} \in P$ for which $\omega \sim_\Omega \tilde{\omega}$. By assumption, the restriction of $\succeq^*$ to Ω agrees with the restriction of $\succeq$ to Ω so $\omega \sim^* \tilde{\omega}$. State Monotonicity implies that if $(p_1, \ldots, p_k) \in \Pi$ and $\omega_1, \ldots, \omega_k \in \Omega$ then

$$\sum p_i \omega_i \sim \sum p_i \tilde{\omega}_i \text{ and} \sum p_i \omega_i \sim^* \sum p_i \tilde{\omega}_i.$$

Now fix any pair of lotteries $\sum p_i \omega_i, \sum q_j \omega_j \in L_\Pi(\Omega)$. We obtain the following chain of equivalences:

$$\begin{array}{ccc}
\sum p_i \omega_i & \succeq^* & \sum q_j \omega_j \\
 & \Updownarrow & \\
\sum p_i \tilde{\omega}_i & \succeq^* & \sum q_j \tilde{\omega}_j \\
 & \Updownarrow & \\
\sum p_i \tilde{\omega}_i & \succeq_{L_\Pi(P)} & \sum q_j \tilde{\omega}_j \\
 & \Updownarrow & \\
\sum p_i \omega_i & \succeq & \sum q_j \omega_j
\end{array}$$

(The first equivalence follows from State Monotonicity and Transitivity for $\succeq^*$; the second equivalence follows from the assumption that $\succeq^*$ and $\succeq$ agree on $L_\Pi(P)$; the third equivalence follows from State Monotonicity and Transitivity for $\succeq$.) Taken together, this string of equivalences asserts that

$$\sum p_i\omega_i \succeq^* \sum q_j\omega_j \iff \sum p_i\omega_i \succeq \sum q_j\omega_j.$$

Because this is true for all pairs of lotteries $\sum p_i\omega_i, \sum q_j\omega_j \in L_\Pi(\Omega)$, we conclude that $\succeq^* = \succeq$, as asserted. ■

Proof of Proposition 2 Fix a budget line $\mathcal{B} \in \mathbb{B}$. By definition, $\langle x, y\rangle \in \arg\max_{\mathcal{B}}(\succeq_{\mathcal{PR}})$ in the Personal Risk domain if and only if $\langle x, y\rangle \succeq_{\mathcal{PR}} \langle \hat{x}, \hat{y}\rangle$ for every $\langle \hat{x}, \hat{y}\rangle \in \mathcal{B}$. Unwinding the notation, this means that

$$\langle x, y\rangle = \tfrac{1}{2}(0, x) + \tfrac{1}{2}(0, y) \succeq_{\mathcal{PR}} \tfrac{1}{2}(0, \hat{x}) + \tfrac{1}{2}(0, \hat{y}) = \langle \hat{x}, \hat{y}\rangle$$

for every $\langle \hat{x}, \hat{y}\rangle \in \mathcal{B}$. If preferences $\succeq_{\mathcal{SC}}$ are selfish $(x, y) \sim_{\mathcal{SC}} (0, y)$ for all (x, y); this is true if and only if

$$[x, y] = \tfrac{1}{2}(y, x) + \tfrac{1}{2}(x, y) \succeq_{\mathcal{SR}} \tfrac{1}{2}(\hat{y}, \hat{x}) + \tfrac{1}{2}(\hat{x}, \hat{y}) = [\hat{x}, \hat{y}]$$

for every $[\hat{x}, \hat{y}] \in \mathcal{B}$. We conclude that preferences $\succeq_{\mathcal{PR}}$ in the Personal Risk domain coincide with preferences $\succeq_{\mathcal{SR}}$ in the Social Risk domain. Given that preferences in the two domains coincide, choice behavior coincides as well. ■

Proof of Proposition 3 Assume preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain are symmetric. Consider two choices $[x, y]$ and $[\hat{x}, \hat{y}]$ in the Social Risk domain and assume that $[\hat{x}, \hat{y}] \succ_{\mathcal{SR}} [x, y]$. Expressed explicitly in terms of lotteries, this means

$$\tfrac{1}{2}(\hat{x}, \hat{y}) + \tfrac{1}{2}(\hat{y}, \hat{x}) \succ_{\mathcal{SR}} \tfrac{1}{2}(x, y) + \tfrac{1}{2}(y, x).$$

State Monotonicity implies that either $(\hat{x}, \hat{y}) \succ_{\mathcal{SC}} (x, y)$ or $(\hat{y}, \hat{x}) \succ_{\mathcal{SC}} (y, x)$; Impartiality implies that if either of these is true then both are true. Hence, if $[\hat{x}, \hat{y}] \succ_{\mathcal{SR}} [x, y]$ in the Social Risk domain then $(\hat{x}, \hat{y}) \succ_{\mathcal{SC}} (x, y)$ in the Social Choice domain. Conversely if $(\hat{x}, \hat{y}) \succ_{\mathcal{SC}} (x, y)$ in the Social Choice domain then

$$\tfrac{1}{2}(\hat{x}, \hat{y}) + \tfrac{1}{2}(\hat{y}, \hat{x}) \succ_{\mathcal{SR}} \tfrac{1}{2}(x, y) + \tfrac{1}{2}(y, x).$$

That is: $[\hat{x}, \hat{y}] \succ_{\mathcal{SR}} [x, y]$ in the Social Risk domain. Putting these together we conclude that preferences $\succeq_{\mathcal{SC}}$ in the Social Choice domain coincide with preferences $\succeq_{\mathcal{SR}}$ in the Social Risk domain. Given that preferences in the two domains coincide, choice behavior coincides as well. ■

Proof of Theorem 2: Sufficiency If every social state is indifferent to some personal state, then Theorem **1** tells us that $\succeq$ is the unique extension of $\succeq_0$ that satisfies our axioms, so it is certainly the unique extension of $\succeq_1$ that satisfies our axioms. Hence this condition is sufficient.

Necessity If there is some social state X that the $\mathcal{DM}$ does *not* find indifferent to any personal state, we need to construct a preference relation that agrees with $\succeq$ on $L(P) \cup \Omega$ but not on all of $L(\Omega)$; the argument is almost the same as in the proof of Theorem **1**, but with a small twist. As in the proof of Theorem **1**, we will need to ensure that the preference relation we construct obeys Continuity and State Monotonicity. As before, we use Debreu's (1954) representation theorem to find a continuous utility function $u : L(\Omega) \to [0, 1]$ that represents $\succeq$. We construct a new utility function $U : L(\Omega) \to [0, 1]$ that agrees with u on $L(P) \cup \Omega$—and hence induces the same preference ordering as $\succeq_1$ on $L(P) \cup \Omega$—but does not induce the same ordering as $\succeq$ on $L(\Omega)$.

