Essays on Decision Theory

David Dillenberger

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

ECONOMICS

June 2008

Abstract

The first chapter, "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior", is motivated by experimental evidence that suggests that individuals are more risk averse when they perceive risk gradually. We address these findings by studying a decision maker (DM) who has recursive preferences over compound lotteries and who cares about the way uncertainty is resolved over time. DM has preferences for one-shot resolution of uncertainty if he always prefers any compound lottery to be resolved in a single stage. We establish an equivalence between dynamic preferences for one-shot resolution of uncertainty and static preferences that are identified with the behavior observed in Allais-type experiments. The implications of this equivalence on preferences over information systems are examined. We define the gradual resolution premium and demonstrate its magnifying effect when combined with the usual risk premium. In an intertemporal context, preferences for one-shot resolution of uncertainty capture narrow framing.

In the second chapter, "Ashamed to be Selfish" (jointly written with Philipp Sadowski), we study a two-stage choice problem, where alternatives are allocations between the decision maker (DM) and a passive recipient. The recipient observes choice behavior in stage two, while stage one choice is unobserved. Choosing selfishly in stage two, in the face of a fairer available alternative, may inflict shame on DM. DM has preferences over sets of alternatives that represent period two choices. We axiomatize a representation that identifies DM's selfish ranking, her norm of fairness and shame. Altruism is the most prominent motive that can explain non-selfish choice. We identify a condition under which shame to be selfish can mimic altruism, when only stage-two choice is observed by the experimenter. An additional condition implies that the norm of fairness can be characterized as the Nash solution of a bargaining game induced by the second-stage choice problem. The representation is generalized to allow for finitely many recipients and applied to explain a social decision maker's incentive for obfuscation.

Acknowledgements

Contents

Chapter 1

Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior

1.1 Introduction

Experimental evidence suggests that individuals are more risk averse when they perceive risk that is gradually resolved over time. In an experiment with college students, Gneezy and Potters [1997] found that subjects invest less in risky assets if they evaluate financial outcomes more frequently. Haigh and List [2005] replicated the study of Gneezy and Potters with professional traders and found an even stronger effect. These two studies allow for flexibility in adjusting investment according to how often the subjects evaluate the returns. Bellemare, Krause, Kröger, and Zhang [2005] found that even when all subjects have the same investment flexibility, variations in the frequency of information feedback alone affects investment behavior systematically. All their subjects had to commit in advance to a fixed equal amount of investment for three subsequent periods. Group A was told that they would get periodic statements (i.e. would be informed about the outcome of the gamble after every draw), whereas group B knew that they would hear only the final yields of their investment. The average investment in group A was significantly lower than in group

B. The authors conclude that "information feedback should be the variable of interest for researchers and actors in financial markets alike." Such interdependence between the way individuals observe the resolution of uncertainty and the amount of risk they are willing to take is not compatible with the standard model of decision making under risk, which is a theory of choice among probability distributions over final outcomes.[1]

In this paper, we make the assumption that the value of a lottery depends not only on its uncertainty, but also on the way this uncertainty is resolved over time. Using this assumption, we provide a choice theoretic framework that can address the experimental evidence above, while pinpointing the required deviations from the standard model. We exploit the structure of the model to identify the link between the temporal aspect of risk aversion, a static attitude towards risk, and intrinsic preferences for information.

In order to facilitate exposition, we mainly consider a decision maker (DM) whose preferences are defined over the set of two-stage lotteries, namely lotteries over lotteries over outcomes. Following Segal [1990], we replace the *reduction of compound lotteries axiom* (an axiom that imposes indifference between compound lotteries and their reduced single-stage counterparts) with the following two assumptions: *time neutrality*, which says that DM does not care about the time in which the uncertainty is resolved as long as resolution happens in a single stage, and *recursivity*, which says that the ranking of second-stage lotteries is unaffected by the first stage. Under these assumptions, any two-stage lottery is *subjectively* transformed into a simpler, one-stage lottery. In particular, there is a single preference relation defined over the set of one-stage lotteries that fully determines preferences over the richer domain of two-stage lotteries.

As a first step to link behavior in both domains, we introduce and formally define the following two properties: the first is dynamic while the second is static.

- *Preferences for one-shot resolution of Uncertainty* (PORU). DM has PORU if he always prefers any two-stage lottery to be resolved in a single stage. PORU implies an aversion to receiving partial information. This notion formalizes an idea first raised

---

[1] All lotteries discussed in this paper are objective, that is, the probabilities are known. Knight [1921] proposed distinguishing between "risk" and "uncertainty" according to whether the probabilities are given to us objectively or not. Despite this distinction, we will interchangeably use both notions to express the same thing.

by Palacios-Huerta [1999] (to be further discussed in the literature review section). Such preferences capture the idea that "the frequency at which the outcomes of a random process are evaluated" is a relevant economic variable.

- *Negative certainty independence* (NCI). NCI states that if DM prefers lottery p to the (degenerate) lottery that yields the prize x for certain, then this ranking is not reversed when we mix both options with any common third lottery q. This axiom is similar, but it is less demanding than Kahneman and Tversky's [1979] "Certainty effect" hypothesis, since it does not imply that people weight probabilities non-linearly. NCI imposes weak restrictions on preferences, just enough to explain commonly observed behavior in Allais-type experiments.

Theorem 1, our main result, establishes a tight connection between the two behavioral properties just described; NCI is a sufficient condition to PORU, and within the class of betweenness-satisfying preferences (Dekel [1986]), it is also necessary.

On the one hand, numerous replications of the Allais paradox in the last fifty years prove NCI to be one of the most prominently observed preference patterns. On the other hand, empirical and experimental studies involving dynamic choices and experimental studies on preference for uncertainty resolution are still rather rare. The disproportional amount of evidence in favor of each property strengthens the importance of theorem 1, since it provides new theoretical predictions for dynamic behavior, based on robust (static) empirical evidence.

Within the betweenness class, axiom NCI has its own static implications. First, it is equivalent to the following geometrical condition that is imposed on the map of indifference curves in every unit probability triangle (a diagram that represents the set of all lotteries over three fixed prizes):

- *Steepest middle slope* property: for every triple $x_3 > x_2 > x_1$, the indifference curve that passes through the origin (the lottery that yields x_2 for certain) is the steepest.

Since this geometrical condition is relatively easy to verify, it proves to be an applicable tool. Second, in theorem 2 we show that NCI is incompatible with the assumption

3

that preferences are at least twice differentiable. When coupled with such a smoothness assumption, NCI turns out to be equivalent to the vNM-independence axiom.

In an extended model, we allow DM to take (just before the second-stage lottery is acted out, but after the realization of the first-stage lottery) intermediate actions that might affect his ultimate payoff. The primitive in such a model is a preference relation over *information systems*, which is induced from preferences over compound lotteries. An immediate consequence of Blackwell's [1953] seminal result is that in the standard expected utility class, DM always prefers to have perfect information before making the decision, which allows him to choose the optimal action corresponding to the resulting state. Safra and Sulganik [1995] left open the question of whether there are other preference relations for which, when applied recursively, a perfect information system is always the most valuable. We show that this property is equivalent to PORU. As a corollary, axiom NCI fully characterizes, within the betweenness class, such preferences for information.

The idea that individuals prefer one-shot resolution of uncertainty can be quantified. The *gradual resolution premium* of any compound lottery is the amount that DM would pay to replace that lottery by its single-stage counterpart. Similarly to the regular risk premium (the amount that DM would pay to replace one-stage lottery by its expected value), the gradual resolution premium is measured in monetary terms. The signs of these two variables need not agree, that is, positive risk premium does not imply and is not implied by positive gradual resolution premium. In the case where DM is both risk averse and displays PORU, however, these two forces magnify each other. We use this observation to explain why people often purchase dynamic insurance contracts, such as periodic insurance for electrical appliances and cellular phones, at much more than the actuarially fair rates.

The gradual resolution premium can be very significant, in the sense that if the resolution process is "long" enough, it might imply an extreme degree of risk aversion. To illustrate this, we first extend our results to preferences over arbitrary n-stage lotteries. We interpret the parameter n as the "resolution sensitivity" of an individual. It describes the frequency with which an individual updates information in a fixed time interval. Qualitatively, the results remain intact; DM who has preferences for one-shot resolution of uncertainty prefers to replace each (compound) sub-lottery with its single-stage counterpart. We then look

at preferences of the disappointment aversion class (Gul [1991]). Such preferences satisfy NCI, and therefore, in a dynamic context, PORU. We show that for any one-stage lottery, there exists a multi-stage lottery (with the same probability distribution over the terminal prizes) whose value approximately equals the value of getting the worst prize for sure. While referring to the problem of repeated investment, Gollier [2001] states that "the central theoretical question of the link between the structure of the utility function and the horizon-riskiness relationship remained unsolved." The result above shows that preferences that display PORU may lead to excessively conservative investment strategies.

1.1.1 Related literature

Palacios-Huerta [1999] was the first to raise the idea that the form of the timing of resolution of uncertainty might be an important economic variable. By working out an example, he demonstrates that DM with Gul's [1991] disappointment aversion preferences will be averse to the sequential resolution of uncertainty, or, in the language of this paper, will be displaying PORU. He also discusses a lot of potential applications. Ang, Bekaert and Liu [2005] use recursive disappointment aversion preferences to study a dynamic portfolio choice designed to maximize final wealth. The general theory we suggest provides an insightful way to understand exactly which attribute of Gul's preferences accounts for the resulting behavior. It also makes a clear distinction between two notions of disappointment: The common static notion of disappointment, as it appears in the literature, and the dynamic version implied by PORU (see section 3).

Loss aversion with narrow framing (also known as "myopic loss aversion") is a combination of two motives: loss aversion (Kahneman and Tversky [1979]), that is, people's tendency to be more sensitive to losses than to gains, and narrow framing (Kahneman and Tversky [1984]), that is, a dynamic aggregation rule that argues that when making a series of choices, individuals "bracket" them by making each choice in isolation. Benartzi and Thaler [1995] were the first to use this approach to suggest explanations for several economic "anomalies", such as the equity premium puzzle (Mehra and Prescott [1985]). Barberis and Huang [2005] and Barberis, Huang and Thaler [2006] generalize Benartzi and Thaler's work by assuming that DM derives utility directly from the outcome of a gamble

over and above its contribution to total wealth.

The model presented here can be used to address similar phenomena. The combination of the folding-back procedure and a specific form of non-smooth atemporal preferences implies that individuals behave *as if* they intertemporally perform narrow framing. The gradual resolution premium quantifies this effect. The two approaches are conceptually different: Loss aversion with narrow framing brings to the forefront the idea that individuals evaluate any new gamble separately from its cumulative contribution to total wealth, while we maintain the assumption that terminal wealth matters, and identify narrow framing as a subjective temporal effect. In addition, we set aside the question of why individuals are sensitive to the way uncertainty is resolved (i.e. why they narrow frame), and construct a model that reveals the (context independent) behavioral implications of such considerations.

Rabin [2000] and Safra and Segal [2006] use calibration results to criticize a broad class of models of decision making under risk. They point out that modest risk aversion over small stakes gambles necessarily implies absurd levels of risk aversion over large stakes gambles. Our model resists these critiques. If most uncertainty resolves gradually, then it cannot be compounded into a single lottery. Our model implies first order risk aversion over each realized gamble, and therefore neither Rabin's nor Safra and Segal's arguments apply.

In this paper, we study time's effect on preferences by distinguishing between "one-shot" and "gradual" resolution of uncertainty. A different, but complementary, approach is to study intrinsic preferences for "early" or "late" resolution of uncertainty. This research agenda was initiated by Kreps and Porteus [1978], and later extended by Epstein and Zin [1989] and Epstein and Chew [1989] among others. Grant, Kajii and Polak [1998, 2000] connect preferences for the timing of resolution of uncertainty to intrinsic preferences for information. We believe that both aspects of intrinsic time preferences play a role in most real life situations. For example, an anxious student might prefer to know as soon as possible his final grade in an exam, but still prefers to wait (impatiently) rather than to get the grade of each question separately. The motivation to impose time neutrality is to demonstrate the role of the "one-shot" versus "gradual" effect, which has been neglected in the literature to date.

The remainder of the paper is organized as follows: we start section 2 by establishing our basic framework, after which we introduce the main behavioral properties of the paper and state our main characterization result. In section 3, we elaborate on the static implications of our model and provide examples. Section 4 first extends our results to preferences over compound lotteries with an arbitrarily finite number of stages. We then define the gradual resolution premium and illustrate its effect. In section 5, we relate our approach to the notion of loss aversion with narrow framing. Section 6 comments on the implications of our model on preferences over information systems. Section 7 is devoted to an application of our model to the area of investment under uncertainty. We present our concluding remarks in section 8. Most proofs are relegated to the appendix.

1.2   The model

1.2.1   Groundwork

Consider an interval $[w, b] = X \subset R$ of monetary prizes. Let $\mathcal{L}(X)$, or simply $\mathcal{L}^1$, be the set of all simple lotteries (lotteries with a finite number of outcomes) over X. Typical elements of $\mathcal{L}^1$ are denoted by p, q and r. If p, $q \in \mathcal{L}^1$ and $\alpha \in (0, 1)$, then the mixture $\alpha p + (1 - \alpha) q \in \mathcal{L}^1$ is the lottery that yields each x with probability $\alpha p_x + (1 - \alpha) q_x$. We denote by δ_x the lottery that gives the prize x with certainty.

Denote by $\mathcal{L}(\mathcal{L}(X))$, or simply by $\mathcal{L}^2$, the set of all simple lotteries over $\mathcal{L}^1$. A typical element of $\mathcal{L}^2$ is $Q = \langle \alpha_1, q^1; ...; \alpha_l, q^l \rangle$ with $\alpha_j \geqslant 0$, $\sum_{j=1}^{l} \alpha_j = 1$ and $q^j \in \mathcal{L}^1$, $j = 1, 2, ..., l$. We call elements of $\mathcal{L}^2$ two-stage lotteries. We think of each $Q \in \mathcal{L}^2$ as a dynamic two-stage process where, in the first stage, a lottery q^j is chosen with probability α_j, and, in the second stage, a prize is obtained according to q^j.

Two special subsets of $\mathcal{L}^2$ are $\Gamma = \left\{ \langle 1, q \rangle \mid q \in \mathcal{L}^1 \right\}$ and $\Delta = \left\{ \langle \alpha_{x_i}, \delta_{x_i} \rangle_{i=1}^{m}, \ x_i \in X \right\}$. All lotteries in Γ and Δ are fully resolved in a single stage; in every member of Γ, no uncertainty is resolved in the first stage, whereas the uncertainty of every lottery in Δ is fully resolved in the first stage. Note that both Γ and Δ are isomorphic to $\mathcal{L}^1$.

Let $\mathcal{V}$ denote the set of all continuous and strictly monotone preference relations over (sets isomorphic to) $\mathcal{L}^1$, with a generic element $\succeq_1$. Each $\succeq_1 \in \mathcal{V}$ is represented by some

continuous function $V : \mathcal{L}^1 \rightarrow \mathbb{R}.^2$

Given V, the certainty equivalent of lottery p is a prize $c_V(p)$ satisfying $p \sim_1 \delta_{c_V(p)}$, where $\sim_1$ is the indifference relation induced from $\succeq_1$. By continuity and monotonicity, $c_V : \mathcal{L}^1 \rightarrow X$ is well defined.

Let $\succeq$ be a preference relation over $\mathcal{L}^2$. Let $\succeq_\Gamma$ and $\succeq_\Delta$ be the restriction of $\succeq$ to Γ and Δ respectively. We assume throughout the paper that both $\succeq_\Gamma$ and $\succeq_\Delta$ are in $\mathcal{V}$. On $\succeq$ we impose the following axioms:

A1 (*time neutrality*): $\forall q \in \mathcal{L}^1$, $\langle 1, q \rangle \sim \langle q_{x_i}, \delta_{x_i} \rangle_{i=1}^m$

A2 (*recursivity*):

$$\left\langle \alpha_1, q^1; ..; \alpha_i, q^i; ..; \alpha_l, q^l \right\rangle \succeq \left\langle \alpha_1, q^1; ..; \alpha_i, \widetilde{q}^i; ..; \alpha_l, q^l \right\rangle \Longleftrightarrow \langle 1, q^i \rangle \succeq \langle 1, \widetilde{q}^i \rangle$$

By postulating **A1**, we assume that DM does not care about the time in which the uncertainty is resolved as long as it happens in a single stage. **A2** assumes that preferences are recursive. It states that preferences over two-stage lotteries respect the preference relation over single-stage lotteries, in the sense that two compound lotteries that differ only in the outcome of a single branch are compared exactly as these different outcomes would be compared separately.

Proposition (Segal [1990]): $\succeq$ *satisfies* **A1** *and* **A2** *iff it can be represented by a continuous function* $W : \mathcal{L}^2 \longrightarrow \mathbb{R}$ *of the following form:*

$$W \left(\left\langle \alpha_1, q^1; ...; \alpha_l, q^l \right\rangle \right) = V \left(\alpha_1 \delta_{c_V(q^1)} + ... + \alpha_l \delta_{c_V(q^l)} \right)$$

Note that under $A1$ and $A2$, the preference relation $\succeq_1 = \succeq_\Gamma = \succeq_\Delta$ fully determines $\succeq$.

---

[2] (i) A preference relation $\succeq$ on a set Z is a complete and transitive binary relation on Z.

(ii) A real valued function V represents the preference relation $\succeq$ on a set Z if for all $z_1, z_2 \in Z$, $z_1 \succeq z_2 \Leftrightarrow V(z_1) \geq V(z_2)$

(iii) Continuity is in the topology of weak convergence.

(iv) Monotonicity is with respect to the relation of first-order stochastic dominance.

The decision maker evaluates two-stage lotteries by first calculating the certainty equivalent of every second-stage lottery using the preferences represented by V, and then calculating (using V again) the first-stage value by treating the certainty equivalents of the former stage as the relevant prizes. As only the function V matters, we drop its index from the certainty equivalents in the remainder of the paper. Furthermore, we slightly abuse notation by writing $V(Q)$, instead of $W(Q)$, for the value of the two-stage lottery Q. Lastly, since under the above assumptions $V(p) = V(\langle 1, p \rangle) = V\left(\langle q_{x_i}, \delta_{x_i} \rangle_{i=1}^m\right)$ for all $p \in \mathcal{L}^1$, we simply write $V(p)$ for this common value.

1.2.2   Main properties

We now introduce and motivate our two main behavioral assumptions. The first is dynamic, whereas the second is static.

Preference for one-shot resolution of uncertainty

We model an individual, DM, whose concept of uncertainty is multi-stage and who cares about the way uncertainty is resolved over time. In this section, we define consistent preferences to have all uncertainty resolved in "one-shot" rather than "gradually" or vice versa.

Fix $p \in \mathcal{L}^1$ and denote its support by $S(p)$, that is, $S(p) = \{x \,|\, p_x > 0\}$. Let

$$\mathcal{P}(p) := \left\{ \langle \alpha_i, p^i \rangle_{i=1}^K \in \mathcal{L}^2 \,\middle|\, K \in \mathbb{N} \text{ and } \forall x \in S(p) = \underset{i}{\cup} S(p_i), \ p_x = \sum_{i=1}^K \alpha_i p_x^i \right\}$$

$\mathcal{P}(p)$ is the set of all two-stage lotteries that induce the same probability distribution over final outcomes as p does. For example, if p is a lottery that gives the prize x_1 with probability 0.3 and the prize x_2 with the remaining probability, then the two-stage lottery $Q = \langle 0.6, q; 0.4, r \rangle$, where q gives both prizes with equal probability and r yields x_2 for sure, is in $\mathcal{P}(p)$.

Let

$$\mathcal{P}^O\left(p\right) := \left\{Q \mid Q \in \mathcal{P}\left(p\right) \cap \left(\Gamma \cup \Delta\right)\right\} = \left\{\langle 1, p\rangle, \ \langle p_x, \delta_x\rangle_{x \in S(p)}\right\}$$

$\mathcal{P}^O\left(p\right)$ contains all elements of $\mathcal{P}\left(p\right)$ that are resolved in a single stage.

Definition: $\succeq$ displays *preference for one-shot resolution of uncertainty* (PORU) if $\forall p \in \mathcal{L}^1$ and $\forall Q \in \mathcal{P}\left(p\right)$, $R \in \mathcal{P}^O\left(p\right)$ implies $R \succeq Q$. If, subject to the same qualifiers, $R \in \mathcal{P}^O\left(p\right)$ implies $Q \succeq R$, then $\succeq$ displays *preference for gradual resolution of uncertainty* (PGRU).

PORU implies an aversion to receiving partial information. If uncertainty is not fully resolved in the first stage, DM prefers to remain fully unaware till the final resolution is available. PGRU implies the opposite. As we will argue in later sections, these notions render "the frequency at which the outcomes of a random process are evaluated" a relevant economic variable.

The Allais paradox and axiom NCI

In a generic Allais-type questionnaire,[3] subjects choose between A and B, where $A = \delta_{300}$ and $B = 0.8\delta_{400} + 0.2\delta_0$. They also choose between C and D, where $C = 0.25\delta_{300} + 0.75\delta_0$ and $D = 0.2\delta_{400} + 0.8\delta_0$. The majority of subjects tend to *systematically* violate expected utility by choosing the pair A and D.

Since Allais's [1953] original work, numerous versions of his questionnaire have appeared, most of which contain one lottery that does not involve any risk. Kahneman and Tversky use the term "certainty effect" to explain the commonly observed behavior. Their idea is that individuals tend to put more weight on certain events in comparison with very likely, yet uncertain, events. Although verbally it might appear to be intuitive reasoning, it is behaviorally translated into a nonlinear probability-weighting function, $\pi : [0, 1] \to [0, 1]$, that individuals are assumed to use when evaluating risky prospects. In particular, this function has a steep slope near –or even a discontinuity point at– 0 and 1. As we remark below, this implication has its own limitations. We thus suggest a direct behavioral property that is motivated by similar insights, but is less restrictive. Consider the following axiom on $\succeq_1$:

---

[3] Also known as "common-ratio effect with a certain prize."

Negative Certainty Independence (NCI):[4] $\forall p, q, \delta_x \in \mathcal{L}^1$ and $\lambda \in [0, 1]$, $p \succeq_1 \delta_x$ implies $\lambda p + (1 - \lambda)q \succeq_1 \lambda\delta_x + (1 - \lambda)q$.

The axiom states that if the sure outcome x is not enough to compensate DM for the risky prospect p, then mixing it with any other lottery, thus eliminating its certainty appeal, will not result in the mixture of x being more attractive than the corresponding mixture of p. The implication of this axiom on responses in Allais questionnaire above is: If you choose B, then you must choose D. This prediction is empirically rarely violated (see for example "pattern 2" in Conlisk [1989]). As stated above, the intuition behind NCI is that the sure outcome loses relatively more (or gains relatively less) than any other lottery from the mixture with the other lottery, q, but it does not imply any probabilistic distortion. This becomes relevant in experiments like those of Conlisk [1989], who studies the robustness of Allais-type behavior to small perturbations of the questionnaire which remove boundary effects. Although violations in that case were no longer systematic, a nonlinear probability function, as suggested above, predicts that this increase in consistency would be the result of fewer subjects choosing (the slightly perturbed) A over B, and not because more subjects choose (the slightly perturbed) C over D. In fact, the latter occurred, which is consistent with NCI.