As before, define $\mathcal{A}, \mathcal{B}, \mathcal{X}$ to be the sets of social states that are strictly preferred to X, strictly dis-preferred to X and indifferent to X, respectively. If $\mathcal{A} \neq \emptyset$, let $A \in \mathcal{A}$ be a minimal element; if $\mathcal{B} \neq \emptyset$, let $B \in \mathcal{B}$ be a maximal element. For each lottery $\Gamma = \sum p_i \omega_i \in L(\Omega)$ define $\Gamma_{\mathcal{A}}$, $\Gamma_{\mathcal{B}}$, $\Gamma_{\mathcal{X}}$ and $x(\Gamma)$ exactly as before and note, as before, that $\Gamma = \Gamma_{\mathcal{A}} + \Gamma_{\mathcal{B}} + \Gamma_{\mathcal{X}}$.

We distinguish the same three cases as in Theorem **1**; the arguments for Case (ii) and Case (iii) are identical to those in the proof of Theorem **1**, but the argument for Case (i) requires a twist because we must be careful to construct the utility function U to preserve the relationship between the social state X (and all those social states in $\mathcal{X}$ that are indifferent to X) and personal *lotteries*, as well as personal *states*. (No additional care is required in Case (ii) because X is preferred to all personal lotteries, or in Case (iii) because X is dis-preferred to all personal lotteries.)

Assume therefore that neither $\mathcal{A}$ nor $\mathcal{B}$ is empty (so that we are in Case (i)). Use Continuity to choose $\gamma, \zeta \in (0, 1)$ such that

$$\begin{aligned} X &\sim \gamma A + (1-\gamma) B \\ \tfrac{1}{2} A + \tfrac{1}{2} X &\sim \zeta A + (1-\zeta) B. \end{aligned}$$

As before, define auxiliary functions $f, g : L(\Omega) \to \mathbb{R}$ by

$$\begin{aligned} f(\Gamma) &= u(\Gamma_{\mathcal{A}} + x(\Gamma) A + \Gamma_{\mathcal{B}}) \\ g(\Gamma) &= u(\Gamma_{\mathcal{A}} + x(\Gamma) B + \Gamma_{\mathcal{B}}). \end{aligned}$$

We are no longer free to choose U to be an arbitrary convex combination $U = \lambda f + (1-\lambda) g$ because we require

$$U(X) = U(\gamma A + (1-\gamma) B) = u(\gamma A + (1-\gamma) B).$$

This would completely determine λ and it might happen that, for this particular λ, the preference relation $\succeq_U$ coincides with $\succeq$. To get around this, we need to combine f, g in a different way.

Let $H : [0,1] \times [0,1] \to [0,1]$ be any continuous function which is strictly increasing in each variable separately and for which $H(s,s) = s$ for every $s \in [0,1]$. (Note that for any $\lambda \in (0,1)$ the the function $H(s,t) = \lambda s + (1-\lambda)t$ has these properties.) Given such an H, define a utility function U_H on $L(\Omega)$ by

$$U_H(\Gamma) = H\left(f(\Gamma), g(\Gamma)\right)$$

and let $\succeq_H$ be the preference relation on $L(\Omega)$ induced by U_H. Because U_H is a function, the preference relation $\succeq_H$ is Complete and Transitive. Because f, g, H are all continuous, so is U_H, and hence the preference relation $\succeq_H$ is continuous. Because H is strictly increasing in each variable separately, the preference relation $\succeq_H$ respects FOSD. Because $H(s,s) = s$ for every $s \in [0,1]$, $U_H(\Gamma) = u(\Gamma)$ for every $\Gamma \in L(\Omega)$ and $U_H(\omega) = u(\omega)$ for every $\omega \in \Omega$, so $\succeq_H$ is an extension of $\succeq_1$.

It remains only to choose an appropriate function H for which $\succeq_H \neq \succeq$; this is easy to do. (Indeed there are infinitely many such functions.) For each $\alpha, \beta \in (0,1)$ define the function $H_{\alpha,\beta} : [0,1] \times [0,1] \to [0,1]$ by

$$H_{\alpha,\beta}(s,t) = \tfrac{1}{2}\left[\alpha s^2 + (1-\alpha)t^2\right]^{1/2} + \tfrac{1}{2}\left[\beta s + (1-\beta)t\right].$$

Note that

$$f(X) = u(A) > u(X) = u(\gamma A + (1-\gamma)B > u(B) = g(X)$$

and

$$f(X) = u(A) > u(\tfrac{1}{2}A + \tfrac{1}{2}X) > u(\tfrac{1}{2}A + \tfrac{1}{2}B) > u(B) = g(X).$$

Choose the parameters α, β so that

$$\begin{aligned} U_{H_{\alpha,\beta}}(X) &= H_{\alpha,\beta}(f(X), g(X)) \\ &= u(X). \\ U_{H_{\alpha,\beta}}(\tfrac{1}{2}A + \tfrac{1}{2}X) &= H_{\alpha,\beta}(u(A), u(\tfrac{1}{2}A + \tfrac{1}{2}B)) \\ &\neq u(\zeta A + (1-\zeta)B) \\ &= u(\tfrac{1}{2}A + \tfrac{1}{2}X). \end{aligned}$$

For such a choice of α, β the preference relation $\succeq_{H_{\alpha,\beta}}$ is an extension of $\succeq_1$ that differs from $\succeq$, as desired. ■

Proof of Theorem 3: Sufficiency The argument for sufficiency is parallel to the argument for sufficiency in Theorem 1, except that we substitute personal lottery equivalents for personal state equivalents.

Necessity Assume that every social state ω admits a personal lottery equivalent. We must show that $\succeq$ is the unique preference relation that is Complete, Continuous, Transitive, obeys the Independence Axiom and agrees with $\succeq_1$ on $L(P) \cup \Omega$. To this end, suppose $\succeq^*$ is any such preference relation; we must show that $\succeq^* = \succeq$; that is, for every pair of lotteries $\sum p_i\omega_i$ and $\sum q_i\omega_i$ we have the equivalence

$$\sum p_i\omega_i \succeq \sum q_i\omega_i \iff \sum p_i\omega_i \succeq^* \sum q_i\omega_i.$$

By assumption, each social state ω_i admits a personal state equivalent Γ_ω; that is, $\omega_i \sim_1 \Gamma_\omega$. Because $\succeq$ and $\succeq^*$ agree with $\succeq_1$ on $L(P) \cup \Omega$, it follows that $\omega_i \sim \Gamma_{\omega_i}$ and $\omega_i^* \!\sim \Gamma_{\omega_i}$. Independence implies that $\sum p_i\omega_i \sim \sum p_i\Gamma_{\omega_i}$ and $\sum q_i\omega_i \sim \sum q_i\Gamma_{\omega_i}$; Independence of $\succeq^*$ implies that

$$\sum p_i\omega_i \sim^* \sum p_i\Gamma_{\omega_i} \text{ and} \sum q_i\omega_i \sim^* \sum q_i\Gamma_{\omega_i}.$$

Transitivity provides a chain of equivalences:

$$\begin{array}{rcl} \sum p_i\omega_i & \succeq & \sum q_i\omega_i \\ & \Updownarrow & \\ \sum p_i\Gamma_{\omega_i} & \succeq & \sum q_i\Gamma_{\omega_i} \\ & \Updownarrow & \\ \sum p_i\Gamma_{\omega_i} & \succeq_1 & \sum q_i\Gamma_{\omega_i} \\ & \Updownarrow & \\ \sum p_i\Gamma_{\omega_i} & \succeq^* & \sum q_i\Gamma_{\omega_i} \\ & \Updownarrow & \\ \sum p_i\omega_i & \succeq^* & \sum q_i\omega_i \end{array}$$

It follows that $\sum p_i\omega_i \succeq \sum q_i\omega_i$ if and only if $\sum p_i\omega_i \succeq^* \sum q_i\omega_i$. Because the lotteries $\sum p_i\omega_i, \sum q_i\omega_i$ were arbitrary, we conclude that $\succeq = \succeq^*$, as asserted. ■

References

Afriat, S. N. (1967). "The construction of a utility function from expenditure data." *International Economic Review*, 8(1): 67–77.

Afriat, S. N. (1972). "Efficiency estimation of production functions." *International Economic Review*, 13(3): 568–598.

Afriat, S. N. (2012). "Afriat's theorem and the index number problem." *Economic Journal*, 122(560): 295–304.

Ahn, David, et al. (2014). "Estimating ambiguity aversion in a portfolio choice experiment." *Quantitative Economics*, 5(2): 195–223.

Andreoni, James and John Miller (2002). "Giving according to GARP: An experimental test of the consistency of preferences for altruism." *Econometrica*, 70(2): 737–753.

Becker, Gary S. (1962). "Irrational behavior and economic theory." *Journal of Political Economy*, 70(1): 1–13.

Bland, J. Martin and Douglas G. Altman (1995). "Multiple significance tests: the Bonferroni method." *BMJ*, 310(6973): 170.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels (2005). "Fair procedures: Evidence from games involving lotteries." *The Economic Journal*, 115(506): 1054–1076.

Bolton, Gary E. and Axel Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 90(1): 166–193.

Brock, J Michelle, Andreas Lange, and Erkut Y Ozbay (2013). "Dictating the risk: Experimental evidence on giving in risky environments." *American Economic Review*, 103(1): 415–37.

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay (2016). "Dictating the risk: Experimental evidence on giving in risky environments: reply." *American Economic Review*, 106(3): 840–842.

Bronars, Stephen G. (1987). "The power of nonparametric tests of preference maximisation." *Econometrica*, 55(3): 693–698.

Camerer, Colin F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, NJ: Princeton University Press.

Cappelen, Alexander W., et al. (2013). "Needs versus entitlements: An international fairness experiment." *Journal of the European Economic Association*, 11(3): 574–598.

Cappelen, Alexander W., et al. (2023). "The development gap in economic rationality of future elites." *Games and Economic Behavior*, 142: 866–878.

Chambers, Christopher P. and Federico Echenique (2016). *Revealed Preference Theory.* Cambridge University Press.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests." *Quarterly Journal of Economics*, 117(3): 817–869.

Chew, S. H. (1989). "Axiomatic utility theories with the betweenness property." *Annals of Operations Research*, 19(1): 273–298.

Choi, Syngjoo, et al. (2007a). "Consistency and heterogeneity of individual behavior under uncertainty." *American Economic Review*, 97(5): 1921–1938.

Choi, Syngjoo, et al. (2007b). "Revealing preferences graphically: An old method gets a new tool kit." *American Economic Review*, 97(2): 153–158.

Choi, Syngjoo, et al. (2014). "Who is (more) rational?" *American Economic Review*, 104(6): 1518–1550.

de Clippel, Geoffroy and Kareen Rozen (2023). "Relaxed Optimization: How Close Is a Consumer to Satisfying First-Order Conditions?" *Review of Economics and Statistics*, 105(4): 883–898.

Cox, James C., Daniel Friedman, and Vjollca Sadiraj (2008). "Revealed altruism." *Econometrica*, 76(1): 31–69.

Debreu, Gerard (1954). "Representation of a preference ordering by a numerical function." In R.M. Thrall, C. H. Coombs, and R. L. Davis, eds., "Decision Processes," Wiley.

Dekel, Eddie (1986). "An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom." *Journal of Economic Theory*, 40(2): 304–318.

Diewert, W. Erwin (2012). "Afriat's theorem and some extensions to choice under uncertainty." *Economic Journal*, 122(560): 305–331.

Echenique, Federico, Taisuke Imai, and Kota Saito (2023). "Approximate Expected Utility Rationalization." *Journal of the European Economic Association*, 21(5): 1821–1864.

Echenique, Federico, Sangmok Lee, and Matthew Shum (2011). "The money pump as a measure of revealed preference violations." *Journal of Political Economy*, 119(6): 1201–1223.

Fehr, Ernst and Klaus M. Schmidt (1999). "A theory of fairness, competition and cooperation." *Quarterly Journal of Economics*, 114(3): 817–868.

Feldman, Paul and Kristian López Vargas (2024). *Disentangling Risk and Other-Regarding Preferences.* Technical report. Available at SSRN: https://dx.doi.org/10.2139/ssrn.4319590.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv (2015a). "How did distributional preferences change during the Great Recession?" *Journal of Public Economics*, 128: 84–95.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv (2017). "Distributional preferences and political behavior." *Journal of Public Economics*, 155: 1–10.

Fisman, Raymond, et al. (2015b). "The distributional preferences of an elite." *Science*, 349(6254): aab0096.

Fisman, Raymond, et al. (2023). "The distributional preferences of Americans, 2013–2016." *Experimental Economics*, 26(4): 727–748.

Fisman, Raymond J., Shachar Kariv, and Daniel Markovits (2007). "Individual preferences for giving." *American Economic Review*, 97(5): 1858–1876.

Fudenberg, Drew and David K. Levine (2012). "Fairness, risk preferences and independence: Impossibility theorems." *Journal of Economic Behavior and Organization*, 81(2): 606–612.

Good, Phillip (2005). *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Springer, 3 edition.

Grant, Simon, Atsushi Kajii, and Ben Polak (1992). "Many good choice axioms: When can many-good lotteries be treated as money lotteries?" *Journal of Economic Theory*, 56(2): 313–337.

Gul, Faruk (1991). "A theory of disappointment aversion." *Econometrica*, 59(3): 667–686.

Hadar, Josef and William R. Russel (1969). "Rules for ordering uncertain prospects." *American Economic Review*, 59(1): 25–34.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill (2018). "Parametric recoverability of preferences." *Journal of Political Economy*, 126(4): 1558–1593.

Holt, Charles A. and Susan K. Laury (2002). "Risk aversion and incentive effects." *American Economic Review*, 92(5): 1644–1655.