Proposition 1: *Under* **A1** *and* **A2**, *if* $\succeq_1$ *satisfy NCI, then* $\succeq$ *display PORU*

Proof : We need to show that an arbitrary two-stage lottery, $\left\langle \alpha_1, q^1; ...; \alpha_l, q^l \right\rangle$, is never preferred to its single-stage counterpart, $\left\langle 1, \sum_{i=1}^{l} \alpha_i q^i \right\rangle$. Using **A1** *and* **A2** we have:

$$\left\langle \alpha_1, q^1; ...; \alpha_l, q^l \right\rangle \overset{(A_2)}{\sim} \left\langle \alpha_1, \delta_{c(q^1)}; ...; \alpha_l, \delta_{c(q^l)} \right\rangle \overset{(A_1)}{\sim} \left\langle 1, \sum_{i=1}^{l} \alpha_i \delta_{c(q^i)} \right\rangle$$

---

[4]We use the word "negative" since this axiom can, equivalently, be stated as: $\forall p, q, \delta_x \in \mathcal{L}^1$ and $\lambda \in [0, 1)$, $\delta_x \not\succ_1 p$ *implies* $\lambda\delta_x + (1 - \lambda)q \not\succ_1 \lambda p + (1 - \lambda)q$. Here $\succ_1$ is the asymmetric part of $\succeq_1$, and $\not\succ_1$ is its negation.

And by repeatedly applying NCI,

$$\sum_{i=1}^{l} \alpha_i \delta_{c(q^i)} = \alpha_1 \delta_{c(q^1)} + (1 - \alpha_1) \left(\sum_{i \neq 1} \frac{\alpha_i}{(1 - \alpha_1)} \delta_{c(q^i)} \right) \overset{\text{(NCI)}}{\preceq}_1$$

$$\alpha_1 q^1 + (1 - \alpha_1) \left(\sum_{i \neq 1} \frac{\alpha_i}{(1 - \alpha_1)} \delta_{c(q^i)} \right) =$$

$$\alpha_2 \delta_{c(q^2)} + (1 - \alpha_2) \left(\frac{\alpha_1}{(1 - \alpha_2)} q^1 + \sum_{i \neq 1,2} \frac{\alpha_i}{(1 - \alpha_2)} \delta_{c(q^i)} \right) \overset{\text{(NCI)}}{\preceq}_1$$

$$\alpha_1 q^1 + \alpha_2 q^2 + \sum_{i \neq 1,2} \alpha_i \delta_{c(q^i)} = ... =$$

$$\alpha_l \delta_{c(q^l)} + (1 - \alpha_l) \left(\sum_{i \neq l} \frac{\alpha_j}{(1 - \alpha_l)} q^i \right) \overset{\text{(NCI)}}{\preceq}_1 \sum_{i=1}^{l} \alpha_i q^i$$

Therefore, $\left\langle \alpha_1, q^1; ...; \alpha_l, q^l \right\rangle \sim \left\langle 1, \sum_{i=1}^{l} \alpha_i \delta_{c(q^i)} \right\rangle \preceq \left\langle 1, \sum_{i=1}^{l} \alpha_i q^i \right\rangle$ $\blacksquare$

The idea behind proposition 1 is simple: the second step of the folding-back procedure involves mixing all certainty equivalents of the corresponding second-stage lotteries. Applying NCI repeatedly implies that each certainty equivalent suffers from the mixture at least as much as the original lottery that it replaces would.

Proposition 1 states that NCI is a sufficient condition for PORU. To show necessity, we need to impose more structure. For the rest of the section, we confine our attention to a class of preferences $\succeq \in \mathcal{V}$ that satisfy the betweenness axiom.

A3 (*single-stage betweenness*) $\forall \ p, q \in \mathcal{L}^1$ and $\alpha \in [0, 1]$, $p \succeq_1 q$ implies $p \succeq_1 \alpha p + (1 - \alpha) q \succeq_1 q$

A3 is a weakened form of the vNM-independence axiom. It implies neutrality toward randomization among equally-good prizes/lotteries. It yields the following representation:

Proposition (Dekel [1986]): $\succeq_1 \in V$ *satisfies* **A3** *iff there exists a local utility function* $u : X \times [0, 1] \to [0, 1]$, *which is continuous in both arguments, strictly increasing in the first argument and satisfies* $u(w, v) = 0$ *and* $u(b, v) = 1$ *for all* $v \in [0, 1]$, *such that for all* $p \in \mathcal{L}^1$, $V(p)$ *is defined implicitly by:*

$$V(p) = \sum_{x \in X} u(x, V(p)) p_x$$

NCI in the probability triangle

The betweenness axiom (**A3**), along with monotonicity, implies that indifference curves in any unit probability triangle are positively sloped straight lines. To demonstrate this result using the representation theorem, note that for any lottery p over a given triple $x_3 > x_2 > x_1$, $V(p) = p_1 u(x_1, V(p)) + (1 - p_1 - p_3) u(x_2, V(p)) + p_3 u(x_3, V(p))$. The slope of any indifference curve in the corresponding two-dimensional space, $\Delta := \{(p_1, p_3) \mid p_1, p_3 \geq 0, \ p_1 + p_3 \leq 1\}$ is given by:

$$\mu\left(V|x_3, x_2, x_1\right) = \frac{u(x_2, v) - u(x_1, v)}{u(x_3, v) - u(x_2, v)}$$

which is positive and independent of the vector of probabilities. By definition, the slope represents the marginal rate of substitution between p_3 and p_1, and as explained by Machina [1982], changes in the slope express local changes in attitude towards risk: the greater the slope, the more risk averse DM is.

Definition: $\succeq_1$ has the *steepest middle slope* property if for every triple $x_3 > x_2 > x_1$ and for all $v \in (V(\delta_{x_1}), V(\delta_{x_3}))$,

$$\mu\left(V(\delta_{x_2})|x_3, x_2, x_1\right) \geq \mu\left(V|x_3, x_2, x_1\right)$$

that is to say, this property holds if for every three prizes $x_3 > x_2 > x_1$, the indifference curve through δ_{x_2} is the steepest.

Observe that NCI implies the steepest middle slope property. To see this, let $I_{V(\delta_{x_2})} := \{p' \in \Delta : p' \sim_1 \delta_{x_2}\}$ and let $\overline{\mu} := \mu\left(V(\delta_{x_2})|x_3, x_2, x_1\right)$. Take any lottery $p \in I_{V(\delta_{x_2})}$. For any $\lambda \in [0, 1]$ and $q \in \Delta$, both $\lambda p + (1 - \lambda)q$ and $\lambda \delta_x + (1 - \lambda)q$ are in Δ (a convex set) and by the triangle proportional sides theorem, the line segment that connects them has a slope that equals $\overline{\mu}$. But NCI requires that $\lambda p + (1 - \lambda)q \succeq_1 \lambda \delta_x + (1 - \lambda)q$ and since indifference curves are upward sloping, the indifference curve that passes through $\lambda p + (1 - \lambda)q$ must have a slope no greater than $\overline{\mu}$. Since λ and q were arbitrary, the result follows.

1.2.3 Characterization

Definition : $\succeq_2$ is *betweenness-recursive* if it satisfies $\mathbf{A1} - \mathbf{A2}$ and its restrictions to $\mathcal{V}$ satisfy $\mathbf{A3}$.

Theorem 1: *For any betweenness-recursive preferences, the following three statements are equivalent:*

$(i) \succeq$ *displays PORU.*

(ii) $\succeq_1$ *satisfies NCI.*

$(iii) \succeq_1$ *has the steepest middle slope property.*

A characterization of PGRU is analogously obtained by reversing the weakly preferred sign in NCI, and replacing steepest with flattest in (iii) .

The detailed proof is in the appendix. The main step in it is to establish, using certain properties of preferences from the betweenness class, that PORU is equivalent to the following condition:

$$\mathbf{C_1} : \left[\sum\nolimits_{x \in S(p)} u\left(x, v\right) p_x - u\left(\mathrm{c}\left(p\right), v\right) \right] \geqslant 0 \ \forall p \in \mathcal{L}^1 \text{ and } \forall v \in V\left(\mathcal{L}^1\right)$$

where $V\left(\mathcal{L}^1\right) := \left\{ V \middle| \exists p \in \mathcal{L}^1 \text{ with } v = V(p) \right\}.$[5] We interpret $\mathbf{C_1}$ by exploiting the main idea behind the construction of the local utility function, $u\left(x, v\right)$. As explained by Dekel [1986], and demonstrated in figure 1, one can think of $u\left(x, v\right)$ as a collection of functions that are derived in the following way: Fix an indifference hyperplane with a value v (denoted by I_v in figure 1a) and construct a collection of parallel hyperplanes relative to it. This collection can be taken to represent some expected utility preferences with an associated

---

[5]The specific normalization $V\left(\mathcal{L}^1\right) = [0, 1]$ is inessential for this result.

Bernoulli function $u_v(x)$. For every lottery p, we can then calculate $V(p,v) := E_p[u_v(x)]$, its expected utility relative to the value v (figure 1b). Repeat this construction for every value of v (which is bounded above and below, since X is bounded) to get the collection of functions $\{u_v(x)\}_{v \in V(\mathcal{L}^1)}$ that are equal to $u(x,v)$. $\mathbf{C_1}$ then implies that DM becomes the most risk averse at the true lottery value. That is, if relative to $V(p)$, the true utility level, DM is just indifferent between p and the certain prize $c(p)$, then relative to any other value v, he (weakly) prefers the lottery. The graphical illustration of $\mathbf{C_1}$ in the probability triangle is precisely item (iii) in the theorem (figure 1c), whereas item (ii) is its direct behavioral interpretation. The proof is completed by ensuring sufficiency of item (iii) to $\mathbf{C_1}$ and using proposition 1.
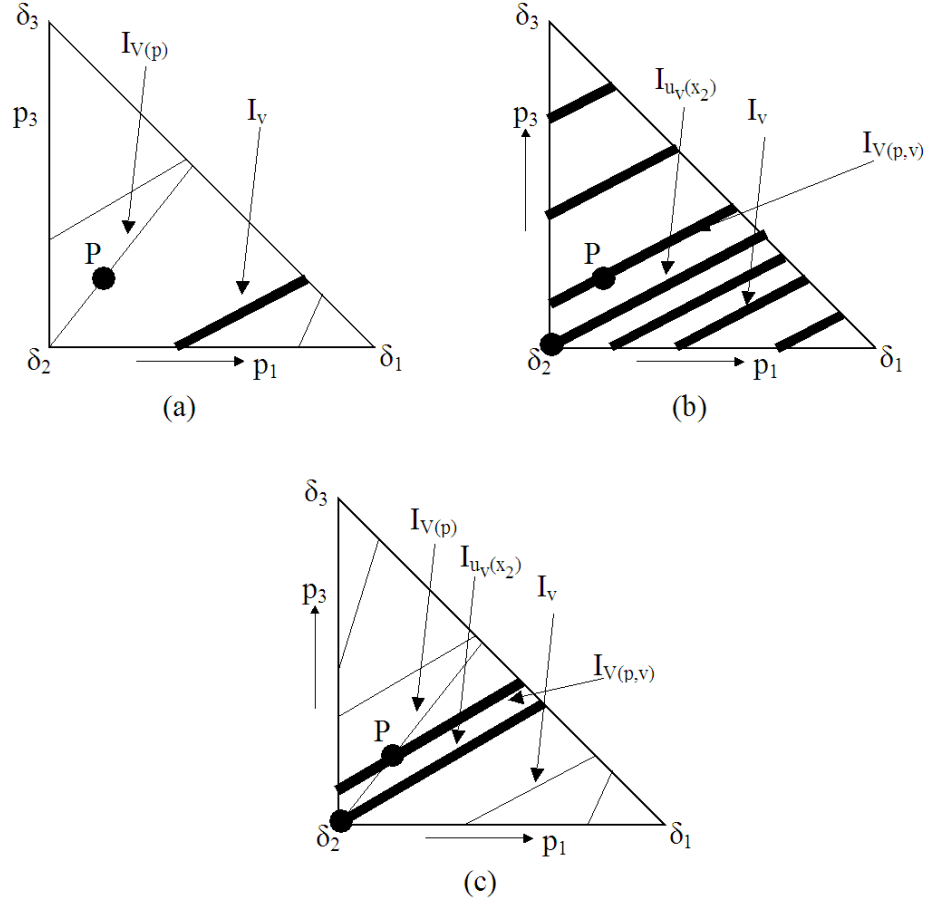

Figure 1: 1a: Fixing an indifference curve of level v. 1b: Constructing the local utility function $u_v(x)$. 1c: Putting them together, $E_p[u(x, V(p))] = u(x_2, V(p))$, but
$$E_p[u(x,v)] > u(x_2, v) \text{ for } v \neq V(p).$$

15

Theorem 1 ties together three notions that are defined on different domains: PORU is a dynamic property, NCI is a static property, and the third item is a geometrical condition, which applies to single-stage lotteries with at most three prizes in their support. The core of the theorem is the equivalence of PORU and NCI, which suggests that being prone to Allais-type behavior and being averse to the gradual resolution of uncertainty are synonymous. This assertion justifies the proposed division of the space of two-stage lotteries into the one-shot and gradually resolved lotteries. On the one hand, numerous replications of the Allais paradox in the last fifty years prove that the availability of a certain prize in the choice set is important and affects behavior in a systematic way. Moreover, we have no firm evidence of a consistent attitude towards lotteries, all of which lie in the interior of a probability triangle. On the other hand, empirical and experimental studies involving dynamic choices and experimental studies on preference for uncertainty resolution are still rather rare. Theorem 1 thus provides new theoretical predictions for dynamic behavior, based on robust (static) empirical evidence.

The applicability of the steepest middle slope property stems from its simplicity. In order to detect violation of PORU, one need not construct the (potentially complicated) exact choice problem. Rather, it is sufficient to introspect the slopes of one-dimensional indifference curves. This, in turn, is a relatively simple task, at least once a local utility function is given.

1.3   Static implications

1.3.1   NCI and differentiability

In most economic applications, it is assumed that individuals' preferences, and therefore the utility functions that represent them, are not only continuous, but also at least twice differentiable.[6] The following result demonstrates that among the betweenness class, smoothness and NCI are inconsistent, in the sense that coupling them leads us back to expected utility.

---

[6]Debreu [1972] provides, for any $k > 0$, a formal definitions of kth- order differentiable preferences.

Theorem 2: *Suppose $u(x,v)$ is at least twice differentiable with respect to both its arguments, and that all derivatives are continuous and bounded. Then preferences satisfy NCI if and only if they are expected utility.*

Expected utility preferences are characterized by the independence axiom that implies NCI. To show the other direction, we fix $\overline{v}$ and denote by $x(\overline{v})$ the unique x satisfying $\overline{v} = u(x,\overline{v})$. Combining the geometrical characterization (theorem 1 item (iii)) of NCI with differentiability implies that for any $x > x(\overline{v}) > w$, the derivative with respect to v of the slope of an indifference curve on the corresponding probability triangle must vanish at $\overline{v}$. We use the fact that this statement is true for any $x > x(\overline{v})$ and that $\overline{v}$ is arbitrary to get a differential equation with a solution on $\{(x,v)\,|v < u(x,v)\}$ given by $u(x,v) = h^1(v)\,g^1(x) + f^1(v)$, and $h^1(v) > 0$. We perform a similar exercise for $x < x(\overline{v}) < b$ to uncover that on the other region, $\{(x,v)\,|v < u(x,v)\}$, $u(x,v) = h^2(v)\,g^2(x) + f^2(v)$, and $h^2(v) > 0$. Continuity and differentiability then imply that the functional form is equal in both regions, therefore for all x, $u(x,v) = h(v)\,g(x) + f(v)$, and $h(v) > 0$. The uniqueness theorem for betweenness representations establishes the result.

1.3.2   Examples

Expected utility preferences are a trivial example of preferences that in a dynamic context satisfy PORU; DM with such preferences is just indifferent to the way uncertainty is resolved. The following is an important class of preferences for which, when applied recursively, PORU is a meaningful concept:

Preferences that satisfy the mixed-fan hypothesis. This set consists of all preferences whose indifference curves, in any unit probability triangle, have the following pattern: Moving northwest, they first get steeper ("fanning out") in the lower-right sub triangle (the less-preferred region), and then get flatter ("fanning in") in the upper-left sub triangle (the more-preferred region). Before giving examples from this class, we first state sufficient restrictions on the local utility function to satisfy the mixed-fan hypothesis.[7]

---

[7]Neilson [1992] provides sufficient conditions for smooth (in the sense of theorem 2) betweenness preferences to satisfy the mixed-fan hypothesis. The additional requirement, that the switch between "fanning out" and "fanning in" always occurs at the indifference curve that passes through the origin (the lottery that yields the middle prize for certain), renders those conditions empty, as is evident from theorem 2.

17

Denote by $L(x) := \left\{ p \in \mathcal{L}^1 : \delta_x \succsim_1 p \right\}$ the lower contour set of $x \in X$.

Sufficient conditions for mixed fan: *If $u(x,v)$ is a local utility function of the form*

$$u(x,v) - v = \begin{cases} u^1(x,v) & V^{-1}(v) \in L(x) \\ u^2(x,v) & V^{-1}(v) \notin L(x) \end{cases}$$

with the following restrictions:

(1) $\frac{\partial}{\partial x}\frac{\partial}{\partial v} u^1(x,v) \leqslant 0$,

(2) $\frac{\partial}{\partial x}\frac{\partial}{\partial v} u^2(x,v) \leqslant 0$, *and*

(3) $\inf\limits_{x} \frac{\partial}{\partial v} u^1(x,v) \geqslant \sup\limits_{x} \frac{\partial}{\partial v} u^2(x,v)$

then preferences satisfy the mixed-fan hypothesis.

Chew [1989] axiomatizes semi-implicit weighted utility. The local utility function he considers is

$$u(x,v) - v = \begin{cases} \overline{w}(x)(x-v) & x > v \\ \underline{w}(x)(x-v) & x \leqslant v \end{cases}$$

with $\underline{w}(x) > 0$, $\overline{w}(x) > 0$, $\underline{w}'(x) \geqslant 0$, $\overline{w}'(x) \geqslant 0$. To ensure that these preferences satisfy the mixed-fan hypothesis, we add the restriction that $\inf\limits_{x}\underline{w}(x) > \sup\limits_{x}\overline{w}(x)$.

Gul [1991] proposes a theory of disappointment aversion. He derives the local utility function

$$u(x,v) = \begin{cases} \frac{\phi(x)+\beta v}{1+\beta} & \phi(x) > v \\ \phi(x) & \phi(x) \leqslant v \end{cases}$$

with $\beta > 0$ and $\phi : X \to \mathbb{R}$ increasing.

Gul's notion of disappointment aversion amounts to dividing the support of each lottery into two groups, the elated outcomes and the disappointed outcomes, and giving the disappointed outcomes a uniformly greater weight when calculating the expected utility of the lottery.[8] For these preferences, the sign of β, the coefficient of disappointment aversion, un-

---

[8]Although Gul's preferences imply probability transformation, this transformation is done endogenously.

ambiguously determines whether preferences satisfy PORU or PGRU (see Artstein-Avidan and Dillenberger [2006]).

PORU can be interpreted as dynamic disappointment aversion. As suggested by Palacios-Huerta [1999], one may argue that being exposed to the resolution process bears the risk of perceiving intermediate outcomes as disappointing or elating, and if DM is more sensitive to disappointments, he would prefer to know only the final result. The term "disappointment aversion preferences" usually refers to Gul's static model. Our dynamic notion of disappointment aversion is translated into a strong restriction on indifference maps across probability triangles. Although Gul's model satisfies both, it is a boundary case. To emphasize the distinction between these two notions, we provide examples of other betweenness-satisfying preferences that were suggested as one-parameter generalizations of Gul's static model but, nevertheless, dynamically violate PORU. As implied by theorem 1, to track down violations of PORU, it is enough to show that neither of the preferences below satisfy the steepest middle slope property.[9]

Nehring [2005] suggests preferences which are represented by an implicit utility function of the following form:

$$u\left(x,v\right) - v = \begin{cases} \left(\phi\left(x\right) - \phi\left(v\right)\right)^{\alpha} & x > v \\ -\beta\left(\phi\left(v\right) - \phi\left(x\right)\right)^{\alpha} & x \leqslant v \end{cases}$$

with $\alpha, \beta > 0$. Gul's model corresponds to the case of $\alpha = 1$, and disappointment aversion implies $\beta > 1$.

Nehring interprets $u\left(x,v\right) - v$ as relative utilities (outcomes are evaluated psychologically relative to a certain reference point) and $\phi\left(x\right)$ as absolute utilities. He shows that such a class is uniquely characterized by the "bi-linearity" property: There exists a monotonic and continuous function $\pi : [0,1] \rightarrow [0,1]$ and a mapping $\phi : X \rightarrow \mathbb{R}$, such that preferences restricted to binary lotteries are represented by the function:

It is the value of each elated prize, and not its probability, which is explicitly down-weighted.

[9]Gul's preferences are one parameter (β) richer than expected utility preferences. The economic interpretation of β in a dynamic context is not evident. Indeed, one of Gul's axioms (axiom 4 in his paper) is necessary to identify β, but is unrelated to NCI. It is imposed in order to rule out further deviations from the expected utility model.

$V(p\delta_x + (1-p)\delta_y) = \pi(p)\phi(x) + (1-\pi(p))\phi(y)$, for $x > y$. Unless $\alpha = 1$ (and $\beta \geqslant 1$), no member of this class of preferences satisfies NCI.

Routledge and Zin [2004] provide a different one-parameter extension of Gul's model, enabling the identification of outcomes as disappointing only when they lie sufficiently below the (implicit) certainty equivalent. They derive the representation:

$$\phi\left(\mathrm{c}\left(p\right)\right) = \sum_{x_i} p\left(x_i\right)\phi\left(x_i\right) - \beta \sum_{x_i \leqslant \delta c(p)} p\left(x_i\right)\left[\phi\left(\delta \mathrm{c}\left(p\right)\right) - \phi\left(x_i\right)\right]$$

with $\delta \leqslant 1$. Note that in Gul's model $\delta = 1$ (where $\beta = 0$ corresponds to expected utility). Unless $\delta = 1$, these preferences also do not satisfy NCI.

1.4 Gradual resolution premium

For further purposes, we first extend our results to finite-stage lotteries.

1.4.1 Extension to n-stage lotteries

Fix $n \in \mathbb{N}$ and denote the space of finite n-stage lotteries by $\mathcal{L}^n$. We interpret the parameter n as the "resolution sensitivity" of an individual. It describes the frequency with which an individual updates information in a fixed time interval, which is a characteristic of preferences. The extension of our setting to $\mathcal{L}^n$ is the following (a formal description is given in the appendix): Occupied with a continuous and increasing function $V : \mathcal{L}^1 \to \mathbb{R}$, DM evaluates any n-stage lottery by folding back the probability tree and applying the same V in each stage. Preferences for one-shot resolution of uncertainty implies that DM prefers to replace each (compound) sub-lottery with its single-stage counterpart. The equivalence between PORU and NCI remains intact. In what follows, we will continue simplifying notation by writing $V(Q)$ for the value of any multi-stage lottery Q. We sometimes write Q^n to emphasize that we consider an n-stage lottery.