Houtman, Martijn and J. A. H. Maks (1985). "Determining all maximal data subsets consistent with revealed preference." *Kwantitatieve Methoden*, 6(19): 89–104.

Karni, Edi and Zvi Safra (2002). "Individual sense of justice: A utility representation." *Econometrica*, 70(1): 263–284.

Karni, Edi and Zvi Safra (2008). "Moral sentiments and social choice." *Social Choice and Welfare*, 30(3): 427–446.

Krawczyk, Michal and Fabrice Le Lec (2008). "Social decisions under risk. Evidence from the probabilistic dictator game." Manuscript.

Krawczyk, Michal and Fabrice Le Lec (2016). "Dictating the risk: Experimental evidence on giving in risky environments: Comment." *American Economic Review*, 106(3): 836–839.

Li, Jing, William H. Dow, and Shachar Kariv (2017). "Social preferences of future physicians." *Proceedings of the National Academy of Sciences*, 114(48): E10291–E10300.

Li, Jing, et al. (2022). “Experimental evidence of physician social preferences.” *Proceedings of the National Academy of Sciences*, 119(28): e2112726119.

López-Vargas, Kristian (2015). “Risk attitudes and fairness: Theory and experiment.” In “Essays in Behavioral and Experimental Economics (PhD dissertation),” chapter 1, pp. 1–37. University of Maryland.

Mas-Colell, A., M.D. Whinston, and J.R. Green (1995). *Microeconomic Theory*. Oxford University Press.

Nelson, Michael (2018). *The Presidency and the Political System*. Thousand Oaks, CA: CQ Press, 11th edition.

von Neumann, John and Oskar Morgenstern (1947). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 2nd edition.

Nishimura, Hiroki, Efe A. Ok, and John K.-H. Quah (2017). “A comprehensive approach to revealed preference theory.” *American Economic Review*, 107(4): 1239–1263.

Polisson, Matthew, John K. H. Quah, and Ludovic Renou (2020). “Revealed preferences over risk and uncertainty.” *American Economic Review*, 110(6): 1782–1820.

Quiggin, John (1982). “A theory of anticipated utility.” *Journal of Economic Behavior & Organization*, 3(4): 323–343.

Quiggin, John (1990). “Stochastic dominance in regret theory.” *Review of Economic Studies*, 57(3): 503–511.

Quiggin, John (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model*. Kluwer Academic Publishers.

Quirk, James P. and Rubin Saposnik (1962). “Admissibility and measurable utility.” *Review of Economic Studies*, 29(2): 140–146.

Saito, Kota (2013). “Social preferences under risk: Equality of opportunity versus equality of outcome.” *American Economic Review*, 103(7): 3084–3101.

Savage, Leonard J. (1954). *The Foundations of Statistics*. New York: Wiley.

Schildberg-Hörisch, Hannah (2010). “Is the veil of ignorance only a concept about risk? An experiment.” *Journal of Public Economics*, 94(11-12): 1062–1066.

Trautmann, Stefan T. (2009). “A tractable model of process fairness under risk.” *Journal of Economic Psychology*, 30(5): 803–813.

Tversky, Amos and Daniel Kahneman (1992). “Advances in prospect theory: Cumulative representation of uncertainty.” *Journal of Risk and Uncertainty*, 5(5): 297–323.

Varian, Hal R. (1982). "The nonparametric approach to demand analysis." *Econometrica*, 107(4): 945–972.

Varian, Hal R. (1983). "Non-parametric tests of consumer behavior." *Review of Economic Studies*, 50(1): 99–110.

Varian, Hal R. (1990). "Goodness-of-fit in optimizing models." *Journal of Econometrics*, 46(1-2): 125–140.

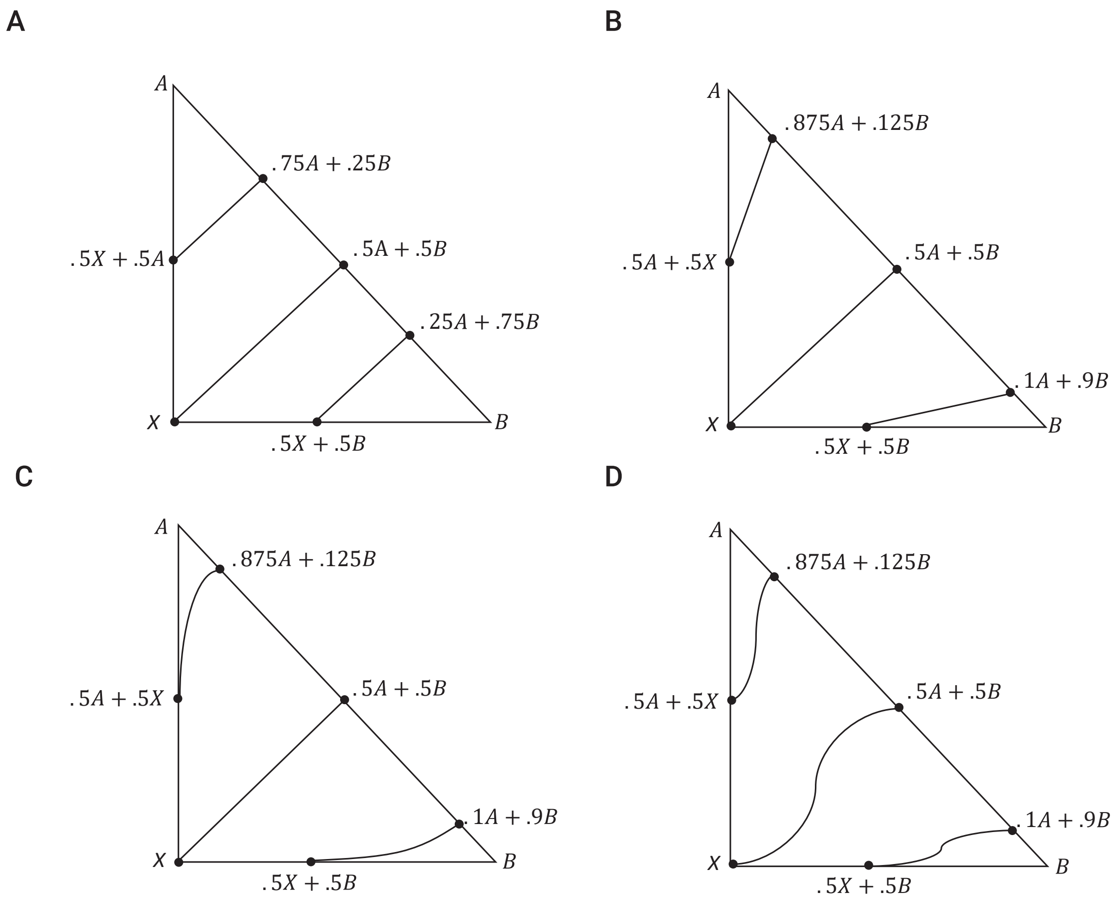
Varian, Hal R. (1991). *Goodness of Fit for Revealed Preference Tests*. Technical Report 13, Michigan – Center for Research on Economic & Social Theory.

Varian, Hal R. (2012). "Revealed preference and its applications." *Economic Journal*, 122(560): 332–338.

Vermeulen, Frederic (2012). "Foundations of revealed preference: Introduction." *Economic Journal*, 122(560): 287–294.

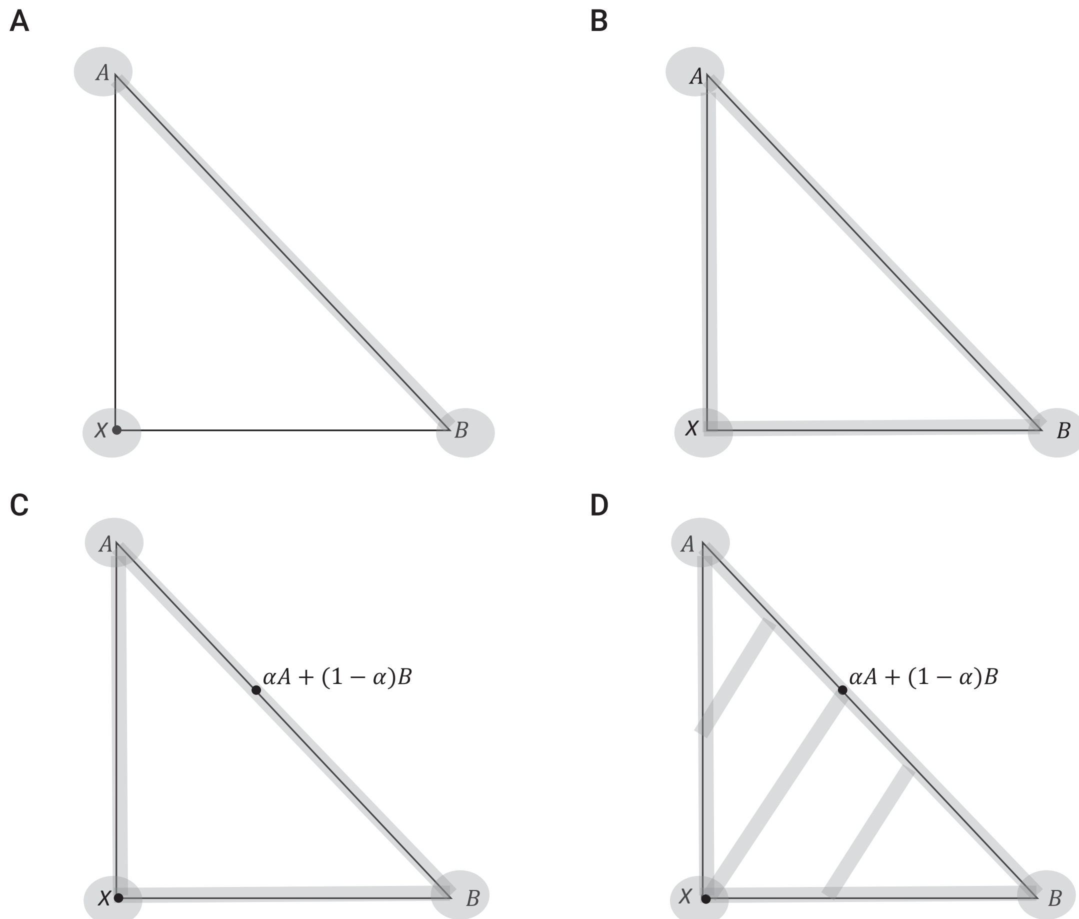
Wakker, Peter (1993). "Savage's axioms usually imply violation of strict stochastic dominance." *Review of Economic Studies*, 60(2): 487–493.

Figure 1: An illustration of the axioms in the Marschak–Machina triangle



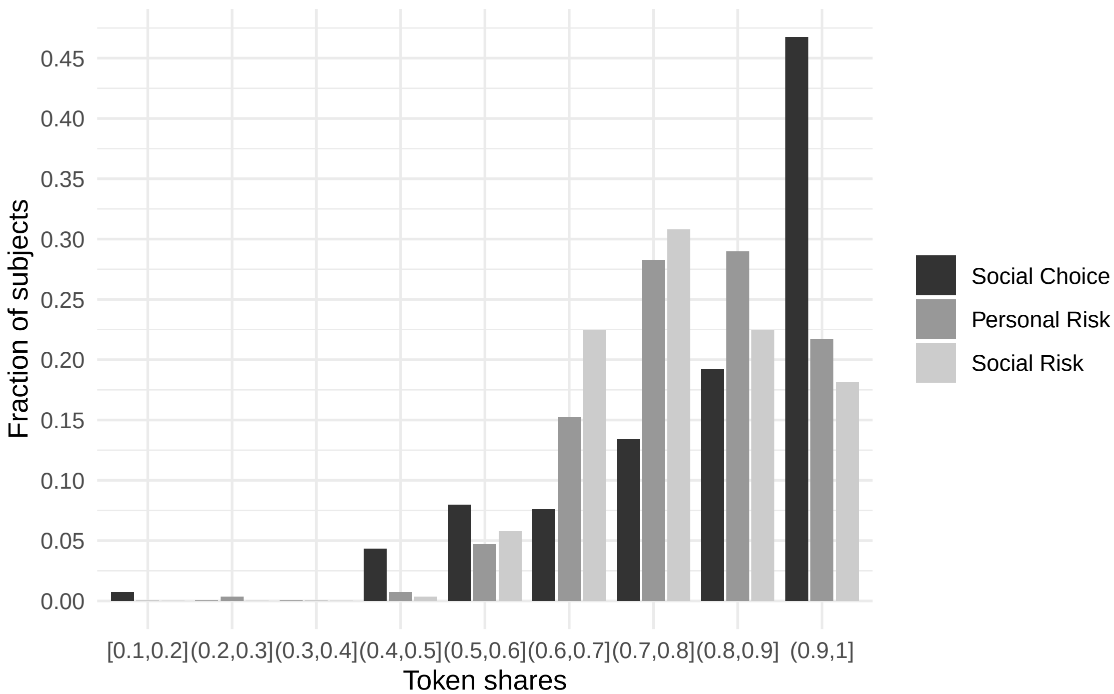
Note: (A) The preference relation $\succeq$ admits an Expected Utility representation. The indifference curves in the triangle are parallel straight lines. (B) The preference relation $\succeq$ only admits a Weighted Expected Utility representation. The indifference curves are straight lines but they need not be parallel. (C) The indifference curves are not straight lines as in Rank Dependent Utility and Prospect Theory. (D) The preference relation $\succeq$ obeys only State Monotonicity and—in the presence of the other axioms—respects FOSD. The indifference curves are "upward sloping" but can otherwise be quite arbitrary.