1.4.2 Definitions

Denote by $e(p)$ the expectation of a lottery $p \in \mathcal{L}^1$, that is, $e(p) = \sum_x x p_x$. Let $G(p,x) := \sum_{z \geqslant x} p_z$. We say that lottery p second-order stochastically dominates lottery q, and denote it by p sosd q, if for all $t < K$, $\sum_{k=0}^{t} [G(p, x_{k+1}) - G(q, x_{k+1})] [x_{k+1} - x_k] \geqslant 0$, where $x_0 < x_1 < ... < x_K$ and $\{x_0, x_1, ..., x_K\} = S(p) \cup S(q)$. DM is risk averse if $\forall p, q \in \mathcal{L}^1$ with $e(p) = e(q)$, p sosd q implies $p \succeq_1 q$.

For any $p \in \mathcal{L}^1$, the risk premium of p, denoted by $\mathrm{rp}(p)$, is the number satisfying $\delta_{e(p) - \mathrm{rp}(p)} \sim_1 p$. $\mathrm{rp}(p)$ is the amount that DM would pay to replace p with its expected value. By definition, $\mathrm{rp}(p) \geqslant 0$ whenever DM is risk averse.[10]

Definition: Fix $p \in \mathcal{L}^1$. For any $Q \in \mathcal{P}(p)$, the *gradual resolution premium* of Q, denoted by $\mathrm{grp}(Q)$, is the number satisfying $\langle 1, \delta_{c(p) - \mathrm{grp}(Q)} \rangle \sim Q$.

$\mathrm{grp}(Q)$ is the amount that DM would pay to replace Q with its single-stage counterpart. By definition, PORU implies $\mathrm{grp}(Q) \geqslant 0$. Since $c(p) = e(p) - \mathrm{rp}(p)$, we can, equivalently, define $\mathrm{grp}(Q)$ as the number satisfying $\langle 1, \delta_{e(p) - \mathrm{rp}(p) - \mathrm{grp}(Q)} \rangle \sim Q$.[11]

Observe that the signs of the two variables above, $\mathrm{rp}(p)$ and $\mathrm{grp}(Q)$, need not agree. In other words, (global) risk aversion does not imply, and is not implied by, PORU. Indeed, Gul's symmetric disappointment aversion preferences (see section 3) are risk averse if and only if $\beta \geqslant 0$ and $\phi : X \to \mathbb{R}$ is concave (Gul's [1991] theorem 3). However, for sufficiently small $\beta \geqslant 0$ and sufficiently convex ϕ, one can find a lottery p with $\mathrm{rp}(p) < 0$, whereas $\beta \geqslant 0$ is sufficient for $\mathrm{grp}(Q) \geqslant 0$ for any $Q \in \mathcal{P}(p)$. On the other hand, if $\lambda'(v) > 0$ and

---

[10] Weak risk aversion is defined as follows: For all p, $\delta_{e(p)} \succeq p$. This definition is not appropriate once we consider preferences that are not expected utility. The definition of the risk premium, on the other hand, is independent of the preferences considered.

[11] Similarly to the risk premium, the complete resolution premium is measured in monetary units. For this reason, these two premiums are different from the timing premium for early resolution, as suggested by Chew and Epstein [1989], which is measured in terms of probabilities.

$\lambda(v) > 1$ for all v,[12] then the local utility function

$$u(x, v) = \begin{cases} x & x > v \\ v - \lambda(v)(v - x) & x \leqslant v \end{cases}$$

has the property that $u(\cdot, v)$ is concave for all v. Therefore, DM is globally risk averse (Dekel's [1986] property 2), and hence $\mathrm{rp}(p) \geqslant 0 \ \forall p \in \mathcal{L}^1$. However, these preferences do not satisfy NCI,[13] meaning that there exists a lottery p and $Q \in \mathcal{P}(p)$ for which $\mathrm{grp}(Q) < 0$.

1.4.3   The magnifying effect

In the case where DM is both risk averse and has PORU, these two forces, as reflected in the two premiums previously defined, magnify each other. Understanding this, insurance companies, when offering dynamic insurance contracts, can require much greater premiums than the actuarially fair ones and still be sure of consumers' participation. This can explain why people often buy periodic insurance for moderately priced objects, such as electrical appliances and cellular phones, at much more than the actuarially fair rates. [14]

To illustrate, consider the following insurance problem: An individual with Gul's preferences, with a linear ϕ and a positive coefficient of disappointment aversion β, owns an appliance (e.g. a cellular phone) that he is about to use for n periods. The individual gets utility 1 in any period the appliance is used and 0 otherwise. In each period, there is an exogenous probability $(1 - p)$ that the appliance will not work (it might be broken, fail to get reception, etc.). The individual can buy a periodic insurance, which guarantees the availability of the appliance, for a price $z \in (1 - p, 1)$. Therefore, if he buys insurance for some period, he gets a certain utility of $(1 - z)$, and otherwise he faces the lottery in which

---

[12]The condition that $\lambda(v)$ is non-decreasing is both necessary and sufficient for $u(,)$ to be a local utility function. See Nehring [2005].

[13]Look at the slope of an indifference curve for values $x_3 > v > x_2 > x_1$. We have: $\mu(V|x_3, x_2, x_1) = \frac{\lambda(v)(x_2 - x_1)}{x_3 - v + \lambda(v)(v - x_2)}$. In this region, the slope is increasing in v if $x_3 > \frac{\lambda(v)(\lambda(v) - 1)}{\lambda'(v)} + v$. For a given v, we can always choose arbitrarily large x_3 that satisfies the condition, and construct, by varying the probabilities, a lottery whose value is equal to v. Apply this argument in the limit where $v = x_2$ to violate condition (iii) of theorem 1.

[14]A popular example is given by Tim Harford ("The Undercover Economist", *Financial Times*, May 13, 2006): "There is plenty of overpriced insurance around. A popular cell phone retailer will insure your \$90 phone for \$1.70 a week—nearly \$90 a year. The fair price of the insurance is probably closer to \$9 a year than \$90."

with a probability p he gets 1, and with the remaining probability he gets 0. For simplicity, assume that the price of a replacement appliance is 0, so that the individual either still has it from the last period or gets a new one for free in the beginning of any period.

Let $\widehat{p}$ be the probability distribution over final outcomes (without insurance). Denote by X the total number of periods in which the appliance works. Since X is a binomial random variable, $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, ..., n$. Applying Gul's formula, one obtains:

$$V_{\beta,n}(\widehat{p}) = \frac{\sum_{k=h+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} k + (1+\beta) \sum_{k=0}^{h} \binom{n}{k} p^k (1-p)^{n-k} k}{1 + \beta \sum_{k=0}^{h} \binom{n}{k} p^k (1-p)^{n-k}}$$

where $h(p, \beta, n)$ is the unique natural number such that all prizes greater than it are elated and all those smaller than it are disappointed.

Let Q be the corresponding gradual (n-stage) lottery *as perceived by DM*. Its value is:

$$V_{\beta,n}(Q) := \frac{1}{(1 + \beta(1-p))^n} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (1+\beta)^{n-k} k$$

Using standard backward induction arguments, it can be shown that DM will buy insurance for all periods if $\beta > \frac{z - (1-p)}{(1-z)(1-p)} > 0$. In that case, $z < 1 - \frac{V_{\beta,n}(Q)}{n}$. Nevertheless, if β is not too high,[15] we have $1 - p < 1 - \frac{V_{\beta,n}(\widehat{p})}{n} < z$, meaning that DM would not buy insurance at all *if he could avoid being aware of the gradual resolution of uncertainty*.[16] This observation explains why and how the attractiveness of a lottery depends not only on the uncertainty embedded in it, but also on the way this uncertainty is resolved over time.

Since $V_{\beta,n}(\widehat{p})$ decreases with β, $\text{rp}(\beta \,|p, n) := np - V_{\beta,n}(\widehat{p})$ is a strictly increasing function of β. The behavior of the gradual resolution premium, $\text{grp}(\beta \,|p, n) := V_{\beta,n}(\widehat{p}) - V_{\beta,n}(Q)$ is more subtle. We have the following result:

---

[15] The condition is: $1 + \beta < \min \left\{ \frac{p^n}{p^n + n(1-p) - 1}, \frac{p^n z}{(1-z)(1-p^n) - p(1-p^{n-1}) - 1} \right\}$.

[16] Nayyar [2004] termed such a situation an "insurance trap". Note that DM still acts rationally given that without insurance he is forced to be exposed to Q^n rather than to p.

23

Proposition 2: *In the insurance problem described above:*

(i) Strict PORU in the interior: $\mathrm{grp}\left(\beta\,|p,n\right) > 0 \; \forall \beta \in (0,\infty)$

(ii) Weak PORU in the extreme: $\mathrm{grp}\left(0\,|p,n\right) = 0$ *and* $\lim_{\beta \to \infty} \mathrm{grp}\left(\beta\,|p,n\right) = 0$

(iii) Single-peakness: *There exists* $\beta^*\left(p,n\right) < \infty$ *such that either* $0 < \beta < \beta' < \beta^*$ *or* $\beta^* < \beta' < \beta$ *implies*

$$\mathrm{grp}\left(\beta\,|p,n\right) < \mathrm{grp}\left(\beta'\,|p,n\right) < \mathrm{grp}\left(\beta^*\,|p,n\right)$$

See figure 1.1.
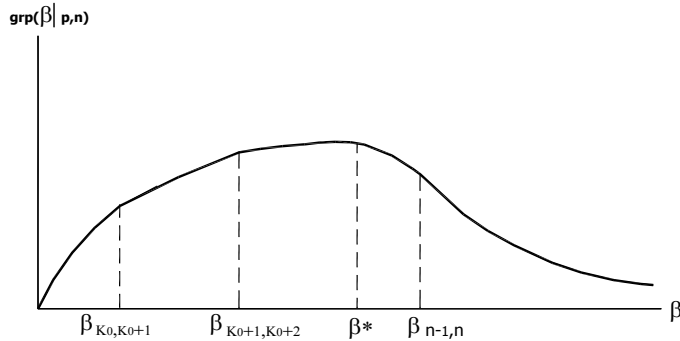
Figure 1.1: $\mathrm{grp}(\beta\,|p,n)$. $\beta_{k,k+1}$ is the value of β where $h\left(\beta\,|p,n\right)$ decreases from $(n-k)$ to $(n-(k+1))$. $\mathrm{grp}(\beta\,|p,n)$ is non-differentiable in each such $\beta_{k,k+1}$. k_0 is the smallest natural number that solves: $\max_{k' > n(1-p)} \frac{n-k'}{n}$

Recall that in Gul's model, the sign of the parameter β unambiguously determines whether preferences display PORU or PGRU. In its original context, greater β implies greater disappointment aversion (as well as greater risk aversion). Since we argued that PORU can be interpreted as dynamic disappointment aversion, it might seem intuitive to expect the gradual resolution premium to be an increasing function of β. This intuition is wrong and, in fact, item (ii) remains valid independent of the decision problem under consideration. In order to see this, note that $\mathrm{grp}(\beta\,|p,n)$ is defined as the difference of two functions, both strictly decreasing with β. When $\beta = 0$, DM cares only about the expected

value of the lottery. When β is sufficiently large, all prizes but 0 become elated, and hence the value of p converges to 0. Correspondingly, the value of the gradual lottery converges to the value of the worst sub-lottery that by itself approaches 0. Since item (i) reinforces the result of theorem 1 and states that $\mathrm{grp}(\beta\,|p,n)$ is actually strictly positive on the positive reals, and since $\mathrm{grp}(\beta\,|p,n)$ is a continuous function, there must exist a finite β, denoted β^* in figure 2, in which $\mathrm{grp}(\beta\,|p,n)$ is maximized. Item (iii) sheds further light on the behavior of moderate disappointment-averse individuals. It suggests that $\beta^*\,(p,n)$ is unique, and that $\mathrm{grp}(\beta\,|p,n)$ is single-peaked. Behaviorally speaking, a moderately disappointment-averse individual is more inclined to pay a higher premium, whereas individuals, who are either approximately disappointment-indifferent or very disappointment-averse, would not pay a substantial premium.

The analysis of the insurance problem suggests that, given n, extreme values of β neutralize the magnifying effect of the gradual resolution premium. In general, this premium can be very significant. By varying the parameter n, we change the frequency at which DM updates information. Our next result shows that high frequency of information updates might inflict an extreme cost on DM; a particular splitting of a lottery drives down its value to the value of the worst prize in its support.

Proposition 3: *Consider disappointment aversion preferences with some $\phi\,:\,X\,\rightarrow\,\mathbb{R}$ and $\beta > 0$. For any $\varepsilon > 0$, and for any lottery $p = \sum_{j=1}^m p_j \delta_{x_j}$, there exists $T < \infty$ and a multi-stage lottery $Q^T \in \mathcal{P}\,(p)$ such that $V\left(Q^T\right) < \min\limits_{x_j \in \mathrm{supp}(p)} \phi\,(x_j) + \varepsilon.$*

Let p be a binary lottery that yields 0 and 1 with equal probabilities. Consider n tosses of an unbiased coin. Define a series of random variables $\{z_i\}_{i=1}^n$ with $z_i = 1$ if the ith toss is "heads" and $z_i = 0$ if it is "tails". Let the terminal nodes of the n-stage lottery be:

$$
\begin{array}{ll}
1 & \text{if } \sum_{i=1}^n z_i > \frac{n}{2} \\
0.5\delta_1 + 0.5\delta_0 & \text{if } \sum_{i=1}^n z_i = \frac{n}{2} \\
0 & \text{if } \sum_{i=1}^n z_i < \frac{n}{2}
\end{array}
$$

Note that the value of this n-stage lottery, calculated using recursive disappointment aversion preferences, is identical to the value calculated using recursive expected utility and probability $\frac{0.5}{1+\beta 0.5} < 0.5$ for "heads" in each period. Applying the weak law of large numbers,

$$\Pr\left(\sum_{i=1}^{n} z_i < \frac{n}{2}\right) \to 1$$

and therefore, for n large enough, the value approaches $\phi(0)$. We use a similar construction to establish that this result holds true for any lottery.

Ignoring the dynamic aspect of risk aversion might be misleading. We have already argued that a substantial fraction of many insurance premiums we observe in daily life can be attributed to the gradual resolution premium. Proposition 3 proves that this effect is quantitatively important, if the parameter n is sufficiently large.

1.5   PORU, "loss aversion with narrow framing" and the final-wealth hypothesis

Loss aversion with narrow framing (also known as "myopic loss aversion") is a combination of two motives: loss aversion (Kahneman and Tversky [1979]), that is, people's tendency to be more sensitive to losses than to gains, and a dynamic aggregation rule, narrow framing (Kahneman and Tversky [1984]), that argues that when making a series of choices, individuals "bracket" them by making each choice in isolation. When applied to behavior in financial markets, narrow framing means that individuals tend to evaluate long-term investments according to their short-term returns. Benartzi and Thaler [1995] were the first to use this approach and suggest explanations for several economic "anomalies", such as the equity premium puzzle (Mehra and Prescott [1985]). Barberis and Huang [2005] and Barberis, Huang and Thaler [2006] generalize Benartzi and Thaler's work by assuming that DM derives utility directly from the outcome of a gamble over and above its contribution to total wealth.

The model presented in this paper can be used to address the same phenomena addressed with the loss aversion with narrow framing approach. Both models assume time

neutrality. The combination of a specific form of non-smooth atemporal preferences and the folding-back procedure accounts for PORU. In an intertemporal context, these two features are analogous to loss aversion and narrow framing, respectively. The gradual resolution premium is the cost an individual incurs from frequently evaluating the outcomes of a dynamic random process.

The loss aversion with narrow framing approach challenges the hypothesis that only final wealth matters. Rabin [2000] and Safra and Segal [2006] give a parallel critique on a broad class of smooth models of decision making under risk. These authors use calibration results to argue that modest risk aversion over small stakes gambles necessarily implies absurd levels of risk aversion over large stakes gambles. Both Safra and Segal [2006] and Barberis Huang and Thaler [2006] argue that if DM faces some background risk, then a similar problem persists even if preferences are non-differentiable (i.e. if preferences display first-order risk aversion[17]); merging new gambles with preexisting ones eliminates the effect of first-order risk aversion.

Our model is consistent with risk aversion over small stakes gambles and only moderate risk aversion over large stakes gambles even if individuals face background risks. If most risks resolve gradually, then they cannot be compounded into a single lottery. Our model then implies first order risk aversion over each realized gamble. In other words, the mere existence of other risks is not enough to apply Rabin-type critique. Such an argument is only compelling if DM compounds risks that are resolved over a long period.

The conceptual difference between the two approaches is twofold. First, loss aversion with narrow framing brings to the forefront the idea that individuals evaluate any new gamble separately from its cumulative contribution to total wealth. Both the reference points relative to which gains and losses are computed and the way they dynamically adjust are usually set exogenously.[18] We, on the other hand, maintain the assumption that terminal wealth matters, and identify narrow framing as a preference parameter. The similarity between "disappointment aversion" and "loss aversion" has already been pointed out in

---

[17]First order risk aversion means that the premium a risk averse DM is willing to pay to avoid an actuarially fair random variable $t\widetilde{\epsilon}$ is proportional, for small t, to t. It implies "kinked" indifference curves along the main diagonal in a states-of-the-world representation (Segal and Spivak [1990]).

[18]Köszegi and Rabin [2006] offer a model in which the reference point is determined endogenously.

Gul [1991] and stimulates further comparisons between these two notions. The novel insight provided by proposition 3 is that the (temporal) effect of narrow framing can be achieved even without giving up the assumption that utility depends on overall wealth, and that this effect is quantitatively important. Second, we set aside the question of why individuals are sensitive to the way uncertainty is resolved (i.e. why they narrow frame),[19] and construct a model that reveals the (context independent) behavioral implications of such considerations.

1.6 PORU and the value of information

We now reconsider the case of two-stage lotteries ($n = 2$). Let us suppose that just before the second-stage lottery is played, but after the realization of the first-stage lottery, DM can take, in the face of the remaining uncertainty, some action that might affect his ultimate payoff. The primitive in such a model is a preference relation over information systems (as we formally define below), which is induced from preferences over compound lotteries. Assume throughout this section that preferences over compound lotteries satisfy **A1** and **A2**. An immediate consequence of Blackwell's [1953] seminal result is that in the standard expected utility class, DM always prefers to have perfect information before making the decision, which allows him to choose the optimal action corresponding to the resulting state. Schlee [1990] shows that if $\succeq_1$ is of the rank-dependent utility class (Quiggin [1982]), then the value of perfect information will always be non-negative. This value is computed relative to the value of having no information at all, and therefore Schlee's result is salient about the comparison between getting complete and partial information. Safra and Sulganik [1995] left open the question of whether there are preference relations, other than expected utility, for which perfect information is always the most valuable. We show below that such preferences are fully characterized by PORU. Combining this result with theorem 1 reveals its implication on betweenness-recursive preferences.

More formally, let $S = \{s_1, \ldots, s_N\}$ be some finite set of states. Each state $s \in S$ occurs with probability p_s. The outcome of a lottery will depend both on the resulting state and

---

[19]Barberis and Huang [2006] suggest two different underlying sources of narrow framing. The first is based on a non-consumption utility, such as regret, and the second relates narrow framing to the "accessibility" of the uncertainty people confront. As these authors mention, each such motive, if taken literally, predicts different duration of narrow framing.

on an action DM has made. For this we let $A = \{a_1, ..., a_M\}$ be a finite set of actions. Let $u : A \times S \to R$ be a function that gives the outcome $u(a, s)$ if action $a \in A$ is taken and the realized state is $s \in S$. (This outcome corresponds to the final prize $x \in X$.)

The first-stage lottery can be thought of as a randomization over a set $J = \{j_1, ..., j_m\}$ of signals indexed by j (where signal j indicates that p^j was selected in the first-stage lottery). Let $\pi : S \times J \to [0, 1]$ be a function such that $\pi(s, j)$ is the probability of getting the signal $j \in J$ when the prevailing state is $s \in S$. We naturally require that for all $s \in S$, $\sum_{j \in J} \pi(s, j) = 1$ (so that when the prevailing state is s, there is some probability distribution on the signal DM might get). The function π is called an information structure. It automatically induces a splitting of the lottery into two stages, where with probability $\alpha_j(\pi) = \sum_{s' \in S} \pi(s', j) p_{s'}$, p^j is the second-stage lottery.

A full information structure, I, is a function such that for all $s \in S$ there exists $j(s) \in J$ with $\Pr(s | j(s)) = \frac{\pi(s, j(s)) p_s}{\sum_{s' \in S} \pi(s', j(s)) p_{s'}} = 1$, and for all $j \neq j(s)$ one has $\Pr(s | j) = 0$. In other words, in the sum above defining α_j, there is only one summand. The null information structure, ϕ, is a function such that $\Pr(s | j) = \Pr(s)$ for all $s \in S$ and $j \in J$.

Define $a^*(s)$ as the optimal action if you know that the prevailing state is s, that is, $u(a^*(s), s) := \max_{a \in A} u(a, s)$. Let $V^p(I)$ be the value of the lottery that assigns probability p_s to the outcome $u(a^*(s), s)$. After a signal j has been given, DM chooses the best a under the circumstances, namely a that maximizes the value of the lottery that assigns probability p_s^j to gain the outcome $u(a, s)$. We let $V(p^{j*})$ stand for the value of the jth lottery maximized over the choice of an action $a \in A$. Finally, let $V^p(\pi)$ be the value of the lottery where the action is taken after receiving signal j, that is, the compound lottery assigning probability $\alpha_j(\pi) = \sum_{s' \in S} \pi(s', j) p_{s'}$ to p^{j*}.

Definition: $\succeq$ displays *preferences for perfect information* if for any information structure π and for any payoff function u, $V^p(I) \geq V^p(\pi)$.

Proposition 4: *If $\succeq$ satisfies* **A1** *and* **A2***, then the two statements below are equivalent:*

(i) $\succeq$ *displays PORU*

(ii) $\succeq$ *displays preferences for perfect information.*

Analogously, PGRU holds if and only if for any information structure π and for any payoff function u, $V^p(\pi) \geq V^p(\phi)$

Showing that (i) is necessary for (ii) is immediate. For the other direction, we note that two forces reinforce each other: First, getting full information means that the underlying lottery is of the "one-shot resolution" type, since uncertainty is completely resolved by observing the signal. Second, better information enables better planning; using it, a decision maker with monotonic preferences is sure to take the optimal action in any state. The proof distinguishes between the two prime motives for getting full information: The former, which is captured by PORU, is intrinsic, whereas the latter, which is reflected via the monotonicity of preferences with respect to outcomes, is instrumental. The result for PGRU is similarly proven. The null information structure is of the "one-shot resolution" type and it has no instrumental value.