Figure 2: The difference between observing $\succeq_0$ and $\succeq_1$ in the Marschak–Machina triangle



Note: The areas in the triangle shaded gray represent the ordering of lotteries that can be inferred. (A) If we observe $\succeq_0$ we observe the ordering $A \succ_0 X \succ_0 B$ and the ordering of lotteries between A, B—but no others. (B) If we observe $\succeq_0$ and we assume State Monotonicity we can *infer* the ordering of lotteries between A, X and lotteries between X, B. (C) Continuity assures us that X is indifferent to *some* lottery $aA + (1-a)B$, but if we observe only $\succeq_0$ then we do not observe *which lottery*, but if we observe $\succeq_1$ then we *do* observe which lottery. (D) If we observe $\succeq_1$ and assume that $\succeq$ obeys Independence then observing *which lottery* completely determines $\succeq$.

Figure 3: The distributions of the average token shares in the three domains



Note: The distributions (across individuals) of the average token shares across all choices. Social Choice domain: the average fraction of tokens kept by *self*. Personal Risk and Social Risk domains: the average fraction of tokens allocated to the *cheaper* account (that is, to x when $p_x < p_y$ and to y otherwise).

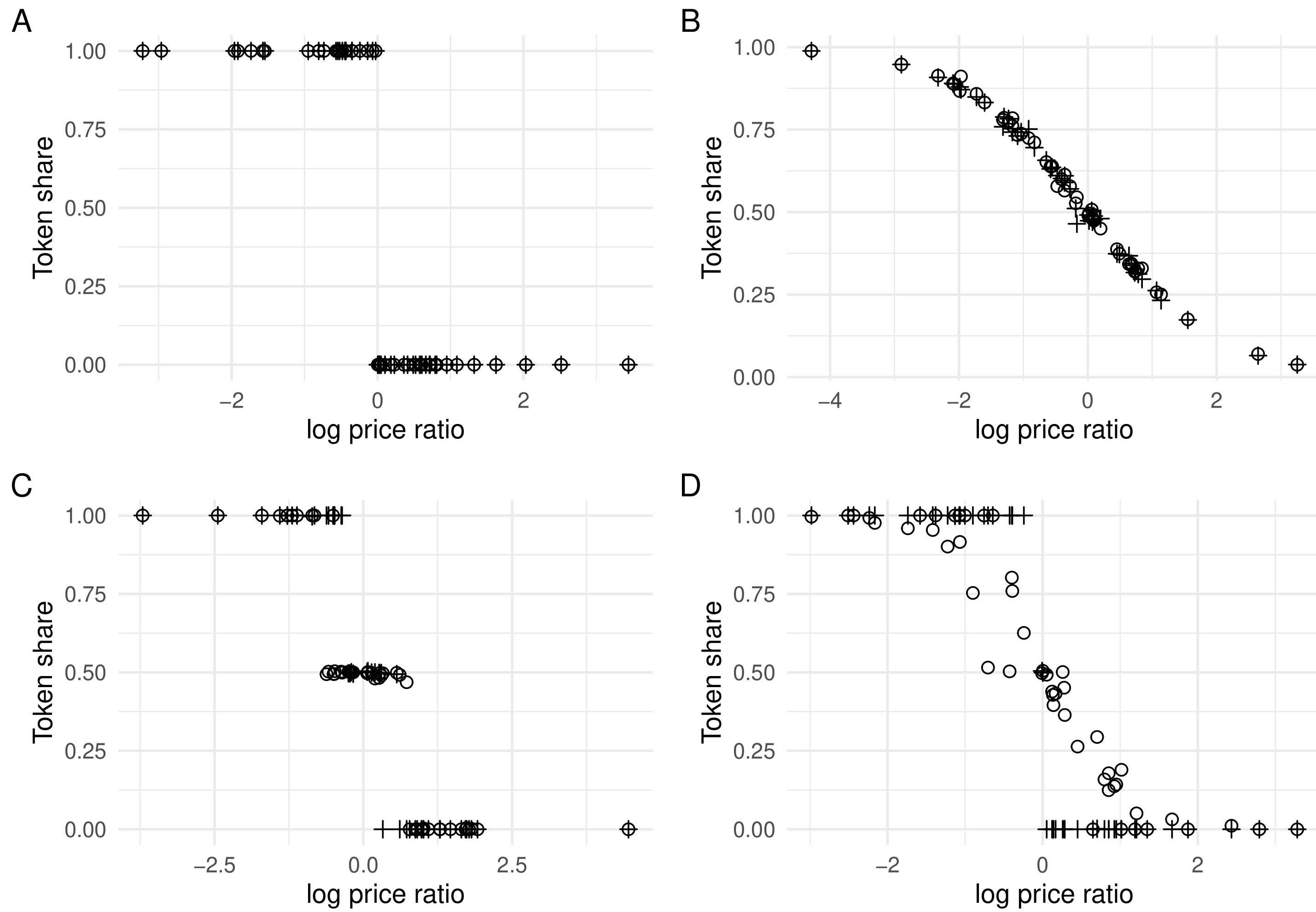
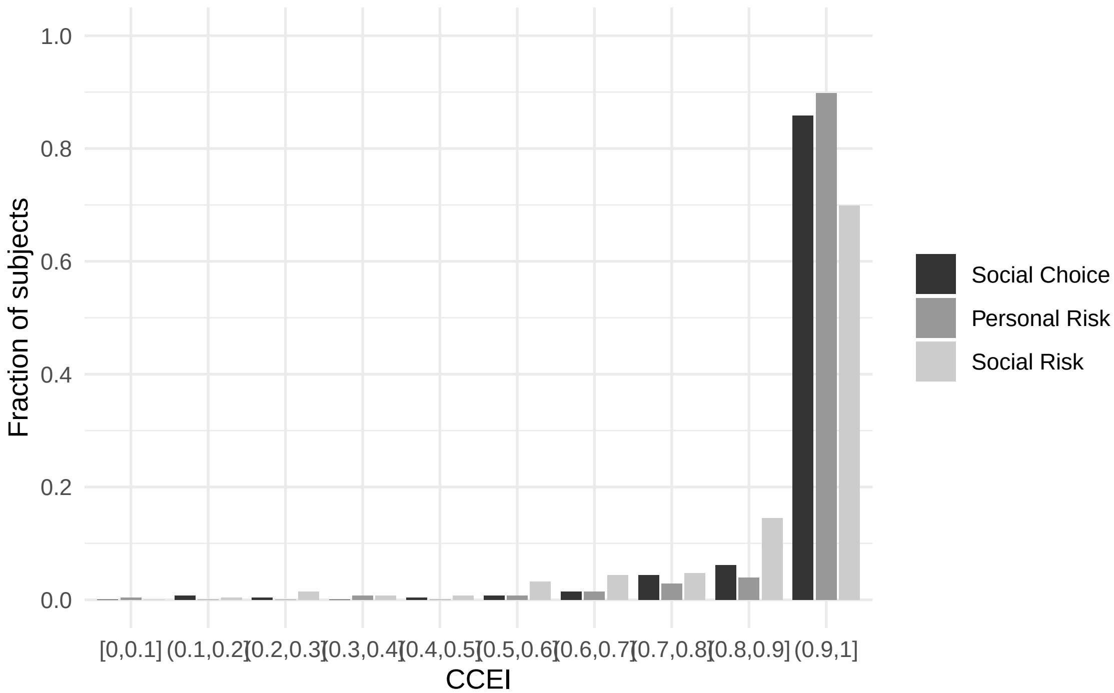