Corollary: *If $\succeq$ satisfies* **A1** *and* **A2***, then $\succeq$ displays preferences for perfect information whenever $\succeq_1$ satisfies NCI.*

Proposition 4 is independent of **A3**. By adding **A3** as a premise we get:

Corollary: *For any betweenness-recursive preferences, $\succeq$ displays preferences for perfect information iff $\succeq_1$ satisfies NCI.*

1.7 Application to investment under uncertainty

The concept of option value was initially demonstrated by Arrow and Fischer [1974], and later recognized in the works of McDonald and Siegel [1986], Pindyck [1991], and Dixit and Pindyck [1994]. These authors point out that if an investor has a choice over when to implement an (irreversible) investment decision, then investing according to the net present value (NPV) of a project is not adequate; waiting leaves room for new information that DM might use to make better decisions. In other words, the availability of future signals always favors delaying the investment.

This result rests on the assumption that decision makers are expected utility maximizers, and it ceases to hold once we relax that assumption. In particular, PORU suggest another effect that should be taken into account: The harmful effect of the gradual resolution of uncertainty, induced by an informative signal, can outweigh the benefit of getting more information from this same signal.

To illustrate, consider the following three-period investment problem.[20] In the first period, an investor decides whether or not to invest in a certain machine. The investment requires an immediate cost of C dollars. If he chooses to invest (option A), he will be able to produce in both the second and the third period. In the second period, the demand is certain and the investor is sure to receive variable profits $\pi^2 > 0$. In the third period, the demand is uncertain and the profits are determined by the realization of a finite random variable π^3. Denote by p the probability distribution over the third-period profits π_i^3, $i = 1, .., n$. If the investor decides not to invest in the first period (option B), he may still invest in the second period. In that case he waives π^2, but he is able, before making the investment decision, to learn the realization of a signal j that is correlated with π^3 and comes from a finite set $J = \{j_1, ..., j_m\}$. Thus, if he invests in the second period after receiving the signal j, his third-period's profits are distributed according to the conditional distribution p^j. Let α_j be the unconditional probability of getting the signal $j \in J$. The discount rate between any two successive periods is r.

Assume that DM has disappointment aversion preferences with linear ϕ and positive β.

---

[20]Based on an example given in Gollier [2001]

The value of option A, V^A, is given by:

$$V^A = -C + \frac{1}{1+r}\pi^2 + \frac{1}{(1+r)\,2}V(p)$$

Let $z_j := \max\left\{0, -C + \frac{1}{1+r}V(p^j)\right\}$. The value of option B is denoted V^B and is the unique value v that solves:

$$v = \frac{1}{1+r}\frac{\sum_{j:z_j>v}z_j\alpha_j + (1+\beta)\sum_{j:z_j\leq v}z_j\alpha_j}{1+\beta\sum_{j:z_j\leq v}\alpha_j}$$

Let $Q = \langle\alpha_j, p^j\rangle_{j=1}^m$ be the compound lottery such that for each i, $p_i = \sum_j \alpha_j p_i^j$. Denote its value by $V(Q)$. Further let $\Delta V := V^A - V^B$. The investor chooses option A if and only if:

$$\Delta V = \frac{1}{(1+r)^2}\left[\underbrace{(1+r)\left(\pi^2 - rC\right)}_{\text{net present value}} -(1+\beta)\underbrace{\frac{\sum_{j:V(p^j)\leqslant\min[V(Q),C(1+r)]}\alpha_j\left(C - \frac{V(p^j)}{(1+r)}\right)}{1+\beta\sum_{j:V(p^j)\leqslant V(Q)}\alpha_j}}_{\text{option value}} + \underbrace{(V(p) - V(Q))}_{\text{gradual resolution premium}}\right] \geq 0$$

The first component is the regular NPV rule: Invest today if the forgone second-period profits are larger than the interest gained due to delaying the investment. This would be the decision criterion in the absence of a signal for the case $\beta = 0$, when DM is risk neutral and simply maximizes the NPV of the investment.

The second component is the flexibility value, or the option value, of delaying the investment. It reflects the idea that occupied with more information, DM can refrain from investing if he learns that the demand is likely to be too low. This term is positive and is an increasing function of β.

Since $Q \in \mathcal{P}(p)$, the last term is the gradual resolution premium, $\mathrm{grp}(Q)$: investing today saves DM the need to be aware of the gradual resolution of uncertainty. This term is non-negative for $\beta \geqslant 0$.

As we mentioned above, if a standard expected utility maximizer prefers to invest in period two, even under the null-information system, then he clearly does so when the information system is finer. However, for strictly positive values of β this is not necessarily true. For example, suppose $\beta = 1$, $C = 50$, $r = 0.1$, $\pi^2 = 5$ and $\pi_i^3 \in \{0, 1000, 2000\}$. Let $p = \frac{1}{3}\delta_0 + \frac{1}{3}\delta_{1000} + \frac{1}{3}\delta_{2000}$, and $Q = \langle \delta_0, 1/3; \frac{1}{2}\delta_{1000} + \frac{1}{2}\delta_{2000}, 2/3 \rangle \in \mathcal{P}(p)$. Since $\pi^2 = rC$, in the absence of a signal, DM is simply indifferent between investing in period 1 and investing in period 2. The signal is useful in that if the investor learns that the quantity demanded is zero, he will choose not to invest. Nevertheless, this option value is not sufficient to compensate him for the compound lottery he must face in case he chooses to view a signal, and he strictly prefers option A. Therefore, the NPV rule should be twice adjusted, taking into account both the second and the third effects.

In the general case, it is not obvious which effect, the option value or PORU, dominates. Similarly to the assertion in proposition 2, there exists a finite value β^* in which the gradual resolution premium is maximized.[21] This observation implies the following:

(i) There exists $\overline{\beta}$, such that for all $\beta > \overline{\beta}$ the option value is dominant.

(ii) If the option value is dominant at β^*, so it is dominant for all $\beta > \beta^*$.

(iii) There exists $\underline{\beta}$, such that the option value is dominant for all $\beta \in \left(0, \underline{\beta}\right]$.

The setting above can be used to distinguish between decision makers with PORU and decision makers who have preferences for early or late resolution of uncertainty. The availability of an informative signal would induce decision makers with preferences for early resolution of uncertainty to choose option B. Independently of its instrumental value, a signal leads to an earlier, yet not complete, resolution of uncertainty. Therefore, the only possible confusion would be between the behavior of individuals with PORU and individuals who prefer late resolution of uncertainty. This confusion can be avoided by altering

---

[21] Since $\lim\limits_{\beta \to \infty} (V(p) - V(Q)) = 0$, there exists $\widehat{\beta} := \max\left\{\beta \,\middle|\, (V(p) - V(Q))\,|_\beta = \frac{\epsilon}{2}\right\}$ and $\widehat{\beta} < \infty$. Thus $(V(p) - V(Q))$ is a continuous function on the compact interval $\left[0, \widehat{\beta}\right]$, and hence achieves its maximum on this domain.

the resolution process in option A. Suppose that under that option, the uncertainty about period 3 returns would already have been resolved in period 2. Due to the time neutrality assumption, such a change has no effect on individuals with PORU. Individuals with preferences for late resolution of uncertainty, on the other hand, would be worse off under this alternative.

1.8   Conclusion

Searching for a better understanding of decision-making under risk, and disentangling decision makers' attitude towards risk and time have been two active fields of research in economics. This paper contributes to both fields. We study preferences over multi-stage lotteries and explicitly assume that the way uncertainty is resolved over time matters. Being exposed to the resolution process bears the risk of perceiving intermediate outcomes as disappointing or elating. Individuals who are more sensitive to disappointment suffer from getting partial information and, therefore, strictly prefer *ex-ante* all uncertainty to be resolved in a single point in time. Behaviorally, these individuals will display higher risk aversion if uncertainty is resolved gradually. We formally define such dynamic *preferences for one-shot resolution of uncertainty* (PORU), and show that they can be modeled using a single, static preference relation. Our main result states that to characterize PORU, one needs to impose on these static preferences a property, *negative certainty independence* (NCI), which is identified with Allais-type behavior, the most compelling argument against the independence axiom. In other words, being prone to Allais-type behavior and being averse to the gradual resolution of uncertainty are synonymous. This equivalence provides clear predictions for dynamic preferences, and calls for further experimental testing to be done. Our model also predicts a specific attitude towards information. Although we accommodate situations where people avoid information that is instrumental to their decision making, perfect information will never be rejected, and will always be preferred to any other information system.

The frequency with which an individual evaluates lotteries over time is a preference parameter in our model, and its effect is measured by the gradual resolution premium. The

more often an individual updates information, the more sensitive he is to gradual resolution. We show that this effect can be quantitatively important, implying extreme degrees of risk aversion.

If most actual risks that individuals face are resolved gradually over time, then these risks cannot be compounded into a single lottery and, therefore, the gradual resolution premium should not be disregarded. Rabin and Thaler [2001] state that "...it is clear that loss aversion and the tendency to isolate each risky choice must both be key components of a good descriptive theory of risk attitudes." Our model shows that in an intertemporal context, both features, and especially the isolation component, can be addressed independently of studying framing effects.

1.9 Appendix

1.9.1 Extension to n-stage lotteries, a formal description

The following is a formal description of any compound lottery, or a probability tree. Let T be a finite set of (chance) nodes. Let $\rhd_p$, "predecessor of", be a partial order on T with $x \rhd_p y$ if x precedes y. For any node $t \in T$, let $PRE(t) = \{x : x \rhd_p t\}$ be the set of predecessors of t. For any $t, t' \in T$, we say that t is an immediate predecessor of t', and denote it by $t \rhd_{ip} t'$, if $t \in PRE(t')$ and $\nexists t'' \in PRE(t')$ such that $t \in PRE\left(t''\right)$. An initial node is any $t \in T$ with $PRE(t) = \varnothing$. A pair $(T, \ \rhd_p)$ is a tree if it has a single initial node, and if for all $t \in T$, $PRE(t)$ is totally ordered by $\rhd_p$ (so that each node t has no more than one immediate predecessor).

We say that T is of length n if each complete path in T is of length n. Denote by T^k the set of stage k's nodes. We have $\bigcup_{k=1}^{n+1} T^k = T$. A node s is an immediate successor of t iff t is an immediate predecessor of s, that is , $s \rhd_{is} t \Longleftrightarrow t \rhd_{ip} s$. Let $F(t) = \{x : x \rhd_{is} t\}$. Let $(\rho_t)_{t \in T}$ be a collection of probability distributions, one for each node, over $F(t)$. If $F(z) = \varnothing$, we say that z is a terminal node. Denote by T^{n+1} the set of all terminal nodes. We identify T^{n+1} as the set of ultimate prizes. For any $k \in \{1, 2, ..., n\}$, we identify $t \in T^k$ as a compound lottery, starting at time k, of length $n + 1 - k$. In order to agree with other notations in the text, we write any such lottery as $Q^{n+1-k}(t)$. Finally, let Γ^l be the set of

lotteries of the following form: For all $j \neq l$, every $t \in T^j$ is a trivial node (i.e. $|F(t)| = 1$). In time l, a certain one-stage lottery is acted out.

Let $\succeq_n$ be a complete and transitive binary relation over $\mathcal{L}^n$, on which we impose the following axioms:

For any $l \in \{1, 2, .., n\}$, let Γ_q^l be the member of Γ^l with the single-stage lottery being q.

A1′ : $\forall q \in \mathcal{L}^1$ and for all $l, l' \in \{1, 2, .., n\}$, $c \sim_n \Gamma_q^{l'}$.

A2′ : Fix $t^* \in T^n$. Suppose that for all $t \in T/\{t^*\}$, $F(t)$ is the same in both Q^n and $Q^{n'}$. If Q^n yields the lottery q in t^* and $Q^{n'}$ yields the lottery q' in t^*, then $Q^n \succeq_n Q^{n'} \iff \Gamma_q^n \succeq_n \Gamma_{q'}^n$.

The implied value of any compound lottery is the following: For any $t \in T^n$, define $W^1(Q^1(t)) = V(Q^1(t))$, and recursively for $k = n-1, n-2..., 1$ and for all $t \in T^k$, let

$$W^{n+1-k}\left(Q^{n+1-k}(t)\right) = V\left(\left\langle \rho_t(s), c\left(Q^{n+1-(k+1)}(s)\right)\right\rangle_{s \in F(t)}\right)$$

where $c(Q^l(s)) \in X$ is the certaity equivalent of $c(Q^l(s))$

Lastly, and using the representation above, we extend the definition of PORU to this richer domain. Let $Q^n, Q^{n'} \in \mathcal{L}^n$ be two compound lotteries that are equal except in one sub-lottery of length $n+1-k$, $k \in \{2, 3, ..., n-1\}$ that originates from some $t^* \in T^k$. Formally, for all $t \in T$ such that $t^* \notin PRE(t)$, $F(t)$ is the same in both Q^n and $Q^{n'}$. Denote the associate (different) sub-lotteries by $Q_{Q^n}^{n+1-k}(t^*)$ and $Q_{Q^{n'}}^{n+1-k}(t^*)$, respectively. Let p be the lottery that for all $s \in F(t^*)$ gives the prize $c_V\left(Q^{n+1-(k+1)}(s)\right)$ with probability $\rho_{t^*}(s)$. Define the set $\mathcal{P}(p)$ just as before.

Definition: $\succeq_n$ display PORU if for all $Q^n, Q^{n'} \in \mathcal{L}^n$ and $p \in \mathcal{L}^1$ as described above, $W^{n+1-k}\left(Q_{Q^n}^{n+1-k}(t^*)\right) = V(p)$ and $W^{n+1-k}\left(Q_{Q^{n'}}^{n+1-k}(t^*)\right) = W(Q^2)$ for some $Q^2 \in \mathcal{P}(p)$ imply $Q^n \succeq_n Q^{n'}$.

Theorem 1': *under* **A1'**, **A2'**, **A3***, theorem 1 remains intact.*

For brevity, we omit the detailed proof. It simply involved a repeated use of **A1'**, **A2'**, and **A3** to transform the problem into the framework of theorem 1.

1.9.2 Proofs

Proof of theorem 1

Define $f_p(v) = \sum_{x \in S(p)} [u(x, v) - v] p_x$. Thus $V(p)$ is the unique solution to $f_p(v) = 0$. Note that whenever $p = \sum_i \alpha_i p^i$ we have $f_p(v) = \sum_i \alpha_i f_{p^i}(v)$. Since $f_p(v) = 0$ has a unique solution and for all $x \in (w, b)$, $u(x, V(\delta_w)) > u(w, V(\delta_w))$ and $u(b, V(\delta_b)) > u(x, V(\delta_b))$, showing that $V(p) \geqslant V(Q) \; \forall Q \in \mathcal{P}(p)$ is equivalent to showing that $f_p(V(Q)) \geqslant 0 \; \forall Q \in \mathcal{P}(p)$. To show the latter, we subtract from it $0 = \sum_i \alpha_i f_{p^i}(V(p^i))$, which does not change the expression, and regroup the terms as follows:

$$
\begin{aligned}
f_p(V(Q)) &= \sum_i \alpha_i f_{p^i}(V(Q)) \\
&= \sum_i \alpha_i \left[f_{p^i}(V(Q)) - f_{p^i}(V(p^i)) \right] \\
&= \sum_i \alpha_i \sum_{x \in S(p^i)} \left[(u(x, V(Q)) - V(Q)) - (u(x, V(p^i)) - V(p^i)) \right] p_x^i \\
&= \sum_i \alpha_i \sum_{x \in S(p^i)} \left[u(x, V(Q)) - u(x, V(p^i)) \right] p_x^i + \sum_i \alpha_i V(p^i) - V(Q) \\
&= \sum_i \alpha_i \sum_{x \in S(p^i)} u(x, V(Q)) p_x^i - \sum_i \alpha_i \sum_{x \in S(p^i)} u(x, V(p^i)) p_x^i \\
&\quad + \sum_i \alpha_i V(p^i) - V(Q) \\
&= \sum_i \alpha_i \sum_{x \in S(p^i)} u(x, V(Q)) p_x^i - \sum_i \alpha_i V(p^i) + \sum_i \alpha_i V(p^i) - V(Q) \\
&= \sum_i \alpha_i \left[\sum_{x \in S(p^i)} u(x, V(Q)) p_x^i - u(c(p^i), V(Q)) \right]
\end{aligned}
$$

Claim 1:

$$
\sum_i \alpha_i \left[\sum_{x \in S(p^i)} u(x, V(Q)) p_x^i - u(c(p^i), V(Q)) \right] \geqslant 0 \; \forall p \text{ and } \forall Q \in \mathcal{P}(p)
$$

iff

$$\forall i, \quad \left[\sum_{x \in S(p^i)} u\left(x, V(Q)\right) p_x^i - u\left(\mathrm{c}\left(p^i\right), V(Q)\right) \right] \geqslant 0 \ \forall p \text{ and } \forall Q \in \mathcal{P}\left(p\right).$$

Proof: The "if" part is obvious. For the "only if" part, assume that for some j and for some $v \neq V\left(p^j\right)$, $u\left(\mathrm{c}\left(p^j\right), v\right) - \sum_{x \in S(p^j)} u\left(x, v\right) p_x^j > 0$. Pick $y \in X$ and $\alpha \in (0,1)$ such that $V\left\langle 1, \left(\alpha \delta_y + (1-\alpha)\,\delta_{\mathrm{c}(p^j)}\right)\right\rangle = v$ (by betweenness and continuity, such y and α exist.) Let $Q = \left\langle \alpha, \delta_y; (1-\alpha), p^j \right\rangle$ (hence $V(Q) = v$). Finally , let $p := \alpha \delta_y + (1-\alpha)\,p^j$. Note that $Q \in \mathcal{P}\left(p\right)$. By construction we have

$$f_p\left(v\right) = (1-\alpha) \left[\sum_{x \in S(p^j)} u\left(x, v\right) p_x^j - u\left(\mathrm{c}\left(p^j\right), v\right) \right] < 0$$

so $V(p) < V(Q).\|$

Since p was arbitrary, we get the following necessary and sufficient condition for PORU:

$$\mathbf{C}_1: \quad \left[\sum_{x \in S(p)} u\left(x, v\right) p_x - u\left(\mathrm{c}\left(p\right), v\right) \right] \geqslant 0 \ \forall p \text{ and } \forall v \in V\left(\mathcal{L}^1\right).$$

Claim 2: $\mathbf{C}_1$ iff for every triple $x_3 > x_2 > x_1$, the indifference curve through δ_{x_2} is the steepest.

Proof: (only if): Fix $x_3 > x_2 > x_1$. By continuity, for every such triple there exists a $p \in (0,1)$ such that $p\delta_{x_3} + (1-p)\delta_{x_1} \sim_1 \delta_{x_2}$. Therefore, the vertex $(0,0)$ that represents the lottery δ_{x_2} and the point $(1-p, p)$ lie on the same indifference curve. This indifference set is of the original preferences, and hence the value attached to it is $V(p\delta_{x_3} + (1-p)\delta_{x_1}) :=$ $V(p) = pu\left(x_3, V(p)\right) + (1-p)u\left(x_1, V(p)\right) = u\left(x_2, V(p)\right)$. By $\mathbf{C}_1$, for any other v, if we pass through $(1-p, p)$ the (artificial) indifference curve corresponding to the value v, it must lie weakly above the curve from the same collection that passes through $(0,0)$. Since the betweenness property implies that indifference curves are straight lines (so their slopes are constant), the result follows.

(if): Take a lottery p with $|S(p)| = n - 1$ that belongs to an indifference set $I_v :=$ $\{p' : \sum_x u(x, v) p'_x = v\}$ in a $(n - 1)$-dimensional unit simplex $\Delta(n)$. Assume further that for some $x_v \in (w, b)$ with $x_v \notin S(p)$, $\langle 1, \delta_{x_v} \rangle \in I_v$[22]. By monotonicity and continuity,[23] p can be written as a convex combination $\alpha r + (1 - \alpha) w$, for some $\alpha \in (0, 1)$ and $r, w \in I_v$ with $|S(r)| = |S(w)| = n - 2$. By the same argument, both r and w can be written, respectively, as a convex combination of two other lotteries with size of support equal $n - 3$ and that belong to I_v. Continue in the same fashion to get an index set J and a collection of lotteries, $\{q^j\}_{j \in J}$, such that for all $j \in J$, $|S(q^j)| = 2$ and $q^j \in I_v$. Note that by monotonicity, if $y, z \in S(q^j)$ then either $z > x_v > y$ or $y > x_v > z$. By construction, for some $\alpha_1, ..., \alpha_J$ with $\alpha_j > 0$ and $\sum_j \alpha_j = 1$, $\sum_j \alpha_j q^j = p$. Let $V(q, v) := \sum_x q_x u(x, v)$. By hypothesis, $V(q^j, v') \geqslant u(x_v, v')$ for all $j \in J$ and for all $v' \in V(\Delta(n))$ and therefore also

$$
\begin{aligned}
V(p, v') &= \sum_j \alpha_j V(q^j, v') = \sum_x \sum_j \alpha_j q_x^j u(x, v') \\
&\geqslant \sum_j \alpha_j u(x_v, v') = u(x_v, v') = u(c(p), v') . \|
\end{aligned}
$$

Claim 3: NCI and $\mathbf{C}_1$ are equivalent.

Proof:

$\mathbf{C}_1 \to$NCI: Assume $p \succeq_1 \delta_x$. Using the observation that for any two lotteries p and q, $V(p) \geqslant V(q)$ is equivalent to $f_p(V(q)) \geqslant 0$, we have $\sum_i p_{x_i} u(x_i, V(p) \geqslant u(x, V(p)$. By $\mathbf{C}_1$ and monotonicity, $\sum_i p_{x_i} u(x_i, v) \geqslant u(x, v)$ for all v, and in particular for $v = V(\lambda p + (1 - \lambda)q)$.[24] Calculating the expected utility of the two lotteries $\lambda p + (1 - \lambda)q$ and $\lambda \delta_x + (1 - \lambda)q$ relative to the value $V(\lambda p + (1 - \lambda)q)$ and using again the observation above, establishes the result.

NCI$\to \mathbf{C}_1$: Suppose not. Then there exists a lottery $p \sim c(p)$ with $\left[\sum_{x \in S(p)} u(x, v) p_x - u(c(p), v) \right] < 0$ for some v. Pick $y \in X$ and $\alpha \in (0, 1)$ such that

---

[22] The analysis would be the same, though with messier notations, even if $|S(p)| = n$, i.e., if $x \in S(p)$.

[23] These two assumptions guarantee that no indifference set terminates in the relative interior of any $k \leqslant n - 1$ dimensional unit simplex.

[24] If $p \sim \delta_x$, the assertion is evident. Otherwise, we need to find p^* that is both first order stochastically dominated by p and satisfies $p^* \sim \delta_x$, and use the monotonicity of $u(, v)$ with respect to its first argument. By continuity such p^* exists.

$V(\alpha p + (1-\alpha)\delta_y) = v$. We have $v < \alpha u\left(c\left(p\right),v\right)+(1-\alpha)u\left(y,v\right) = V\left(\alpha\delta_{c(p)} + (1-\alpha)\delta_y, v\right)$, or $\alpha\delta_{c(p)} + (1-\alpha)\delta_y \succ_1 \alpha p + (1-\alpha)\delta_y$, contradicting NCI.$\|$

Note that by reversing the inequality in $\mathbf{C}_1$ and the weakly-prefer sign in NCI, we derive the analogous conditions for PGRU.$\blacksquare$

Proof of theorem 2

Since for expected utility preferences NCI is always satisfied, it is enough to demonstrate the result for lotteries with at most 3 prizes in their support.

For $x \in [w, b]$, denote by $V(\delta_x)$ the unique solution of $v = u(x, v)$. Without loss of generality, set $u(w, v) = 0$ and $u(b, v) = 1$ for all $v \in [0, 1]$. Fix $\overline{v} \in (0, 1)$. By monotonicity and continuity there exists $x\left(\overline{v}\right) \in (w, b)$ such that $\overline{v} = V\left(\delta_{x(\overline{v})}\right)$. Take any $x > x\left(\overline{v}\right)$ and note that $\mu\left(V|x, x\left(\overline{v}\right), w\right) = \left[\frac{u(x(\overline{v}),v)}{u(x,v)-u(x(\overline{v}),v)}\right]$, the slope of the indifference curves on the space $\{(p_w, p_x) \mid p_w, p_x \geq 0, \ p_w + p_x \leq 1\}$, is continuous and differentiable as a function of v on $[0, V(\delta_x)]$.

Since $\overline{v} \in (0, V(\delta_x))$, theorem 1 implies that $\mu\left(V|x, x\left(\overline{v}\right), w\right)$ is maximized at $v = \overline{v}$. A necessary condition is:

$$\frac{\partial}{\partial v}\left[\frac{u\left(x\left(\overline{v}\right),\overline{v}\right)}{u(x,\overline{v}) - u\left(x\left(\overline{v}\right),\overline{v}\right)}\right] = 0$$

Or,[25] using $\overline{v} = u\left(x\left(\overline{v}\right),\overline{v}\right)$ and denote by u_i the partial derivative of u with respect to its ith argument,

$$u_2\left(x\left(\overline{v}\right),\overline{v}\right)\left[u(x,\overline{v}) - \overline{v}\right] = \left[u_2(x,\overline{v}) - u_2\left(x\left(\overline{v}\right),\overline{v}\right)\right]\overline{v} \tag{1.1}$$

Note that by continuity and monotonicity of $u(x, v)$ in its first argument, for all $x \in \left(x\left(\overline{v}\right), b\right)$ there exists $p \in (0, 1)$ such that $p\delta_w + (1-p)\delta_x \sim_1 \delta_{x(\overline{v})}$, or $u(x, \overline{v})(1-p) = u(x\left(\overline{v}\right),\overline{v}) = \overline{v}$. Therefore, and using again theorem 1, (1) is an identity for $x \in \left(x\left(\overline{v}\right), b\right)$, so we can take the partial derivative of both sides with respect to x and maintain equality.

---

[25] second order conditions would be :

$$\frac{u_{22}\left(x\left(\overline{v}\right),\overline{v}\right)}{u_{22}\left(x,\overline{v}\right)} < \frac{\overline{v}}{u\left(x,\overline{v}\right)} \ (< 1)$$

We get:

$$u_2\left(x\left(\overline{v}\right),\overline{v}\right)u_1(x,\overline{v}) = u_{21}(x,\overline{v})\overline{v}$$

Since u is strictly increasing in its first argument, $u_1(x,\overline{v}) > 0$ and $\overline{v} > 0$. Thus: $\frac{u_{21}(x,\overline{v})}{u_1(x,\overline{v})} = \frac{u_2(x(\overline{v}),\overline{v})}{\overline{v}} = l\left(\overline{v}\right)$ independent of x, or by changing order of differentiation: $\frac{\partial}{\partial v}\left[\ln u_1(x,\overline{v})\right]$ is independent of x.

Since $\overline{v}$ was arbitrary, we have the following differential equation on $\{(x,v) \mid v < u(x,v)\}$:

$$\frac{\partial}{\partial v}\left[\ln u_1(x,v)\right] = l\left(v\right)$$

By the fundamental theorem of calculus, the solution of this equation is:

$$
\begin{aligned}
\frac{\partial}{\partial v}\left[\ln u_1(x,v)\right] &= l\left(v\right) \\
\implies \ln u_1(x,v) &= \ln u_1(x,0) + \int_{s=0}^{v} l\left(s\right) ds \\
\implies u_1(x,v) &= u_1(x,0)\exp\left(\int_{s=0}^{v} l\left(s\right) ds\right) \\
\implies u(x,v) - u(x\left(v\right),v) &= \exp\left(\int_{s=0}^{v} l\left(s\right) ds\right)\int_{x(v)}^{x} u_1(t,0)dt \\
\implies u(x,v) - v &= \exp\left(\int_{s=0}^{v} l\left(s\right) ds\right)(u(x,0) - u(x\left(v\right),0))
\end{aligned}
$$

Note that the term

$$\exp\left(\int_{s=0}^{v} l\left(s\right) ds\right) = \exp\left(\int_{s=0}^{v}\frac{u_2\left(x\left(s\right),s\right)}{s}ds\right)$$

is well defined since by the assumption that all derivatives are continuous and bounded and that $u_1 > 0$, we use L'Hopital's rule and implicit differentiation to show that the term

$$
\begin{aligned}
\lim_{s\to 0}\frac{u_2\left(x\left(s\right),s\right)}{s} &= \lim_{s\to 0} u_{21}\left(x\left(s\right),s\right)x'\left(s\right) + u_{21}\left(x\left(s\right),s\right) \\
&= \lim_{s\to 0} u_{21}\left(x\left(s\right),s\right)\frac{1 - u_2\left(x\left(s\right),s\right)}{u_1\left(x\left(s\right),s\right)} + u_{21}\left(x\left(s\right),s\right)
\end{aligned}
$$

is finite and hence $\left(\int_{s=0}^{v}\frac{u_2(x(s),s)}{s}ds\right)$ is finite as well.

To uncover $u(x,v)$ on the region $\{(x,v) \mid v > u(x,v)\}$, fix again some $\overline{v} \in (0,1)$ and the corresponding $x\left(\overline{v}\right) \in (w,b)$ (with $\overline{v} = u(x\left(\overline{v}\right),\overline{v})$). Take any $x < x\left(\overline{v}\right)$ and note

that $\widehat{\mu}\left(V|b, x\left(\overline{v}\right), x\right) = \left[\frac{u(x(\overline{v}),v)-u(x,v)}{1-u(x(\overline{v}),v)}\right]$, the slope of the indifference curves on the space $\{(p_x, p_b)\mid p_x, p_b \geq 0, \ p_x + p_b \leq 1\}$, is continuous and differentiable as a function of v on $[V(\delta_x), b]$.

Since $\overline{v} \in (V(\delta_x), b)$, by using theorem 1 we have:

$$\frac{\partial}{\partial v}\left[\frac{u(x\left(\overline{v}\right), \overline{v}) - u(x, \overline{v})}{1 - u(x\left(\overline{v}\right), \overline{v})}\right] = 0$$

or,

$$(u_2\left(x\left(\overline{v}\right), \overline{v}\right) - u_2\left(x, \overline{v}\right))\left[1 - \overline{v}\right] = -u_2(x\left(\overline{v}\right), \overline{v})\left[\overline{v} - u(x, \overline{v})\right] \tag{1.2}$$

Using the same argumentation from the former case, (2) holds for all $x \in (w, x\left(\overline{v}\right))$, so we can take the partial derivative of both sides with respect to x and maintain equality. We get:

$$-u_{21}\left(x, \overline{v}\right)\left[1 - \overline{v}\right] = u_1(x, \overline{v})u_2(x\left(\overline{v}\right), \overline{v})$$

Since u is strictly increasing in its first argument, $u_1(x, \overline{v}) > 0$ and $1 - \overline{v} > 0$. Thus: $\frac{u_{21}(x,\overline{v})}{u_1(x,\overline{v})} = -\frac{u_2(x(\overline{v}),\overline{v})}{[1-\overline{v}]} = k\left(\overline{v}\right)$ independent of x, or by changing order of differentiation: $\frac{\partial}{\partial v}\left[\ln u_1(x, \overline{v})\right]$ is independent of x.

Since $\overline{v}$ was arbitrary, we have the following differential equation on $\{(x, v)\mid v > u(x, v)\}$:

$$\frac{\partial}{\partial v}\left[\ln u_1(x, v)\right] = k\left(v\right)$$

Its solution is given by

$$\frac{\partial}{\partial v}\left[\ln u_1(x, v)\right] = k\left(v\right)$$

$$\Longrightarrow \quad \ln u_1(x, 1) - \ln u_1(x, v) = \int_{s=v}^{1} k\left(s\right) ds$$

$$\Longrightarrow \quad \ln u_1(x, v) = \ln u_1(x, 1) - \int_{s=v}^{1} k\left(s\right) ds$$

$$\Longrightarrow \quad u_1(x, v) = u_1(x, 1) \exp\left(\int_{s=v}^{1} k\left(s\right) ds\right)^{-1}$$

$$\Longrightarrow \quad u(x, v) - u(x\left(v\right), v) = \exp\left(\int_{s=v}^{1} k\left(s\right) ds\right)^{-1} \int_{x}^{x(v)} u_1(t, 1) dt$$

$$\Longrightarrow \quad u(x, v) - v = -\left[u(x\left(v\right), 1) - u(x, 1)\right] \exp\left(\int_{s=v}^{1} k\left(s\right) ds\right)^{-1}$$

42

which is again well defined since

$$\exp\left(\int_{s=v}^{1} k\left(s\right) ds\right) = \exp\left(\int_{s=v}^{1} -\frac{u_2\left(x\left(s\right),s\right)}{\left[1-s\right]} ds\right)$$

and

$$\begin{aligned}
\lim_{s\to 1} -\frac{u_2\left(x\left(s\right),s\right)}{\left[1-s\right]} &= \lim_{s\to 1} u_{21}\left(x\left(s\right),s\right) x'\left(s\right) + u_{21}\left(x\left(s\right),s\right) \\
&= \lim_{s\to 1} u_{21}\left(x\left(s\right),s\right) \frac{1-u_2\left(x\left(s\right),s\right)}{u_1\left(x\left(s\right),s\right)} + u_{21}\left(x\left(s\right),s\right)
\end{aligned}$$

is finite, and hence the whole integral is finite.

So far we have:

$$u(x,v) - v = \begin{cases} \left[u(x,0) - u(x\left(v\right),0)\right] \exp\left(\int_{s=0}^{v} \frac{u_2(x(s),s)}{s} ds\right) & x > x\left(v\right) \\ -\left[u(x\left(v\right),1) - u(x,1)\right] \left(\exp\left(\int_{s=v}^{1} -\frac{u_2(x(s),s)}{[1-s]} ds\right)\right)^{-1} & x < x\left(v\right) \end{cases} \quad (1.3)$$

We add the following restrictions:

(i) $u(b,v) = 1$ for all $v \in [0,1]$, which implies:

$$\left[1 - u(x\left(v\right),0)\right] \exp\left(\int_{s=0}^{v} \frac{u_2\left(x\left(s\right),s\right)}{s} ds\right) = 1 - v$$

(ii) $u(w,v) = 0$ for all $v \in [0,1]$, which implies:

$$u(x\left(v\right),1) \left(\exp\left(\int_{s=v}^{1} -\frac{u_2\left(x\left(s\right),s\right)}{\left[1-s\right]} ds\right)\right)^{-1} = v$$

Substituting into (3) to get:

$$u(x,v) - v = \begin{cases} \left[u(x,0) - u(x\left(v\right),0)\right] \frac{1-v}{\left[1-u(x(v),0)\right]} & x > x\left(v\right) \\ -\left[u(x\left(v\right),1) - u(x,1)\right] \frac{v}{u(x(v),1)} & x < x\left(v\right) \end{cases} \quad (1.4)$$

We further require:

(iii) Continuity at $x = x(v)$. This is immediate since

$$\lim_{x \to -x(v)} (u(x,v) - v) = \lim_{x \to +x(v)} (u(x,v) - v) = 0$$

(iv) Differentiability at $x(v)$ for all v:

$$u_1(x(v),0)\frac{1-v}{[1-u(x(v),0)]} = u_1(x(v),1)\frac{v}{u(x(v),1)}$$

or

$$\frac{u_1(x(v),1)}{u_1(x(v),0)} = \frac{[1-u(x(v),v)]\,u(x(v),1)}{[1-u(x(v),0)]\,u(x(v),v)} \tag{1.5}$$

Let $r(x,v) := \frac{-u_{11}(x,v)}{u_1(x,v)}$. Given $v \in (0,1)$, note that

$$r(x,v) = \begin{cases} -\frac{u_{11}(x,0)}{u_1(x,0)} & x > x(v) \\ -\frac{u_{11}(x,1)}{u_1(x,1)} & x < x(v) \end{cases}$$

But since u is continuous and $r(x,v)$ is well defined, $r(x,v)$ must be continuous as well. Therefore, we require:

$$-\frac{u_{11}(x(v),0)}{u_1(x(v),0)} = -\frac{u_{11}(x(v),1)}{u_1(x(v),1)}$$

and since this is true for any v and the function $x(v)$ is onto, we have for all $x \in (w,b)$:

$$-\frac{u_{11}(x,0)}{u_1(x,0)} = -\frac{u_{11}(x,1)}{u_1(x,1)}$$

which implies that for some a and b, $u(x,1) = au(x,0) + b$. But $u(0,1) = u(0,0) = 0$ and $u(1,1) = u(1,0) = 1$, hence, by continuity, $b = 0$ and $a = 1$, or $u(x,1) = u(x,0) := z(x)$ for all $x \in [w,b]$. Plug into (4) to get:

$$u(x,v) - v = \begin{cases} [z(x) - z(x(v))]\frac{1-v}{[1-z(x(v))]} & x > x(v) \\ -[z(x(v)) - z(x)]\frac{v}{z(x(v))} & x < x(v) \end{cases} \tag{1.6}$$

44

and into (5) to get:

$$\frac{u_1(z\,(x))}{u_1(z\,(x))} = 1 = \frac{[1-v]}{[1-z\,(x\,(v))]}\frac{z\,(x\,(v))}{v}$$

or

$$\frac{v}{z\,(x\,(v))} = \frac{[1-v]}{[1-z\,(x\,(v))]} := m\,(v) \qquad (1.7)$$

Substituting (7) into (6) we have:

$$u(x,v) - v = [z\,(x) - z\,(x\,(v))]\,m\,(v) \qquad (1.8)$$

and using the boundary conditions, (i) and (ii), again we find that

$$u(w,v) - v = 0 - v = [0 - z\,(x\,(v))]\,m\,(v)$$

or

$$v - z\,(x\,(v))\,m\,(v) = 0 \qquad (1.9)$$

and

$$u(b,v) - v = 1 - v = [1 - z\,(x\,(v))]\,m\,(v)$$

or

$$1 = m\,(v) + v - z\,(x\,(v))\,m\,(v) = m\,(v) \qquad (1.10)$$

where the second equality is implied by (9). Therefore $m\,(v) = 1$ and using (7) and (8) we have

$$u(x,v) = z\,(x)$$

which implies that the local utility function is independent of v, hence preferences are expected utility.$\blacksquare$

Proof of the sufficient conditions for mixed fan

Note that in the two-dimensional probability simplex, an indifference set is defined by

$v = p_{x_1} u(x_1, v) + (1 - p_{x_1} - p_{x_3}) u(x_2, v) + p_{x_3} u(x_3, v)$. The slope of an indifference curve is then given by: $\frac{\partial p_3}{\partial p_1} = \frac{u(x_2, v) - u(x_1, v)}{u(x_3, v) - u(x_2, v)}$. Using theorem 1, it is evident that the requirement $\sum u(x, v) p_x \geqslant u(c(p), v)$ is equivalent to having the indifference curve through the $(0, 0)$ vertex being the steepest. Denote by v_i the solution to $\phi(x_i, v) - v = 0$. By monotonicity, for any triple $x_3 > x_2 > x_1$, and for any $v \in (v_1, v_3)$, $V^{-1}(v) \in L(x_3)$ and $V^{-1}(v) \notin L(x_1)$ but for the middle prize, x_2, both are possible. Let u_i denotes the partial derivative of u with respect to its ith argument. The derivative (with respect to v) at $\hat{v}$ of $\ln\left(\frac{dp_{x_3}}{dp_{x_1}}\right)$ is given by $\frac{[u_2(x_2, \hat{v}) - u_2(x_1, \hat{v})]}{u(x_2, \hat{v}) - u(x_1, \hat{v})} - \frac{[u_2(x_3, \hat{v}) - u_2(x_2, \hat{v})]}{u(x_3, \hat{v}) - u(x_2, \hat{v})}$. By assumptions (1)-(3), if $V^{-1}(\hat{v}) \in L(x_2)$, this term is positive ("fanning out") whereas If $V^{-1}(\hat{v}) \notin L(x_2)$, it is negative ("fanning in"). In particular, the indifference curve in the level $v = \phi^1(x_2, v) = \phi^2(x_2, v)$ is the steepest.∎

Proof of proposition 2

Let $\triangle V(\beta \,|p, n) := \text{grp}(\beta \,|p, n)$, anb for $k = 2, 3, ..., n - 1$, denote $\triangle V(\beta \,|p, n)$ with $h(\beta \,|p, n) = n - k$ by $\triangle V^{(k)}(\beta \,|p, n)$. It can be shown that

$$
\begin{aligned}
&\triangle V^{(k)}(\beta \,|p, n) \\
&= np\beta(1 - p) \frac{(1 - p)^{k-1}\left(+\beta\left(\sum_{j=0}^{n-(k+1)} \binom{j+k-2}{j} p^j\right) - p^{n-k}\left(\binom{n-2}{n-(k+1)}\beta - \binom{n-1}{n-k}\right)\right) + 1}{(1 + \beta(1 - p))\left(\beta\left(\sum_{j=k-1}^{n-1} \binom{j}{j-(k-1)} p^{j-(k-1)}\right)(1 - p)^k + 1\right)}
\end{aligned}
$$

The denominator of $\triangle V^{(k)}(\beta \,|p, n)$ is always positive, whereas the coefficient $np\beta(1 - p)$ is strictly positive for $\beta > 0$. At $\beta = 0$ the nominator is equal to $1 - \binom{n-1}{n-k}(1 - p)^{k-1} p^{n-k}$ which is positive since $\binom{n-1}{n-k}(1 - p)^{k-1} p^{n-k}$ is simply the probability of $n - k$ successes in $n - 1$ trials of a Bernoulli random variable with parameter p. We then note that the nominator is also increasing with β. Indeed, this is the case if $\left(\sum_{j=0}^{n-(k+1)} \binom{j+k-2}{j} p^j\right) > p^{n-k}\binom{n-2}{n-(k+1)}$, which is true since $p < 1$ and $\sum_{j=0}^{n-(k+1)} \binom{j+k-2}{j} = \binom{n-2}{n-k-1}$. Therefore, item (i) is implied. Since $\beta = 0$ implies expected utility, the first part of item (ii) is immidiate. For the second part of item (ii), observe that as β increases, the value of the sequential lottery $(V(Q^n))$ is (smoothly) strictly decreasing and converges to 0, the value of the worst prize in its support. The value of the one stage lottery $(V(\hat{p}))$ is affected in two ways when

β increases: First, given a threshold $h(\beta \,|p, n)$, the value is (smootly) strictly decreasing with β. Second, $h(\beta \,|p, n)$ itself is a decreasing step-function of β. For β large enough, all prizes but 0 are elated and the value of the lottery is given by $\frac{\sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} k}{1 + \beta (1-p)^n} \xrightarrow[\beta \to \infty]{} 0$.

To show the existence of β^* (item (iii)), pick $\beta' > 0$ such that $\text{grp}(\beta' \,|p, n) = \epsilon > 0$. Since $\lim\limits_{\beta \to \infty} \text{grp}(\beta \,|p, n) = 0$, there exists $\overline{\beta} := \max\{\beta \,|\text{grp}(\beta \,|p, n) = \frac{\epsilon}{2}\}$ and $\overline{\beta} < \infty$. Thus $\text{grp}(\beta \,|p, n)$ is a continuous function on the compact interval $[0, \overline{\beta}]$, and hence achieves its maximum on this domain. For single-peakness, we have the following two claims:

Claim 1: $\forall k = 2, 3, ..., n-1$, $\triangle V^{(k)}(\beta \,|p, n)$ *is either strictly increasing or single-peaked on* $(0, \infty)$.

Proof: By differentiating $\triangle V^{(k)}(\beta \,|p, n)$ with respect to β, one gets:

$$\frac{\partial}{\partial \beta} \triangle V^{(k)}(\beta \,|p, n)$$
$$= np \frac{C\beta^2 + \left(2Ap^k(1-p)^k - 2p^n \binom{n-2}{n-k-1}(1-p)^k\right)\beta + \left((1-p)p^k - p^n \binom{n-1}{n-k}(1-p)^k\right)}{p^k(-\beta + p\beta - 1)^2 \left(B\beta(-p+1)^k + 1\right)^2}$$

Where C is some constant, and $A := \left(\sum_{j=0}^{n-(k+1)} \binom{j+k-2}{j} p^j\right)$.

The roots of $\frac{\partial}{\partial \beta} \triangle V^{(k)}(\beta \,|p, n)$ are the roots of the second-degree polynomial in β that appears in the nominator.

Evaluated at $\beta = 0$, this polynomial is equal to $\left(p^k - pp^k - p^n \binom{n-1}{n-k}(1-p)^k\right)$. Note that

$$\left(p^k - pp^k - p^n \binom{n-1}{n-k}(1-p)^k\right) > 0 \iff 1 > \binom{n-1}{n-k} p^{n-k}(1-p)^{k-1}$$

which is true as claimed before. In addition, the slope of that polynomial at $\beta = 0$ is equal to the coefficient of β, $2Ap^k(1-p)^k - 2p^n \binom{n-2}{n-k-1}(1-p)^k$, which is positive since $\left(\sum_{j=0}^{n-(k+1)} \binom{j+k-2}{j} p^j\right) > p^{n-k} \binom{n-2}{n-k-1}$.

To summarize, both the slope and the intercept of the polynomial in the nominator are positive at $\beta = 0$. Therefore, if $C \geq 0$ then $\frac{\partial}{\partial \beta} \triangle V^{(k)}(\beta \,|p, n)$ has no positive roots, and otherwise it has exactly one positive root.$\|$

Note that $\triangle V(\beta \,|p, n)$ is a continuous function that is not differentiable in the points

47

where $h(\beta | p, n)$ changes. For $k = 2, 3, ..., n - 1$, let $\beta_{k,k+1}$ be the value of β where $h(\beta | p, n)$ decreases from $(n - k)$ to $(n - (k + 1))$. Using the same notations as above, we claim that at the switch point, the slope of the resolution premium decreases.