Figure 4: The choices of four illustrative subjects

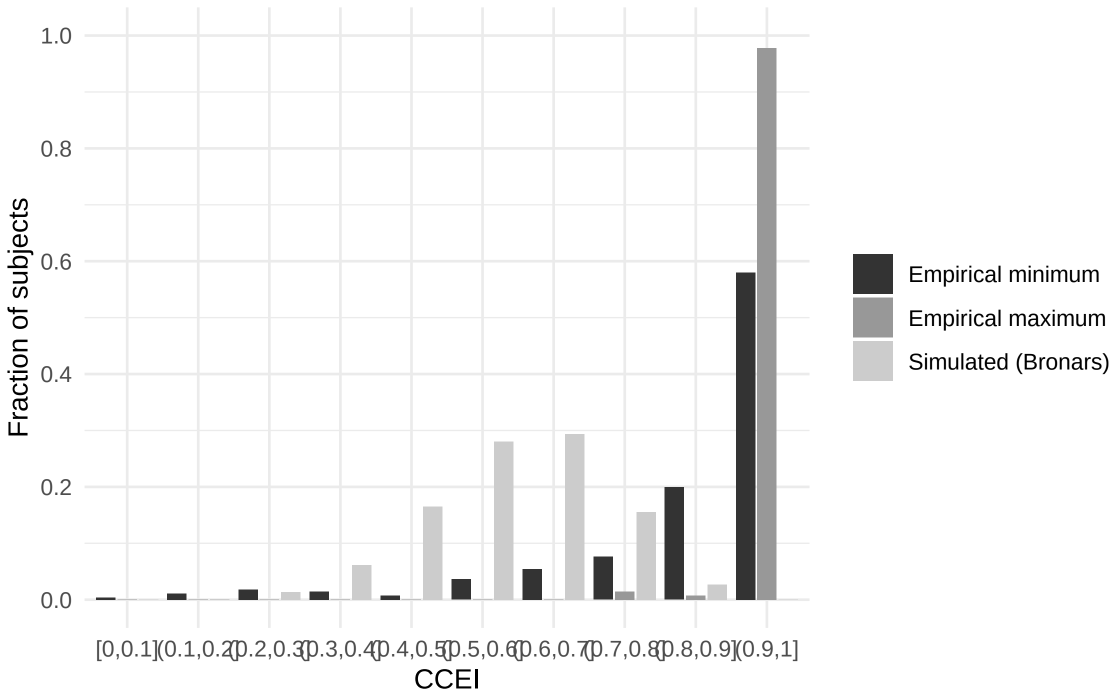
Note: The choices in the Personal Risk and Social Risk domains of four illustrative subjects who we classified as selfish on the basis of their choices in the Social Choice domain. Each panel shows the subject's relative demand $y/(x+y)$ (vertical axis) at a given log-price ratio $\ln(p_y/p_x)$ (horizontal axis) in the Personal Risk (+) and Social Risk (○) domains. These are special cases, where the regularities in the data are very clear but they reveal striking regularities *within* and marked heterogeneity *across* subjects that is characteristic of all our data.

Figure 5: The distributions of CCEI scores in the three domains



Note: The CCEI is bounded between 0 and 1. The closer the CCEI is to 1, the smaller the perturbation of the budget constraints required to remove all violations and thus the closer the data are to satisfying GARP.

Figure 6: The power of the GARP test



Note: The distributions of the minimum and maximum CCEI scores across the three domains for the actual subjects and the distribution of the CCEI scores generated by 100,000 simulated subjects who randomize uniformly on each budget line (Bronars, 1987). Each of the 100,000 random subjects makes 50 choices from randomly generated budget lines in the same way as the human subjects do.