Claim 2: $\lim\limits_{\beta \to -\beta_{k,k+1}} \frac{\partial}{\partial \beta} \triangle V^{(k)}(\beta | p, n) > \lim\limits_{\beta \to +\beta_{k,k+1}} \frac{\partial}{\partial \beta} \triangle V^{(k+1)}(\beta | p, n)$

Proof: Apart from at $\beta = 0$, where $\triangle V^{(k)}(0 | p, n) = \triangle V^{(k+1)}(0 | p, n) = 0$, it can be shown that the two curves cross at exactly one more point, given by

$$\beta_{k,k+1} = \frac{np - (n - k)}{\left(\sum_{j=0}^{n-(k+1)} (n - k - j) \binom{j+k-1}{j} p^j \right) (1 - p)^{k+1}}$$

Note that $\beta_{k,k+1} > 0$ iff $p > \frac{n-k}{n}$. To prove the claim it will be sufficient to show that $\frac{\partial}{\partial \beta} \triangle V^{(k)}(0 | p, n) < \frac{\partial}{\partial \beta} \triangle V^{(k+1)}(0 | p, n)$, since this implies that at $\beta_{k,k+1}$, $\triangle V^{(k+1)}(\beta | p, n)$ crosses $\triangle V^{(k)}(\beta | p, n)$ from above. Now:

$\frac{\partial}{\partial \beta} \triangle V^{(k)}(0 | p, n) = np \frac{\left(p^k - pp^k - p^n \binom{n-1}{n-k}(1-p)^k \right)}{p^k}$ and

$\frac{\partial}{\partial \beta} \triangle V^{(k+1)}(0 | p, n) = np \frac{\left(p^{k+1} - pp^{k+1} - p^n \binom{n-1}{n-k-1}(1-p)^{k+1} \right)}{p^{k+1}}$

so $\quad \frac{\partial}{\partial \beta} \triangle V^{(k+1)}(0 | p, n) > \frac{\partial}{\partial \beta} \triangle V^{(k)}(0 | p, n)$

$\Longleftrightarrow \frac{1}{p^k} n (-p + 1)^k p^n \left(p \binom{n-1}{-k+n} + p \binom{n-1}{-k+n-1} - \binom{n-1}{-k+n-1} \right) > 0$

$\Longleftrightarrow p \binom{n-1}{-k+n} + p \binom{n-1}{-k+n-1} - \binom{n-1}{-k+n-1} > 0$

$\Longleftrightarrow p > \frac{\binom{n-1}{-k+n-1}}{\binom{n-1}{-k+n} + \binom{n-1}{-k+n-1}} = \frac{(n-k)}{n}. \parallel$

To complete the proof we verify that both claims above are also valid for the two extreme cases, $k = 1$ (where only the best prize, n is elation) and $k = n$ (only the worst prize, 0 is disappointment).

$k = 1$: Using the same notation as used above we have:

$$\triangle V^{(1)}(\beta | p, n) = np\beta \left(\sum_{j=0}^{n-2} p^j \right) (p - 1)^2 \frac{\beta + 1}{(1 + (1 - p)\beta)(1 + (1 - p^n)\beta)}$$

48

and

$$\frac{\partial}{\partial\beta}\triangle V^{(1)}\left(\beta\,|p,n\right)=n\left(1-p\right)\left(p-p^n\right)\frac{\left(1-pp^n\right)\beta^2+2\beta+1}{\left(-\beta+p^n\beta-1\right)^2\left(-\beta+p\beta-1\right)^2}>0$$

for all $\beta\geq 0$ so $\triangle V^{(1)}\left(\beta\,|p,n\right)$ is strictly increasing with β (claim 1).

For the second claim, similar calculations as above establish that:

$$\frac{\partial}{\partial\beta}\triangle V^{(2)}\left(0\,|p,n\right)>\frac{\partial}{\partial\beta}\triangle V^{(1)}\left(0\,|p,n\right)\iff p>\frac{n-1}{n}$$

so claim 2 follows as well.

$k=n$:

$$\triangle V^{(n)}\left(\beta\,|p,n\right)=np^2\beta\left(1-p\right)\frac{\left(\sum\limits_{j=1}^{n-1}\binom{n-1}{j}p^{j-1}\left(-1\right)^{j-1}\right)}{\left(1+\beta\left(1-p\right)\right)\left(1+\beta\left(1-p\right)^n\right)}$$

Let $C=\left(\sum\limits_{j=1}^{n-1}\binom{n-1}{j}p^{j-1}\left(-1\right)^{j-1}\right)$, so:

$$\frac{\partial}{\partial\beta}\triangle V^{(n)}\left(\beta\,|p,n\right)=Cnp^2\left(p-1\right)\frac{\beta^2\left(1-p\right)^{n+1}-1}{\left(\beta\left(-p+1\right)^n+1\right)^2\left(-\beta+p\beta-1\right)^2}$$

which is clearly single peaked *on* $(0,\infty)$ (claim 1), and, again by similar calculations:

$$\frac{\partial}{\partial\beta}\triangle V^{(n)}\left(0\,|p,n\right)>\frac{\partial}{\partial\beta}\triangle V^{(n-1)}\left(0\,|p,n\right)\iff p>\frac{1}{n}$$

which is claim 2.

Combining claim 1 and claim 2 ensures that $\triangle V\left(\beta\,|p,n\right)$ is single-peaked on $(0,\infty)$.∎

Proof of proposition 3

We first show that the claim is true for any lotteries of the form $p\delta_x+\left(1-p\right)\delta_y$, with $x>y$.

Case 1, $p=0.5$:

Construct the compund lottery $Q^n\in\mathcal{P}\left(0.5\delta_x+0.5\delta_y\right)$ as follows:

In each period $\Pr\left(\text{"success"}\right)=\Pr\left(\text{"failure"}\right)=0.5$.

Define:

$$z_i = \begin{cases} 1 & [\text{"success"}] \\ 0 & [\text{"failure"}] \end{cases} \quad i = 1, 2, 3, ..$$

The terminal nodes are:

$$\delta_x \qquad \text{if } \sum_{i=1}^n z_i > \frac{n}{2}$$
$$0.5\delta_x + 0.5\delta_y \quad \text{if } \sum_{i=1}^n z_i = \frac{n}{2}$$
$$\delta_y \qquad \text{if } \sum_{i=1}^n z_i < \frac{n}{2}$$

Claim:

$$\lim_{n \to \infty} V(Q^n) = V(\delta_y) = \phi(y)$$

Proof of claim: We use the fact that Value of the lottery using recursive Gul preferences and probability 0.5 for "success" in each period is equal to the value of the lottery using recursive expected utility and probability $\frac{0.5}{1+\beta 0.5}$ for "success" in each period.

Since $z_i's$ are i.i.d random variables, the weak law of large numbers implies:

$$\frac{\sum_{i=1}^n z_i}{n} \xrightarrow{p} \frac{0.5}{1 + \beta 0.5} < 0.5$$

or,

$$\Pr\left(\sum_{i=1}^n z_i < \frac{n}{2}\right) \to 1$$

Therefore

$$\begin{aligned} V(Q^n) &= \phi(x) \Pr\left(\sum_{i=1}^n z_i > \frac{n}{2}\right) + \\ &\quad \frac{0.5\phi(x) + (1 + \beta) 0.5\phi(y)}{1 + \beta 0.5} \Pr\left(\sum_{i=1}^n z_i = \frac{n}{2}\right) + \\ &\quad \phi(y) \Pr\left(\sum_{i=1}^n z_i < \frac{n}{2}\right) \to \phi(y) \end{aligned}$$

case 2, $p < 0.5$:

Take $Q^{n+1} = \langle 2p, Q^n; 1 - 2p, \delta_y \rangle$, with Q^n as defined above.

case 3, $p > 0.5$:

Fix $\varepsilon > 0$. Using the construction in case 1, obtain Q^{T_1} with $V\left(Q^{T_1}\right) \in \left(\phi\left(y\right), \phi\left(y\right) + \frac{\varepsilon}{2}\right)$. Re-construct a lottery as above, but replace δ_y with Q^{T_1} in the terminal node. By the same argument, there exists T_2 and $V\left(Q^{T_1+T_2}\right) \in \left(\phi\left(y\right), \phi\left(y\right) + \varepsilon\right)$. Note that the underlying probability of y in $Q^{T_1+T_2}$ is 0.25. Therefore, by monotonicity, the construction works for any $p < 0.75$. Repeat in the same fashion to show that the assertion is true for $p^k < \frac{3+4k}{4+4k}$, $k = 1, 2, ..$, and note that $p^k \to 1.\|$

Now take any finite lottery $\sum_{j=1}^m p_j \delta_{x_j}$ and order its prizes as $x_1 < x_2 < ... < x_m$. Repeat the construction above for the binary lottery x_{m-1}, x_m to make its value arbitrarily close to $\phi\left(x_{m-1}\right)$. Then mix it appropriately with x_{m-2} and repeat the argument above. Continue in this fashion to get a multi-stage lottery over $x_2, ..., x_m$ with a value arbitrarily close to $\phi\left(x_2\right)$. Conclude by mixing it with x_1 and repeat the construction above.∎

Proof of proposition 4

It is obvious that (i) is necessary for (ii). To show sufficiency, we introduce the intermediate lotteries Q and p^j, where the compound lottery Q assigns probability $\alpha_j(\pi)$ to p^j, and p^j assigns probability p_s^j to the outcome $u\left(a^*\left(s\right), s\right)$. Clearly, since for each state s and for any action a we have $u\left(a, s\right) \leq u\left(a^*\left(s\right), s\right)$, by monotonicity of the value of a lottery with respect to the relation of first-order stochastic dominance, $V(p^{j*}) \leq V(p^j)$, and hence, by the same reason, also $V^p(\pi) \leq V(Q)$.

However, now Q is simply the folding back of the two-stage lottery, which when played in one-shot is the lottery corresponding to full information structure, I. Thus by (i) we have that $V^p(I) \geq V(Q)$. Combining the two inequalities establishes the result.

Similarly, it is obvious that PGRU is necessary for ϕ being the least valuable information structure. To show sufficiency, define $V(a, p)$ as the value of a lottery in which with probability p_s you get the outcome $u\left(a, s\right)$. Let $\underline{a} = \arg\max_a V(a, p)$, then $V^p(\phi) = V(\underline{a}, p)$. Let Q be a two-stage lottery that assigns probability $\alpha_j(\pi)$ to p^j and p^j assigns probability p_s^j to the outcome $u(\underline{a}, s)$. By definition, $V(p^j) \leq V(p^{j*})$ for all j, and therefore, by monotonicity, $V(Q) \leq V^p(\pi)$. However, now Q is simply the folding back of the two-stage

lottery, which when played in one-shot is the lottery corresponding to ϕ. Thus by (i) we have that $V^p(\phi) \leq V(Q)$. Combining the two inequalities establishes the result.∎

Bibliography

[1] Allais, M., (1953), "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine." *Econometrica,* 21, 503-546.

[2] Ang, A., Bekaert,G. and J. Liu., (2005), "Why Stocks May Disappoint." *Journal of Financial Economics*, 76, 471-508.

[3] Arrow, K.J., and C. Fischer, (1974), "Environmental Preservation, Uncertainty and Irreversibility." *Quarterly Journal of Economics,* 88, 312-19.

[4] Artstein-Avidan, S. and D. Dillenberger, (2006), "Don't tell me anything until you know for sure." mimeo.

[5] Barberis, N. and M. Huang, (2005), "The Loss Aversion / Narrow Framing Approach to the Equity Premium Puzzle." NBER Working Paper No. 12378.

[6] Barberis, N., Huang, M. and R. H. Thaler, (2006), "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing." *American Economic Review*, 96, 1069-1090.

[7] Bellemare, C., Krause, M., Kröger, S.and C. Zheng, (2005), "Myopic Loss Aversion: Information Feedback vs. Investment Flexibility." *Economics Letters*, 87, 319-324.

[8] Benartzi, S. and R.Thaler, (1995), "Myopic loss aversion and the equity premium puzzle." *Quarterly Journal of Economics,* 110, 73–92.

[9] Blackwell, D., (1953), "Equivalent Comparison of Experiments." *Annals of Mathematics and Statistics,* 24, 265-272.

[10] Chew, S.H., (1989), "Axiomatic Utility Theories with the Betweenness Property." *Annals of Operations Research*, 19, 273-298.

[11] Chew, S.H. and L.G. Epstein, (1989), "The Structure of Preferences and Attitudes Towards the Timing of The Resolution of Uncertainty." *International Economic Review,* 30, 103-117.

[12] Conlisk, J., (1989), "Three Variations on the Allais Example." *The American Economic Review*, 79, 392-407.

[13] Debreu, G., (1972), "Smooth Preferences." *Econometrica*, 40, 603-615.

[14] Dekel, E., (1986), "An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom." *Journal of Economic Theory,* 40, 304-318.

[15] Dixit, A.K., and R.S. Pindyck, (1994), " Investment Under Uncertainty." Princeton: Princeton Univ. Press.

[16] Epstein, L.G. and S.E. Zin, (1989), "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework." *Econometrica,* 57, 937-969.

[17] Gollier, C., (2001), "The Economics of Risk and Time." The MIT Press, 0-262-07215-7.

[18] Gneezy, U., J. Potters, (1997), "An experiment on risk taking and evaluation periods." *Quarterly Journal of Economics,* 112, 632–645.

[19] Grant,S., Kajii A., and B. Polak, (1998), "Intrinsic Preference for Information." *Journal of Economic Theory*, 83, 233-259.

[20] Grant,S., Kajii A., and B. Polak, (2000), "Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models." *Econometrica*, 68, 425-434.

[21] Gul, F., (1991), "A Theory of Disappointment Aversion." *Econometrica*, 59, 667-686.

[22] Haigh, M.S. and J.A. List, (2005), "Do professional traders exhibit myopic loss aversion? An experimental analysis." *The Journal of Finance*, 60, 523-534.

[23] Kahneman, D. and A.Tversky, (1979), "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, 47, 263-291.

[24] Knight, F., (1921), "Risk, Uncertainty and Profit." Boston, Mass: Houghton Mifflin. Reprint, London: London School of Economics, 1946.

[25] Köszegi, B. and M. Rabin, (2007), "A Model of Reference-Dependent Preferences", forthcoming in *Quarterly Journal of Economics*.

[26] Kreps, D. M. and E.L. Porteus, (1978), "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica,* 46, 185-200.

[27] Machina, M.J., (1982), "'Expected Utility' Analysis without the Independence Axiom." *Econometrica*, 50, 277-323.

[28] Mcdonald, R. and D.Siegel, (1986), "The Value of Waiting to Invest." *Quarterly Journal of Economics,* 101, 707-28.

[29] Mehra, R. and E.C. Prescott, (1985), "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, 15, 145-161.

[30] Nayyar, S., (2004), "Disappointment Aversion and the Insurance Trap." mimeo.

[31] Nehring, K., (2005), "Notes on Expected Utility with Psychological Consequences: The Logic of Betweenness-Based Risk Preferences." mimeo.

[32] Neilson, W.S., (1992), "A Mixed Fan Hypothesis and its Implications for Behavior Towards Risk."*Journal of Economic Behavior and Organization,* 19, 197-211.

[33] Palacious-Huerta, I., (1999), "The Aversion to the Sequential Resolution of Uncertainty." *Journal of Risk and Uncertainty,*18, 249-269.

[34] Pindyck, R., (1991), "Irreversibility, Uncertainty and Investment." *Journal of Economic Literature,* 29, 1110-48.

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

[36] Rabin, M., (2000), "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica,* 68, 1281–1292.

[37] Rabin, M. and R. H. Thaler, (2001), "Anomalies: Risk Aversion." *Journal of Economic Perspectives*, 15, 219–232.

[38] Routledge, B. and S.E. Zin, (2004), "Generalized Disappointment Aversion and Asset prices." mimeo.

[39] Safra, Z. and E. Sulganik, (1995), "On the Nonexistence of Blackwell's Theorem-Type Results with General Preference Relations." *Journal of Risk and Uncertainty*, 10, 187-201.

[40] Safra, Z. and U. Segal, (2006), "Calibration Results for Non-Expected Utility Theories." working paper.

[41] Schlee, E., (1990), "The Value of Information in Anticipated Utility Theory." *Journal of Risk and Uncertainty,* 3, 83–92.

[42] Segal, U., (1990), "Two-stage Lotteries without the Reduction Axiom." *Econometrica,* 58, 349-377.

[43] Segal, U.and A. Spivak, (1990), "First order versus second order risk aversion." *Journal of Economic Theory*, 51, 111-125.

Chapter 2

Ashamed to be Selfish (with Philipp Sadowski)

2.1 Introduction

2.1.1 Motivation

The notions of fairness and altruism have attracted the attention of economists in different contexts. The relevance of these motives to decision making is both intuitively convincing and well documented. For example in a classic "dictator game," where one person gets to anonymously divide, say, $10 between herself and a partner, people tend not to take the whole amount for themselves, but to give a sum of between $0 and $5 to the other player. They act as if they are trading off a concern for fairness or for the other person's incremental wealth and a concern for their own.[1] Thus, preferences for fairness as well as preferences for altruism have been suggested and considered (for example Fehr and Schmidt [1999], Anderoni and Miller [2002], and Charness and Rabin [2002]).

Recent experiments, however, show that this interpretation may be rash: Dana, Cain and Dawes (2006) study a variant of the same dictator game, where the dictator is given the option to exit the game before the recipient learns it is being played. If she opts out, she is given a specified amount of money and the recipient gets nothing, as the game has not taken place. It turns out that about a third of the participants choose to leave the game when offered $9 for themselves and $0 for the recipient. Write this allocation as ($9,

---

[1]See for example Camerer (2003).

$0). Such behavior contradicts altruistic concern regarding the recipient's payoff, because then the allocation ($9, $1) should be strictly preferred. It also contradicts purely selfish preferences, as ($10, $0) would be preferred to ($9, $0). Instead, people seem to suffer from behaving egoistically in a choice situation where they could dictate a fairer allocation. Hence, if they can avoid getting into such a situation, they happily do so. Real-life scenarios with this character could be:

• donating to a charity over the phone but wishing not to have been home when the call came,

• crossing the road to avoid meeting a beggar.

Our explanation of this type of behavior is the following: Whether a person's actions are observed or not plays a crucial role in determining her behavior. We term "shame" the motive that distinguishes choice behavior when observed from choice behavior when not observed. In our model, individuals are selfish when not observed. Thus, concern for another person's payoff is motivated not by altruism, but by avoiding the feeling of shame that comes from behaving selfishly when observed.[2] The interpretation is that, if people are observed, they feel shame when they do not choose the fairest available alternative.[3]

We axiomatically formalize the notion of shame and its interaction with selfishness as described above. To this end, we consider games like the one conceived by Dana et al (2006) as a two-stage choice problem. In the first stage, the decision maker (DM) chooses a "menu," a set of payoff-allocations between herself and the anonymous recipient. This choice is not observed by the recipient. In the second stage, she makes a potentially anonymous choice from the alternatives on this menu, where the recipient observes the chosen alternative in full knowledge of the menu.[4] DM has well-defined preferences over sets of alternatives (menus). Our interpretation of shame as the motivating emotion allows considerations of fairness to

---

[2] To distinguish shame from guilt, note that guilt is typically understood to involve regret, even in private, while, according to Buss (1980), *"shame is essentially public; if no one else knows, there is no basis for shame. [...] Thus, shame does not lead to self-control in private."* We adopt the interpretation that even observation of a selfish behavior without identification of its purveyor can cause shame.

[3] In a parallel work, Neilson (2006-b) entertains a very similar notion of shame. The questions and the methodology of the two works are different. Section 6 comments in more detail.

[4] If the exit option is chosen in the aforementioned experiment by Dana et al, as in our setup, the recipient does not observe that there was a dictator, who could have chosen another allocation. In their experiment, the recipient is further unaware that another person was involved at all. It would be interesting to see how informing the recipient that some other person had received $9 would change the experimental findings. This would correspond to our setup.

impact preferences only through their effect on second-stage choices, where the presence of a fairer option reduces the attractiveness of an allocation. The underlying normative notion of fairness is central to our model, because assumptions on the norm of fairness are indirect assumptions on DM's preferences. Assuming a *particular* norm of fairness is difficult, descriptively as well as normatively. Instead, we pose what we consider minimal normative constraints on fairness.

Our representation results establish a correspondence between DM's norm of fairness and her choice behavior. On the one hand, this illustrates how those minimal constraints on fairness impact choice. On the other hand, the particular norm of fairness used by DM can be elicited from her choice behavior.

2.1.2 Illustration of Results

Denote a typical menu as $A = \{(a_1, a_2), (b_1, b_2), ...\}$, where the first and second components in each alternative are, respectively, the private payoff for DM and for the recipient. We pose axioms on DM's preferences over menus that allow us to establish a sequence of representation theorems. To illustrate our results, consider a special case of those representations:

$$U(A) = \max_{(a_1, a_2) \in A} [u(a_1) + \beta \varphi(a_1, a_2)] - \beta \max_{(b_1, b_2) \in A} [\varphi(b_1, b_2)],$$

where u and φ are increasing in all arguments. u is a utility function over private payoffs and $\varphi(a_1, a_2)$ is interpreted as the fairness of the allocation (a_1, a_2).

Alternatively, if we denote by a^* and b^* the two maximizers above, it can be written as:

$$U(A) = \underbrace{u(a_1^*)}_{\text{value of private payoff}} - \underbrace{\beta(\varphi(b_1^*, b_2^*) - \varphi(a_1^*, a_2^*))}_{\text{shame}}.$$

This representation captures the tension between the impulse to maximize private payoff and the desire to minimize shame from not choosing the fairest alternative within a set. It evaluates a menu by the highest utility an allocation on the menu gets, where this utility depends on the menu itself. The utility function that is used to evaluate allocations is additive and has two distinct components. The first component, $u(a_1)$, gives the value of a

degenerate menu (a singleton set) that contains the allocation under consideration. When evaluating degenerate menus, which leave DM with a trivial choice under observation, we assume her to be *selfish*: she prefers one allocation to another if and only if the former gives her a greater private payoff, independent of the recipient's payoff. The second component is "shame." It represents the cost DM incurs when selecting (a_1, a_2) in the face of the fairest available alternative, (b_1^*, b_2^*).

As shame is evoked whenever this fairest available alternative is not chosen, we can relate choice to a second binary relation "fairer than," which represents DM's private norm of fairness. We assume that DM's private norm of fairness induces a *Fairness Ranking* of all alternatives, which is represented by $\varphi(a_1, a_2)$. We further assume that DM's norm of fairness satisfies *Solvability*, implying that the fairness ranking is never satiated in one player's payoff, and the *Pareto* criterion in payoffs, implying that φ is increasing in all arguments.

In the special case considered here, the shame from choosing (a_1, a_2) in stage two is $\beta(\varphi(b_1^*, b_2^*) - \varphi(a_1, a_2))$. Hence, even alternatives that are not chosen may matter for the value of a set, and larger sets are not necessarily better. To see this, consider the representation above with $u(a_1) = a_1$, $\beta = \frac{1}{2}$ and $\varphi(a_1, a_2) = a_1 a_2$. Compare the sets $\{(10, 1), (4, 3)\}$, $\{(10, 1)\}$ and $\{(4, 3)\}$. Evaluating these sets we find $U\{(10, 1), (4, 3)\} = 9$, $U\{(10, 1)\} = 10$ and $U\{(4, 3)\} = 4$. To permit such a ranking, we assume a version of *Left Betweenness*, which allows smaller sets to be preferred over larger sets. Left Betweenness weakens the Set Betweenness assumption first introduced by Gul and Pesendorfer (2001), henceforth GP. Theorem 1 establishes that our weakest representation, which captures the intuition discussed thus far, is equivalent to the collection of all the above assumptions.