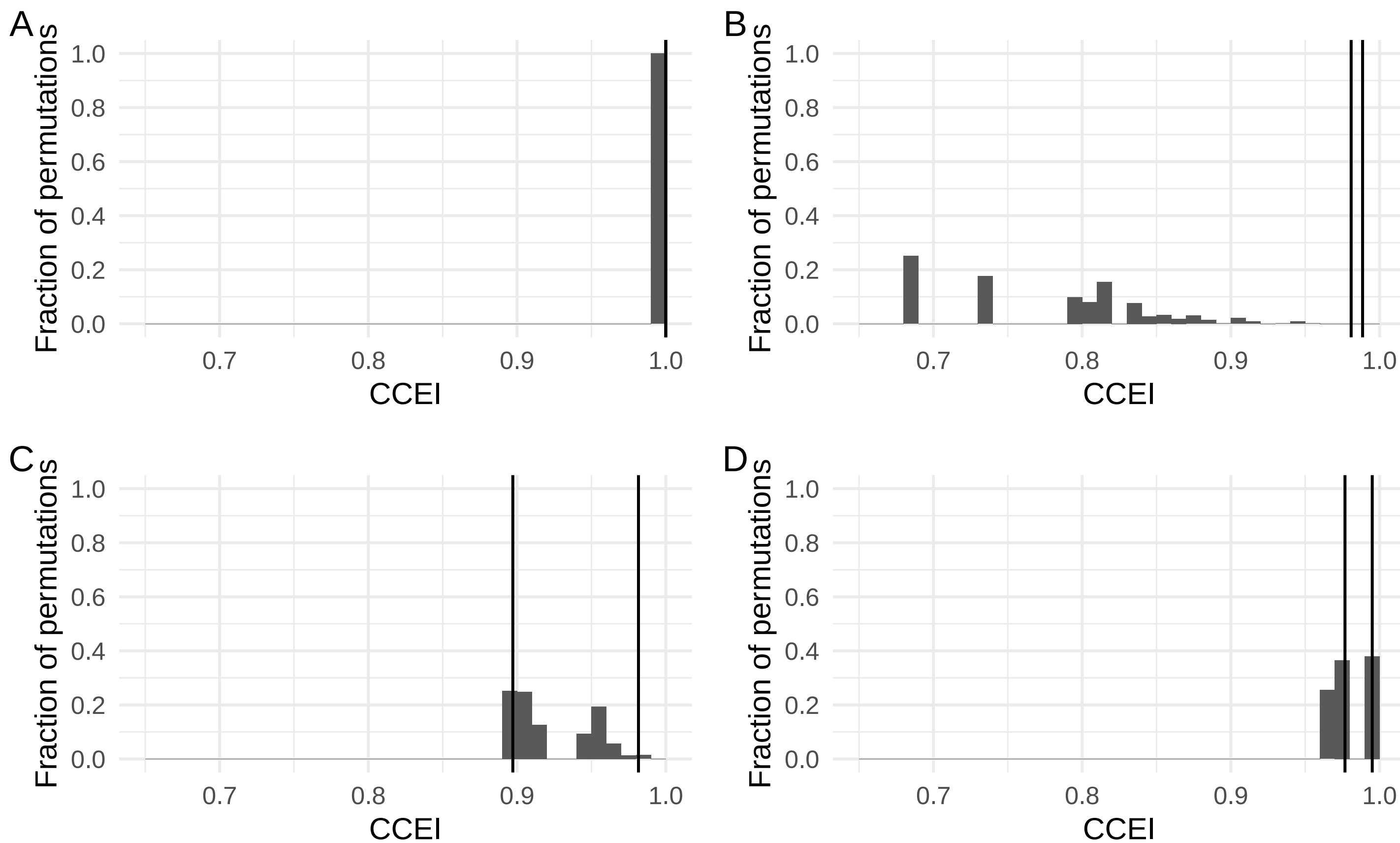


Figure 7: The permutation test for four illustrative subjects

Note: The permutation test for four illustrative subjects who we classified as selfish on the basis of their choices in the Social Choice domain. Each panel presents a histogram of the permuted CCEI scores and the two actual CCEI scores $e^- = \min\{e_{\mathcal{PR}}, e_{\mathcal{SR}}\}$ and $e^+ = \max\{e_{\mathcal{PR}}, e_{\mathcal{SR}}\}$ from the Personal Risk and Social Risk domains for one of these selfish subjects.

Table 1: Population-level summaries of the individual-level test results

A. Subjects classified as selfish (top entry) / non-selfish (bottom entry)

		1% significance level				5% significance level				10% significance level			
Classification	n	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
I	68	**0.059**	1.000	1.000	1.000	**0.206**	1.000	1.000	1.000	**0.235**	1.000	1.000	1.000
	208	0.154	0.750	0.731	0.736	0.231	0.808	0.769	0.837	0.298	0.827	0.798	0.861
II	86	**0.093**	0.977	1.000	1.000	**0.221**	0.988	1.000	1.000	**0.244**	0.988	1.000	1.000
	190	0.147	0.737	0.705	0.711	0.226	0.795	0.747	0.821	0.300	0.816	0.779	0.847
III	103	**0.078**	0.981	1.000	1.000	**0.204**	0.990	1.000	1.000	**0.223**	0.990	1.000	1.000
	173	0.162	0.711	0.676	0.682	0.237	0.775	0.723	0.803	0.318	0.798	0.757	0.832
IV	129	**0.093**	0.977	1.000	1.000	**0.225**	0.992	1.000	1.000	**0.256**	0.992	1.000	1.000
	147	0.163	0.667	0.619	0.626	0.224	0.735	0.673	0.769	0.306	0.762	0.714	0.803

B. Subjects classified as impartial (top entry) / non-impartial (bottom entry)

		1% significance level				5% significance level				10% significance level			
Classification	n	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
V	19	0.158	**0.158**	0.000	0.000	0.263	**0.158**	0.053	0.211	0.263	**0.158**	0.053	0.263
	257	0.128	0.860	0.856	0.860	0.222	0.907	0.883	0.926	0.284	0.922	0.907	0.942
VI	30	0.200	**0.267**	0.133	0.167	0.233	**0.267**	0.167	0.333	0.233	**0.267**	0.167	0.333
	246	0.122	0.878	0.878	0.878	0.224	0.927	0.907	0.943	0.289	0.943	0.931	0.963
VII	33	0.212	**0.333**	0.152	0.212	0.273	**0.333**	0.182	0.394	0.273	**0.333**	0.182	0.394
	243	0.119	0.877	0.885	0.881	0.218	0.926	0.914	0.942	0.284	0.942	0.938	0.963
VIII	37	0.216	**0.351**	0.162	0.270	0.297	**0.351**	0.189	0.432	0.297	**0.351**	0.216	0.459
	239	0.117	0.883	0.895	0.883	0.213	0.933	0.925	0.946	0.280	0.950	0.946	0.962

Note: Panel A (resp. B) tabulates the number of subjects we classify as selfish/non-selfish (resp. impartial/non-impartial) based on their choices in the SOCIAL CHOICE domain. The columns of each panel report the percent of subjects for whom we can reject the null that preferences coincide: (1) $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$, (2) $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$, (3) $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SC}}$, (4) $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$. The tests of the theoretical predictions reported in column (1) in panel A and column (2) in panel B are presented in bold type. In panel A (resp. B), the top entry at each cell is for the selfish (resp. impartial) subjects and the bottom entry is for the non-selfish (resp. non-impartial) subjects. In panel A, we classify a subject to be selfish if in the SOCIAL CHOICE domain (I) $y/(x+y) > 0.99$, (II) $y/(x+y) > 0.975$, (III) $y/(x+y) > 0.95$, and (IV) $y/(x+y) > 0.9$. In panel B, we classify a subject as impartial if in the SOCIAL CHOICE domain (V) $0.45 < y/(x+y) < 0.55$, and (VI)-(VIII) using the nonparametric test that preferences $\succeq_{\mathcal{SC}}$ in the SOCIAL CHOICE domain are symmetric using 10%, 5%, and 1% significance levels, respectively.

Table 2: The power of the test of Propositions 2 and 3

A. Testing Proposition 2, rejection rates

Significance level	Probability of random choice replacement (μ) 0	5%	10%	15%	20%
1%	0.078	0.175	0.379	0.553	0.718
5%	0.204	0.233	0.592	0.786	0.845
10%	0.223	0.408	0.748	0.874	0.913

B. Testing Proposition 3, rejection rates

Significance level	Probability of random choice replacement (μ) 0	5%	10%	15%	20%
1%	0.158	0.158	0.263	0.526	0.632
5%	0.158	0.158	0.526	0.632	0.737
10%	0.158	0.263	0.579	0.684	0.737

Note: Panel A (resp. B) tabulates the percent of selfish (resp. impartial) subjects for whom we reject the null that $\succeq_{\mathcal{PR}}=\succeq_{\mathcal{SR}}$ (resp. $\succeq_{\mathcal{SC}}=\succeq_{\mathcal{SR}}$) at the 1%, 5% and 10% levels when we replace the actual choices with random choices with different probabilities $\mu = 0, 0.05.0.1, 0.15, 0.2$. In panel A (resp. B), the subjects are those who we classify as selfish (resp. impartial) according to classification III in Table 1A (resp. classification VII in Table 1B).