Selfishness leaves no room for altruism. Suppose, however, that only the second stage of the procedure is observed (for example, because DM, as in the classic dictator game, never gets to choose between menus). In this case, our representations might conform with DM behaving as if she had direct interest in the recipient's welfare and had to trade off this altruistic motive with concerns about her private payoff. We argue that it is hard to reconcile such an interpretation with observing any choice reversal in stage two. Thus, when observing stage two in isolation, shame can mimic altruism only if the induced choice

ranking is set independent. Theorem 2 establishes that, given the assumptions made so far, an additional separability assumption on preferences over sets, *Consistency*, is equivalent to the existence of such a ranking. In the special case of our representation considered above, the induced choice behavior satisfies *Consistency*. To see this, regroup the terms as follows:

$$U\left(A\right) = \underbrace{\max_{\left(a_1,a_2\right)\in A}\left[u\left(a_1\right) + \beta\varphi\left(a_1,a_2\right)\right]}_{\text{second stage choice criterion}} - \beta\underbrace{\max_{\left(b_1,b_2\right)\in A}\left[\varphi\left(b_1,b_2\right)\right]}_{\text{effect of fairest alterative}}.$$

We further specify the norm of fairness by assuming that the private payoffs to the two players have *Independent Fairness Contributions*: The fairness contribution of raising one player's payoff can not depend on the level of the other player's payoff. The idea is that interpersonal utility comparisons are infeasible. With this additional assumption, Theorem 3 establishes that there are two utility functions, v_1 and v_2, evaluated in the payoff to DM and the recipient respectively, such that the value of their product represents the fairness ranking, $\varphi\left(a_1,a_2\right) = v_1\left(a_1\right)v_2\left(a_2\right)$. Thus, the fairest alternative within a set of alternatives can be characterized as the Nash Bargaining Solution (NBS) of an associated game. Because the utility functions used to generate this game are private, so is the norm.[5] We argue that when based on true selfish utilities, the NBS is a convincing fairness criterion in our context. Those utilities, however, may not be publicly known, especially in anonymous choice situations, and therefore, DM may not be able to base her evaluation on true selfish utilities. Nevertheless, one can assess the descriptive appeal of the representation by asking whether the utilities comprising the norm at least resemble selfish utilities.

Example: Let $u\left(a_1\right) = a_1$, $\varphi\left(a_1,a_2\right) = v_1\left(a_1\right)v_2\left(a_2\right) = a_1a_2$ and $\beta = \frac{1}{2}$. This implies that selfish utility u is risk neutral and unbounded, and that the utilities v, which are used to generate the fairness ranking, coincide with u. Shame is half the difference between the Nash-product of the fairest and the chosen alternatives. Reconsider the experiment by Dana et al (2006) mentioned above, with the added constraint that only integer values are possible allocations. The set $A = \left\{\left(10,0\right)\left(9,1\right)\left(8,2\right),...,\left(0,10\right)\right\}$ then corresponds

---

[5]Therefore, the fairness ranking could also be represented by a different functional, based on different utilities.

to the dictator game. It induces the imaginary bargaining game with possible utility-allocations $\{(10, 0) (9, 1) (8, 2), ..., (0, 10), (0, 0)\}$, where the imaginary disagreement point is $\lim_{(x,y) \to 0} \left(v_1^{-1}(x), v_2^{-1}(y) \right) = (0, 0)$. According to the NBS, $(5, 5)$ would be the outcome of the bargaining game. Its fairness is $5 \cdot 5 = 25$. To trade off shame with selfishness, DM chooses the alternative that maximizes the sum of private utility and fairness, $a_1 + a_1 a_2$, which is $(6, 4)$. Its fairness is $6 \cdot 4 = 24$ and the shame incurred by choosing it is $\frac{1}{2}$. Hence $U(A) = 5.5$. From the singleton set $B = \{(9, 0)\}$, which corresponds to the exit option in the experiment, the choice is trivial and $U(B) = 9$. This example illustrates both the tension DM is exposed to when choosing from a large set and the reason why she might prefer a smaller menu.

Finally, Theorem 4 extends the former representations by allowing DM to be responsible for the welfare of many other recipients. This extension is then applied to model a social decision maker who is able to alter the transparency of her policies' consequences. Policies create social value, but also have a redistributive component. DM faces a trade-off when choosing the transparency of her policies: More transparency makes it easier for the public to perceive fair choices as such, while less transparency makes it harder for society to detect selfish choices. Shame, therefore, might lead her to implement policies with relatively opaque consequences.

The organization of the paper is as follows: Section 2 presents the basic model and a representation that captures the concepts of fairness and shame. Section 3 isolates a choice criterion from the choice situation. Section 4 further specifies the fairness ranking. Section 5 extends the representation to finitely many other players and suggests an application to a social decision maker. Section 6 points out connections to existing literature and section 7 concludes.

2.2 The Model

Let K be the set of all finite subsets of $\mathbb{R}^2_+$.[6] Any element $A \in K$ is a finite set of alternatives. A typical alternative $\mathbf{a} = (a_1, a_2)$ is interpreted as a payoff pair, where a_1 is the private payoff for DM and a_2 is the private payoff allocated to the (potentially anonymous) other player, the recipient. Endow K with the topology generated by the Hausdorff metric, which is defined for any pair of non-empty sets, $A, B \in K$, by:

$$d_h(A, B) := \max \left[\max_{\mathbf{a} \in A} \min_{\mathbf{b} \in B} d(\mathbf{a}, \mathbf{b}), \max_{\mathbf{b} \in B} \min_{\mathbf{a} \in A} d(\mathbf{a}, \mathbf{b}) \right],$$

where $d : \mathbb{R}^2_+ \to \mathbb{R}_+$ is the standard Euclidian distance.

Let $\succ$ be a continuous preference relation (weak order) over K. We write $A \succ B$ if DM strictly prefers A to B. The associate weak preference, $\succeq$ and the indifference relation, $\sim$ are defined in the usual way.

The choice of a menu $A \in K$ is not observed by the recipient, while the choice *from* any menu is. We call the impact this observation has on choice "shame." Of course various other regarding preferences that are not impacted by observation could be present as well. We do not account for those, as our aim is not to describe a range of possible attitudes toward others, but to derive a tractable representation according to which DM distinguishes the two stages in an intuitive way.

The first axiom specifies DM's preferences over singleton sets.

P_1 **(Selfishness)** $\quad \{\mathbf{a}\} \succ \{\mathbf{b}\}$ *if and only if* $a_1 > b_1$.

A singleton set $\{\mathbf{a}\}$ is a degenerate menu that contains only one feasible allocation, (a_1, a_2). It leaves DM with a trivial choice to be made when being observed in the second stage. Therefore, the ranking over singleton sets can be thought of as the ranking over allocations that are imposed on DM. We contend that there is no room for shame in this situation; choosing between two singleton sets reveals DM's "true" preferences over allocation outcomes. The axiom states that DM is not concerned about the payoff to the

---

[6] With $\mathbb{R}_+$ we denote the positive reals including 0. $\mathbb{R}_{++}$ denotes the positive reals without 0.

second player when evaluating such sets; she compares any pair of alternatives based solely on the first component, her private payoff. If, for example, DM had an altruistic concern for fairness in the dictator game previously described, she would strictly prefer the menu $\{(9,1)\}$ to $\{(9,0)\}$. P_1 rules out such altruistic concerns. Negative emotions regarding the other player, such as spite or envy, are ruled out as well.

The next axiom captures that shame is a mental cost, which is invoked by unchosen alternatives.

P_2 (**Strong Left Betweenness**)   *If $A \succeq B$, then $A \succeq A \cup B$. Further, if $A \succ B$ and $\exists C$ such that $A \cup C \succ A \cup B \cup C$, then $A \succ A \cup B$.*

We assume that adding unchosen alternatives to a set can only increase shame. Therefore, no alternative is more appealing when chosen from $A \cup B$, than when chosen from one of the smaller sets, A or B. Hence, $A \succeq B$ implies $A \succeq A \cup B$.[7] Furthermore, if additional alternatives add to the shame incurred by the original choice from a menu $A \cup C$, then they must also add to the shame incurred by any choice from the smaller menu A. Thus, if there is C such that $A \cup C \succ A \cup B \cup C$ and if $A \succ B$, then $A \succ A \cup B$.

Shame, which is the only motive DM knows beyond selfishness, must refer to some personal norm that determines what the appropriate choice should have been. In our interpretation, this norm is to choose one of the fairest available allocations. Interpreting "fairness" as a property of an allocation, which is independent of the menu it is on, we consider a binary relation $\succ_f$ over $\mathbb{R}^2_+$ as a second primitive.

Definition: If $\mathbf{b} \succ_f \mathbf{a}$, we say that DM considers $\mathbf{b}$ to be *fairer than* $\mathbf{a}$.

Some of the axioms below are imposed on $\succ_f$ rather than on $\succ$ and are labeled by F instead of P. The underlying notion of fairness is at the heart of those assumptions.[8] To

---

[7]This is the "Left Betweenness" axiom. It appears in Dekel, Lipman and Rustichini (2005) and is a weakening of "Set Betweenness" as first posed in GP.

[8]In everyday language, "fair" is sometimes used to capture various different notions. According to the Merriam-Webster Collegiate Dictionary (Tenth Edition, 2001) "*Fair implies an elimination of one's own feelings, prejudices, and desires so as to achieve a proper balance of conflicting interests.*" This is the

make them descriptively intuitive, we emphasize their normative appeal, implying that DM will want her norm of fairness to satisfy them. Making these assumptions directly on $\succ_f$ is natural. The relation $\succ_f$ is not directly observable, but the next axiom relates it to observable choice behavior. One contribution of our work is that the implications of F-axioms on $\succ$ are most easily understood from the representation.

P_3 (**Shame**) *If $\exists A \in K$ with $\mathbf{a} \in A$, such that $A \succ A \cup \{\mathbf{b}\}$, then $\mathbf{b} \succ_f \mathbf{a}$.[9]*

$A \succ A \cup \{\mathbf{b}\}$ implies that $\mathbf{b}$ adds to the shame incurred by the original choice in A. The interpretation is that DM is concerned about not choosing one of the fairest available alternative. Thus, $\mathbf{b}$ must be fairer than any alternative in A, in particular $\mathbf{b} \succ_f \mathbf{a}$.

Definition: We say that DM is *susceptible to shame* if there exists A and B with $A \succ A \cup B$.

F_1 (**Fairness Ranking**) *$\succ_f$ is an anti-symmetric and negatively transitive binary relation.*

Our discussion rests on the assumption that DM can rank alternatives according to their fairness. In $\mathbb{R}_+^2$ and with increasing utility from self-payoffs, this assumption is not unreasonably restrictive.[10]

Combined with P_3, F_1 implies that only one alternative in each menu, the fairest, is responsible for shame.

F_2 (**Pareto**) *If DM is susceptible to shame, then $\mathbf{a} \geq \mathbf{b}$ and $\mathbf{a} \neq \mathbf{b}$ imply $\mathbf{a} \succ_f \mathbf{b}$.*

According to this axiom, absolute, as opposed to relative, well-being matters; the Pareto criterion excludes notions such as "strict inequality aversion." The resulting concept of fairness must have some concern for efficiency. In the case where there truly is no potential for

---

definition of "fair" we base our arguments on.

[9] The notion of "fairer than" is analogous to the definition of "more tempting than" in Gul and Pesendorfer (2005).

[10] If, instead, there were a globally most prefered self-payoff, this assumption would rule out very reasonable preference rankings.

redistribution, we believe that people find the Pareto criterion a reasonable requirement for one allocation to be fairer than another.[11]

F_3 (**Solvability**)   *If* $(a_1, 0) \nsucc_f (b_1, b_2)$ *then* $\exists x$ *such that* $(a_1, x) \sim_f (b_1, b_2)$. *Analogously, if* $(0, a_2) \nsucc_f (b_1, b_2)$ *then* $\exists y$ *such that* $(y, a_2) \sim_f (b_1, b_2)$.

Ignoring the qualifier, the axiom states that in order to make two allocations deemed equally fair, any variation in the level of one person's payoff can always be compensated by appropriate variation in the level of the other person's payoff. This requires $\succ_f$ never to be satiated in any person's payoff. Relying on F_2, the qualifiers take into account that monetary payoffs are bounded below by 0. For example, F_3 implies that there is a sum x, such that $(x, 1) \sim_f (10, 10)$. This assumption captures the insight that any fairness ranking with a concern for efficiency must go beyond the Pareto principle and trade off, in some manner, payoffs across individuals.

As $\succ$ is continuous, $\succ_f$ is continuous in all alternatives for which P_3 relates $\succ$ to $\succ_f$. $F_1 - F_3$ imply that this is the case on $\mathbb{R}_+ \times \mathbb{R}_{++}$.[12] Assuming that $\succ_f$ is continuous even in alternatives for which P_3 does not relate $\succ$ to $\succ_f$ has obviously no implication for choice. For ease of exposition, we assume in all what follows that $\succ_f$ is continuous on all of $\mathbb{R}_+^2$.

---

[11]In many contexts, people would disagree with the statement that the allocation $(1million, 6)$ is fairer than $(5, 5)$. On the basis of the definition in footnote 10, however, we claim that the opposition to $(1million, 6)$ as a fair allocation can only be based on the implicit premise that there must be some mechanism to divide the gains more evenly (Such a mechanism would imply the availability of a third option, which would render both of the above allocations unfair.) In an explicit choice situation this premise cannot be sustained. The Pareto property has indeed been advocated in the philosophical literature on fairness. Rawls (1971), for example, proposes the idea of "original position," a mental exercise whereby a group of rational people must establish a principle of fairness (e.g. when distributing income) without knowing beforehand where on the resulting pecking order they will end up themselves. Requiring that the allocation satisy *Pareto* makes much sense in such an environment.

[12]$\succ_f$ is relevant for choice in alternative **b**, if and only if there is **c** with $\mathbf{c} \prec_f \mathbf{b}$ and $c_1 > b_1$, which requires $c_2 < b_2$. Thus $b_2 > 0$ is necessary for the construction of **c**.

Theorem 1 *If DM is susceptible to shame, then $\succ$ and $\succ_f$ satisfy $P_1 - P_3$ and $F_1 - F_3$ respectively, if and only if there exist continuous and strictly increasing functions $u : \mathbb{R}_+ \to \mathbb{R}$, $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ and a continuous function $g : \mathbb{R}_+^2 \times \varphi\left(\mathbb{R}_+^2\right) \to \mathbb{R}$, weakly increasing in its second argument and satisfying: $g\left(\mathbf{a}, x\right) \gtreqless 0$ whenever $\varphi\left(\mathbf{a}\right) \lesseqgtr x$, such that the function $U : K \to \mathbb{R}$ defined as $U\left(A\right) = \max\limits_{\mathbf{a} \in A} \left[u\left(a_1\right) - g\left(\mathbf{a}, \max\limits_{\mathbf{b} \in A} \varphi\left(\mathbf{b}\right)\right) \right]$ represents $\succ$ and φ represents $\succ_f$.*

If DM is not susceptible to shame, $g \equiv 0$.

All detailed proofs are in the appendix. We now highlight the important steps. As both $\succ$ and $\succ_f$ are continuous binary relations, they can be represented by continuous functions $U : K \to \mathbb{R}$ and $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ respectively. φ is an increasing function as implied by *Pareto* (F_2). The combination of *Strong Left Betweenness* (P_2), *Shame* (P_3) and *Fairness Ranking* (F_1) implies GP's *Set Betweenness* (SB) property: $A \succeq B$ implies $A \succeq A \cup B \succeq B$. GP demonstrate that imposing SB on preferences over sets makes every set indifferent to a certain subset of it, which includes at most two elements (Lemma 2 in their paper). Hence we confine our attention to a subset of our domain that includes all sets with cardinality no greater than 2. *Selfishness* (P_1) and P_3 imply that a set $\{\mathbf{a}, \mathbf{b}\}$ is strictly inferior to $\{\mathbf{a}\}$ if and only if $a_1 > b_1$ and $\mathbf{b} \succ_f \mathbf{a}$. We can then strengthen GP's Lemma 2 and state that any set is indifferent to some two-element set that includes one of the fairest allocations in the original (larger) set. Using *Solvability* (F_3) we show the continuity of the second component, the function g, in the representation.

The representation in Theorem 1 highlights the basic trade-off between private payoff and shame as the only concepts DM may care about. There are at most two essential alternatives within a set, to be interpreted as the "chosen" and the "fairest" alternative, $\mathbf{a}$ and $\mathbf{b}$ respectively. For the latter, its fairness, $\varphi\left(\mathbf{b}\right)$, is a sufficient statistic for its impact on the set's value. DM suffers from shame, measured by $g\left(\mathbf{a}, \varphi\left(\mathbf{b}\right)\right)$, whenever $\varphi\left(\mathbf{a}\right) < \varphi\left(\mathbf{b}\right)$, where $\varphi\left(\mathbf{a}\right)$ is the fairness of the chosen alternative. The representation captures the idea of shame being an emotional cost that emerges whenever the fairest available allocation is not chosen. Its magnitude may depend on the fairness of the chosen allocation.

The main contribution of Theorem 1 is the provision of a way to elicit DM's fairness

ranking, $\succ_f$, from choice behavior: all functions in the representation are continuous and hence, for $\mathbf{b} \in \mathbb{R}_+ \times \mathbb{R}_{++}$ and $\mathbf{b} \succ_f \mathbf{a}$, there is $\mathbf{c}$, such that $U(\{a, c\}) > U(\{a, b, c\})$. Since it is continuous, φ is then uniquely determined on its entire domain, $\mathbb{R}_+^2$.

Note that the properties of the function g and the max operator inside imply that the second term is always a cost (non-positive). The other max operator implies that DM's payoff will never lie below b_1, which is her payoff as suggested by the fairest allocation. Thus, any deviations by DM from choosing the fairest allocation will be in her own favor. These observations justify labeling said cost as "shame."

From the representation, it is easy to see that the induced choice correspondence,

$$C(A) := \left\{ \arg\max_{\mathbf{a} \in A} \left[u(a_1) - g\left(\mathbf{a}, \max_{\mathbf{b} \in A} \varphi(\mathbf{b}) \right) \right] \right\}$$

 may be context dependent in the sense that a higher degree of shame may affect choice. In other words, if we define a binary relation "better choice than," $\succ_c$, by $\mathbf{a} \succ_c \mathbf{b}$ if $\exists B$ with $\mathbf{b} \in B$, such that $B \cup \{\mathbf{a}\} \succ B$, then this binary relation need not be acyclic. This feature may be plausible when shame is taken into account. In the next section we spell out the implications of enforcing a context-independent criterion for choice.

2.3   A Second-Stage Choice Ranking

In many situations, only second-stage choice may be observable. For example, the standard dictator game corresponds only to second-stage choice in our setup. Typical behavior in various versions of this game, where subjects tend to give part of the endowment to the recipient, is often interpreted as motivated by an altruistic motive. We interpret altruism to imply that the recipient's welfare is a good, just as selfishness implies that DM's private payoff is a good.[13] If DM had those two motives, she would have to make a trade-off between them. As in the case of two generic goods, very basic assumptions would lead to a context-independent choice ranking of alternatives. As we point out at the end of section 2, we can define a binary relation "better choice than," $\succ_c$, by $\mathbf{a} \succ_c \mathbf{b}$ if $\exists B$ with $\mathbf{b} \in B$,

---

[13]This interpretation is based on the following definition of altruism (Merriam-Webster Collegiate Dictionary [Tenth Edition, 2001]): "*Unselfish regard for or devotion to the welfare of others.*" We understand this definition as ruling out any considerations that condition on available but unchosen alternatives.

such that $B \cup \{\mathbf{a}\} \succ B$. This binary relation need not be acyclic: Different choice problems, A and B, may lead to different second-stage rankings of $\mathbf{a}$ and $\mathbf{b}$, for $\mathbf{a}, \mathbf{b} \in A \cap B$. If no cycles occur, second-stage behavior might look as if it were generated by, for instance, a trade-off of selfishness and altruism, even though observation of stage-one choice would rule this out. If, on the other hand, cycles are observed in stage-two choice, simple altruistic motives cannot be solely responsible for behavior that is not purely selfish. In this section we identify a condition on preferences that makes DM's second-stage choice independent of the choice set. This implies finding a function $\psi : \mathbb{R}^2_+ \to \mathbb{R}$ that assigns a value to each $\mathbf{a} \in A$, such that $\mathbf{a}$ is a choice from A only if $\psi(\mathbf{a}) \geq \psi(\mathbf{b})$ for all $\mathbf{b} \in A$.

Definition: $X := \{(\mathbf{a}, \mathbf{b}) : \{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}\}$ is the set of all pairs of alternatives generating strict *Set Betweenness*.

For any set of two allocations $\{\mathbf{a}, \mathbf{b}\}$, we interpret the preference ordering $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$ as an indication of a discrepancy between what DM chooses ($\mathbf{a}$) and the alternative she deems to be the fairest ($\mathbf{b}$), which causes her choice to bear shame. This shame, however, is not enough to make her choose $\mathbf{b}$.

Combined with F_1, Shame (P_3) implies that choice between sets depends on the fairness of the fairest alternative in the set. The next axiom relates choice to the fairness of the chosen alternative as well: The fairer DM's choice, the less shame she feels.

P_4 **(Fairer is Better)** *If for $\{\mathbf{a}\} \sim \{\mathbf{a}'\}$ we have $\{(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b})\} \subseteq X$ and $\mathbf{a} \succ_f \mathbf{a}'$, then $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}', \mathbf{b}\}$.*

Axiom P_4 implies that only the fairness of the chosen alternative matters for its impact on shame.

Given $P_1 - P_4$ and $F_1 - F_3$, an additional separability assumption is equivalent to separable shame, and thus to a set-independent choice ranking.

P_5 (**Consistency**) *If*

$$\{(\mathbf{a}, \mathbf{b}), (\mathbf{a}, \mathbf{d}), (\mathbf{a'}, \mathbf{b'}), (\mathbf{a'}, \mathbf{d'}), (\mathbf{c}, \mathbf{b}), (\mathbf{c'}, \mathbf{b'}), (\mathbf{c}, \mathbf{d}), (\mathbf{c'}, \mathbf{d'})\} \subseteq X,$$

then $\{\mathbf{a}, \mathbf{b}\} \sim \{\mathbf{a'}, \mathbf{b'}\}$ *and* $\{\mathbf{a}, \mathbf{d}\} \sim \{\mathbf{a'}, \mathbf{d'}\}$ *imply* $\{\mathbf{c}, \mathbf{b}\} \succ \{\mathbf{c'}, \mathbf{b'}\} \Leftrightarrow \{\mathbf{c}, \mathbf{d}\} \succ \{\mathbf{c'}, \mathbf{d'}\}$.

We make no claim about the normative or descriptive appeal of this assumption. Instead, we view it as an empirical criterion: If the condition is not met, observation of stage-two choice should suffice to distinguish altruism from shame as the motive behind DM's other-regarding behavior. The axiom requires independence between the impact of the chosen and the fairest alternative on the set ranking:

$$\{(\mathbf{a}, \mathbf{b}), (\mathbf{a}, \mathbf{d}), (\mathbf{a'}, \mathbf{b'}), (\mathbf{a'}, \mathbf{d'}), (\mathbf{c}, \mathbf{b}), (\mathbf{c'}, \mathbf{b'}), (\mathbf{c}, \mathbf{d}), (\mathbf{c'}, \mathbf{d'})\} \subseteq X$$

implies that from each of the sets $\{\mathbf{a}, \mathbf{b}\}, \{\mathbf{a}, \mathbf{d}\}, \{\mathbf{a'}, \mathbf{b'}\}, \{\mathbf{a'}, \mathbf{d'}\}, \{\mathbf{c}, \mathbf{b}\}, \{\mathbf{c'}, \mathbf{b'}\}, \{\mathbf{c}, \mathbf{d}\}$ and $\{\mathbf{c'}, \mathbf{d'}\}$, the alternative listed first is chosen in the second stage despite the availability of a fairer alternative, which is listed second. Assume, without loss of generality that $\{\mathbf{a}\} \succ \{\mathbf{a'}\}$. Suppose there are two pairs of fairer and less attractive alternatives, $\mathbf{b}, \mathbf{b'}$ and $\mathbf{d}, \mathbf{d'}$, such that for each of them pairing their members with $\mathbf{a}$ and $\mathbf{a'}$, respectively, gives rise to indifference. In the context of Theorem 1, this implies that both pairs induce the same shame differential, which exactly cancels the selfish preference of $\{\mathbf{a}\}$ over $\{\mathbf{a'}\}$: $\{\mathbf{a}, \mathbf{b}\} \sim \{\mathbf{a'}, \mathbf{b'}\}$ and $\{\mathbf{a}, \mathbf{d}\} \sim \{\mathbf{a'}, \mathbf{d'}\}$. Then, the axiom states that pairing the members of $\mathbf{b}, \mathbf{b'}$ or $\mathbf{d}, \mathbf{d'}$ with any other chosen alternatives $\mathbf{c}$ and $\mathbf{c'}$, respectively, must also lead to the same differential in shame. In particular, $\{\mathbf{c}, \mathbf{b}\} \succ \{\mathbf{c'}, \mathbf{b'}\}$ implies $\{\mathbf{c}, \mathbf{d}\} \succ \{\mathbf{c'}, \mathbf{d'}\}$. Again, the validity of this technical assumption in a given context is an empirical question.

Theorem 2 *If DM is susceptible to shame, then* $\succ$ *and* $\succ_f$ *satisfy* $P_1 - P_5$ *and* $F_1 - F_3$ *respectively, if and only if there exist continuous and strictly increasing functions* $u : \mathbb{R}_+ \to \mathbb{R}$ *and* $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$, *such that the function* $U : K \to \mathbb{R}$ *defined as* $U(A) = \max_{\mathbf{a} \in A} [u(a_1) + \varphi(a_1, a_2)] - \max_{b \in A} [\varphi(b_1, b_2)]$ *represents* $\succ$ *and* φ *represents* $\succ_f$.

The proof constructs a path in the (a_1, a_2)-plane such that the fairness $\varphi(\mathbf{a})$ increases along this path. Then, on two neighboring indifference curves in the $(\mathbf{a}, \varphi(\mathbf{b}))$-space, $\varphi(\mathbf{b})$ increases, as $\mathbf{a}$ varies along the path. Relying on P_5, these indifference curves allow us to rescale $\varphi(\mathbf{b})$ to make the representation of $\succ$ quasi-linear.[14] Separability is then immediate. Since the proof of Theorem 2 is a special case of the proof of Theorem 4, we only go through the more general case in detail in the appendix.

The representation isolates a choice criterion that is independent of the choice problem: DM's behavior is governed by maximizing

$$u(a_1) + \varphi(a_1, a_2).$$

The value of the set is reduced by

$$\max_{b \in A} \varphi(b_1, b_2),$$

a term that depends solely on the fairest alternative in the set. Grouping the terms differently reveals the trade-off between self-payoff, $u(a_1)$, and the shame involved with choosing $\mathbf{a}$ from the set A:

$$\max_{\mathbf{b} \in A} [\varphi(b_1, b_2) - \varphi(a_1, a_2)] \geq 0.$$

Note that now shame takes an additively separable form, depends only on the fairness of both alternatives, and is increasing in the fairness of the fairest and decreasing in that of the chosen alternative. If $P_1 - P_4$ and $F_1 - F_3$ hold, then, according to Theorem 2, P_5 is equivalent to having a set-independent choice ranking.

2.4 Specifying a Fairness Ranking

In this section we impose one more axiom on $\succ_f$ to further characterize the fairness ranking. It asserts that the fairness contribution of one person's marginal payoff cannot depend on

---

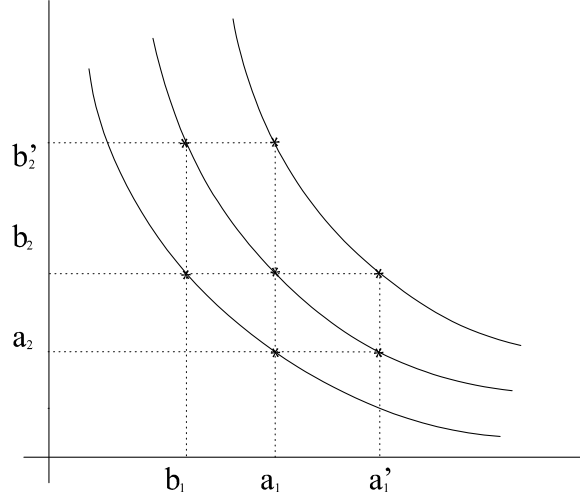[14]A more elaborate discussion on this technique appears after Theorem 3.

Figure 2.1: Independent Fairness Contributions.

the initial payoff levels.

F_4 (**Independent Fairness Contributions**) *If* $(a_1, a_2) \sim_f (b_1, b_2)$ *and* $(a_1', a_2) \sim_f (a_1, b_2)$ $\sim_f (b_1, b_2')$, *then* $(a_1', b_2) \sim_f (a_1, b_2')$.

The axiom is illustrated in figure 1. If $a_1 = a_1'$ or $b_2 = b_2'$, this axiom is implied by F_1, F_2 and the continuity of $\succ_f$. For $a_1 \neq a_1'$ and $b_2 \neq b_2'$, the statement is more subtle. Consider first a stronger assumption:

F_4' (**Strong Independent Fairness Contributions**) $(a_1, a_2) \sim_f (b_1, b_2)$ *and* $(a_1', a_2) \sim_f$ (b_1, b_2') *imply* $(a_1', b_2) \sim_f (a_1, b_2')$.

The fairness contribution of one person's marginal payoff cannot depend on the initial payoff level of the other person: It is unclear to DM how much an increase in monetary payoff means to the recipient, because even if the (marginal) utility of the recipient were known to DM, she could not compare it to her own, as interpersonal utility comparisons are infeasible. The qualifier in F_4' establishes that DM considers the fairness contribution of changing her own payoff from a_1 to a_1' given the allocation (a_1, a_2) to be the same as that of changing the recipient's payoff from b_2 to b_2' given (b_1, b_2). F_4' then states that starting from

the allocation (a_1, b_2), changing a_1 to a_1' should again be as favorable in terms of fairness as changing b_2 to b_2'. This is the essence of *Independent Fairness Contributions*. The stronger qualifier $(b_1, b_2') \sim_f (a_1, b_2) \sim_f (a_1', a_2)$ in F_4 weakens the axiom. For example, the fairness ranking $(a_1, a_2) \succ_f (b_1, b_2)$ if and only if $\min(a_1, a_2) > \min(b_1, b_2)$ is permissible under F_4, but not under F_4'.[15]

Theorem 3 $\succ_f$ *satisfies* $F_1 - F_4$, *if and only if there are continuous, increasing and unbounded functions* $v_1, v_2 : \mathbb{R}_+ \to \mathbb{R}_{++}$, *such that* $\varphi(\mathbf{a}) = v_1(a_1) v_2(a_2)$ *represents* $\succ_f$.

Luce and Tukey (1964) prove the necessity and sufficiency of *Solvability* and the *Corresponding Trade-offs Condition* (the label they use for F_4) to admit an additive representation. To show how a proof works, we repeatedly use axiom F_4 to establish that if $(a_1, a_2) \sim_f (a_1', a_2')$ and $(a_1, \tilde{a}_2) \sim_f (a_1', \tilde{a}_2')$, then $(\tilde{a}_1, a_2) \sim_f (\tilde{a}_1', a_2') \Leftrightarrow (\tilde{a}_1, \tilde{a}_2) \sim_f (\tilde{a}_1', \tilde{a}_2')$. With this knowledge, we can create a monotone increasing mapping $a_2 \to \gamma(a_2)$ that transforms the original indifference map to be quasi-linear with respect to the first coordinate in the $(a_1, \gamma(a_2))$ plane. Keeney and Raiffa (1976) refer to the procedure we employ as the lock-step procedure. Quasi-linearity implies that there is an increasing continuous function $\xi : \mathbb{R}_+ \to \mathbb{R}$, such that $\varphi(\mathbf{a}) := \xi(a_1) + \gamma(a_2)$ represents $\succ_f$. Define $v_1(a_1) := \exp(\xi(a_1))$ and $v_2(a_2) := \exp(\gamma(a_2))$. Then $v_1, v_2 : \mathbb{R}_+ \to \mathbb{R}_{++}$ are increasing and continuous and if we redefine $\varphi(\mathbf{a}) := v_1(a_1) v_2(a_2)$, it represents $\succ_f$.

This representation suggests an appealing interpretation of the fairness ranking DM is concerned about: She behaves *as if* she had in mind two increasing and unbounded utility functions, one for herself[16] and one for the recipient. By mapping the alternatives within each set into the associated utility space, any choice set induces a finite bargaining game where only the disagreement point is unspecified. DM then identifies the fairest alternative within a set *as if* she also had in mind a disagreement point, that makes this alternative

---

[15] F_4 is refered to as the *Hexagon condition* or the *Corresponding Trade-offs Condition* (Keeney and Raiffa [1976]), F_4' as the *Thomsen condition*. With F_2 and F_3, F_4' is implied by F_4. See Karni and Safra (1998) for a proof.

[16] This utility function need not agree with her true utility for personal payoffs, u. The interpretation is that DM is concerned about the recipient's perception of her choice. The recipient, however, may not know DM's true utility, especially under anonymity.

the Nash Bargaining Solution[17] of the game.[18] Moreover, the fairness of all alternatives can be ranked according to the same functional, namely the Nash product.

Remember that F_3 requires trading off marginal payoffs. The tension of having to trade off marginal payoffs without being able to compare their welfare contribution (F_4) is common in a range of social-choice problems.[19] Our axioms are weak in the sense that they do not constrain DM in this trade-off, as long as she takes into account that the fairness contribution of increasing one person's payoff should not depend on the other's payoff. The power of Theorem 3 is that it bases a representation on these weak assumptions. The downside is that the form of this representation is not unique, as the utilities v_1 and v_2 are not observable independent of the norm of fairness. For example, there is another pair of increasing utility functions such that DM is concerned about their sum, that is, she acknowledges efficiency as the only fairness criterion.

To underline the appeal of the Nash product as a descriptive representation of fairness,[20] we now point out how DM might reason within the constraints of the axioms:

We justified the Pareto criterion, F_2, as a plausible axiom for the fairness ranking. As argued above, concern for fairness requires the acknowledgment of some form of interpersonal comparability of preferences' intensity. If utilities were known cardinally, symmetry in terms of utility payoffs is the other criterion we would expect the ranking to satisfy.[21] In our context, this implies independence of the role people play, dictator or recipient. However, utilities are inherently ordinal, rendering such a comparison infeasible. At best we can, if we assume people to have cardinal utilities that reflect their attitudes toward risk, determine marginal utilities up to scaling. Mariotti (1997), for example, considers a context in which *"interpersonal comparisons of utility are meaningful; that is, there exists an (unknown)*

---

[17]See Nash (1950).

[18]The imaginary disagreement point is determined by $\lim_{(x,y)\to 0} \left(v_1^{-1}(x), v_2^{-1}(y)\right)$. It could be some finite and weakly positive pair of utility payoffs. In particular it could be $(0,0)$, which corresponds to DM imagining that players walk away in the case that no agreement is reached. It could also be negative. This corresponds to DM imagining that players have an extra incentive to find an agreement: there is a cost to disagreement.

[19]For a review, see Hammond (1990).

[20]Even though u and v_1 do not have to agree, our interpretation might be more convincing when they resemble each other empirically. In particular it is more appealing if DM's actual utility from self-payoff u is unbounded.

[21]This reasoning leads Rawls (1971) to suggest *Pareto* and *Symmetry* as the two criteria a decision maker under a veil of ignorance should respect.

rescaling of each person's utility which makes utilities interpersonally comparable." At the same time, however, *"interpersonal comparisons of utility are not feasible."* Assume there is a correct interpersonal utility scaling, but DM cannot determine it. Can she guarantee that for this unknown scaling both symmetry and Pareto are satisfied? They would have to be satisfied for all potential scalings. Mariotti establishes that the NBS is the only criterion with this property.

Even more appealing is an interpretation of the NBS as the fairest allocation that is related to Gauthier's (1986) principle of "moral by agreement": Trying to assess what is fair, but finding herself unable to compare utilities across individuals, DM might refer to the prediction of a symmetric mechanism for generating allocations. In particular, DM might ask what would be the allocation if both she and the recipient were to bargain over the division of the surplus. To answer this question, she does not need to assume the intensities of the two preferences. This is a procedural interpretation that is not built on the axioms: DM is not ashamed of payoffs, but of using her stronger position in distributing the gains. It is, then, the intuitive and possibly descriptive appeal of the NBS in many bargaining situations that makes it normatively appealing to DM in our context.[22] Theorem 3 establishes the behavioral equivalence of this interpretation and our axioms.

The Pareto and the Solvability axioms, F_2 and F_3 respectively, rule out fairness rankings with $(x,0) \sim_f (0,y)$ for all x, y. In particular the Nash product with linear utility functions v_1, v_2 is ruled out as a criterion for fairness. Such orderings could easily be accommodated by posing *Pareto* and *Solvability* only on $\mathbb{R}^2_{++}$. As a consequence, φ would be strictly increasing only on $\mathbb{R}^2_{++}$ and v_1, v_2 would only have to be weakly positive, $v_1, v_2 : \mathbb{R}_+ \to \mathbb{R}_+$.[23] These weaker axioms would still rule out the Maximin as a criterion for fairness.

Remark: Any concern DM has about fairness originates from being observed. Consequently, DM should expect a potentially anonymous observer to share her notion of what is fair: Her private norm of fairness, which we observe indirectly, should reflect her con-

---

[22] The descriptive value of the NBS has been tested empirically. For a discussion see Davis and Holt (1993) pages 247-55. Further, multiple seemingly natural implementations of it have been proposed (Nash [1953] , Osborne and Rubinstein [1994]).

[23] As can be seen in the proof of Theorem 2, this would imply the possibility of $(-\infty, -\infty)$ as an imaginary disagreement point, which corresponds to DM imagining that players have to find an agreement (infinite cost of disagreement).

cern about not violating a social norm. If the observed choice situation is anonymous, DM does not know the recipient's identity and is aware that the recipient does not know hers. Therefore, the ranking cannot depend on either identity. Combining this with the idea that fairness of an allocation should not depend on the role a person plays, whether dictator or recipient, one might want to pose symmetry of the fairness ranking in terms of direct payoffs.

F_5 **(Symmetry)** $(a_1, a_2) \sim_f (a_2, a_1)$.

Adding this assumption constrains $v_1(a) = v_2(a)$ in the representation of Theorem 3. The numerical example given in the introduction features the combination of Theorem 2 and Theorem 3, where all functions involved are the identity. For brevity, we will not repeat it here.

2.5   Multiple Recipients and an Application

In order to expand the range of possible applications of our representation, we first extend our results to finitely many agents.

2.5.1   Multiple Recipients

The underlying idea is that DM (without loss of generality individual 1) is concerned about $N - 1 \geqslant 2$ other individuals, whose payoffs depend on her choice. In analogy to section 2, let K be the set of all finite subsets of $\mathbb{R}_+^N$. Any element $A \in K$ is a finite set of alternatives. A typical alternative $\mathbf{a} = (a_1, a_2, ..., a_N)$ is interpreted as a payoff vector, where a_n is the payoff allocated to individual n. We write, for example, $(a_m, a_n, \mathbf{a}_{-m,n})$ as the alternative with payoff a_m to individual m, payoff a_n to individual n and $\mathbf{a}_{-m,n} \in \mathbb{R}_+^{N-2}$ lists all other individuals' payoffs in order. We endow K with the topology generated by the Hausdorff metric.

Let $\succ$ be a continuous preference relation over K. Most of the axioms we pose on $\succ$ in section 2 can be readily applied to $\succ$ on this new domain. We define $\succ_f$ in analogy to the previous definition. Instead of F_3 we write

F_3^N (**Weak Solvability**) *If* $(a_n, \mathbf{0}) \not\succ_f \mathbf{b}$ *then for all* $m \neq n$, *there exists* a_m *such that* $(a_m, a_n, \mathbf{0}) \sim_f \mathbf{b}$.

The axiom states that it is always possible to equate the fairness of an allocation with payoff to only one individual to that of an initially fairer allocation by giving appropriate payoffs to any second individual. This property requires the fairness ranking never to be satiated in any individual payoff.

Definition: The pair of possible payoffs to individuals m and n is *Preferentially Independent with respect to its Complement* (P.I.C.), if the fairness ranking in the (a_m, a_n)-space is independent of $\mathbf{a}_{-m,n}$.

F_4^N (**Pairwise Preferential Independence**) *For all* $m, n \in \{1, .., N\}$, *the pair of possible payoffs to individuals* m *and* n *is P.I.C.*

Similarly to F_4, this axiom must hold if the contribution of one person's marginal private payoff to the fairness of an allocation cannot depend on another person's private payoff level.

Theorem 4 *Assume $N \geq 3$ and that DM is susceptible to shame.*
(i) $\succ$ and $\succ_f$ satisfy $P_1 - P_5$ and F_1, F_2 and F_3^N respectively, if and only if there exist continuous and strictly increasing functions $u : \mathbb{R}_+ \to \mathbb{R}$ and $\varphi : \mathbb{R}_+^N \to \mathbb{R}$ such that the function $U : K \to \mathbb{R}$ defined as $U(A) = \max\limits_{\mathbf{a} \in A} [u(a_1) + \varphi(a_1, a_2, ..., a_n)] - \max\limits_{\mathbf{b} \in A} [\varphi(b_1, b_2, ..., b_n)]$ represents $\succ$ and φ represents $\succ_f$.
(ii) $\succ_f$ also satisfies F_4^N if and only if there exist continuous and strictly increasing functions $v_1, .., v_N : \mathbb{R}_+ \to \mathbb{R}_{++}$, where $v_1, .., v_N$ are unbounded such that $\varphi(\mathbf{a}) = \prod\limits_{i1}^{N} v_i(a_i)$.

Theorem 4 is analogous to Theorem 2. For the proof, note that the analogue of Theorem 1 can be established by substituting $\mathbf{a}_{-1}$ for a_2 in the theorem and in the proof, where now $\varphi : \mathbb{R}_+^N \to \mathbb{R}$. To establish the analogue of Theorem 3, namely that there are N increas-

ing unbounded functions $v_1, .., v_N$, such that the fairness ranking $\succ_f$ can be represented by $\varphi(\mathbf{a}) = \prod\limits_{i1}^{N} v_i(a_i)$ if and only if it satisfies F_1, F_2, F_3^N and F_4^N, we first state a stronger version of F_3^N:

$F_3^{N\prime}$ (**Solvability**) *If $(a_n, \mathbf{a}_{-n}) \not\succ_f \mathbf{b}$ then for all $m \neq n$, there exists a_m such that $(a_m, a_n, \mathbf{a}_{-m,n}) \sim_f \mathbf{b}$.*

We observe that *Continuity, F_1, F_2* and F_3^N imply *Solvability*. To see this, assume $(a_n, \mathbf{a}_{-n}) \not\succ_f b$. By F_2, $(a_n, \mathbf{0}) \not\succ_f (a_n, \mathbf{a}_{-n})$ and hence (using F_1) $(a_n, \mathbf{0}) \not\succ_f b$. By F_3^N, there exists $\widetilde{a}_m$ such that $(\widetilde{a}_m, a_n, \mathbf{0}) \sim_f \mathbf{b}$. By F_2 again, $(\widetilde{a}_m, a_n, \mathbf{z}) \succeq_f \mathbf{b}$ for all $\mathbf{z} \in \mathbb{R}_+^{N-2}$. Therefore, by *Continuity*, there must be $a_m \in \mathbb{R}_+$ for which $(a_m, a_n, \mathbf{a}_{-m,n}) \sim_f \mathbf{b}$. We can then apply:

Theorem (Luce and Tukey [1964]) *Pairwise Preferential Independence and Solvability imply the existence of an additive representation of $\succ_f$.*

The proof of this theorem can be found in Kranz et al (1971). We illustrate the idea for the case $N = 3$ by showing that F_4^N implies F_4 for (without loss of generality) the pair of individuals 1 and 2, independent of the payoff to individual 3:

For any $\left(a_1^0, a_2^0, a_3^0\right)$ and any a_1^1, define a_2^1 and a_3^1 such that

$$\left(a_1^1, a_2^0, a_3^0\right) \sim_f \left(a_1^0, a_2^1, a_3^0\right) \sim_f \left(a_1^0, a_2^0, a_3^1\right).$$

Applying F_4^N twice implies that

$$\left(a_1^1, a_2^1, a_3^0\right) \sim_f \left(a_1^1, a_2^0, a_3^1\right) \sim_f \left(a_1^0, a_2^1, a_3^1\right).$$

For any a_1^2, define a_2^2 and a_3^2 such that

$$\left(a_1^2, a_2^0, a_3^0\right) \sim_f \left(a_1^0, a_2^2, a_3^0\right) \sim_f \left(a_1^0, a_2^0, a_3^2\right) \sim_f \left(a_1^1, a_2^1, a_3^0\right).$$

We have to show that $\left(a_1^2, a_2^1, a_3\right) \sim_f \left(a_1^1, a_2^2, a_3\right)$ for any value of a_3: $\left(a_1^2, a_2^0, a_3^0\right) \sim_f$ $\left(a_1^1, a_2^0, a_3^1\right)$, so by F_4^N also $\left(a_1^2, a_2^1, a_3^0\right) \sim_f \left(a_1^1, a_2^1, a_3^1\right)$. Similarly $\left(a_1^0, a_2^2, a_3^0\right) \sim_f \left(a_1^0, a_2^1, a_3^1\right)$, so by F_4^N also $\left(a_1^1, a_2^2, a_3^0\right) \sim_f \left(a_1^1, a_2^1, a_3^1\right)$. Using transitivity, $\left(a_1^2, a_2^1, a_3^0\right) \sim_f \left(a_1^1, a_2^2, a_3^0\right)$ and by F_4^N this is independent of a_3^0. Hence $\left(a_1^2, a_2^1, a_3\right) \sim_f \left(a_1^1, a_2^2, a_3\right)$ for any value of a_3.

The existence of utility functions according to which $\succ_f$ is represented by the Nash product follows, as before, where additivity is implied by Luce and Tukey's theorem. We gave the intuition for the remainder of the proof of Theorem 4 after stating Theorem 2.

2.5.2   An Application to Obfuscation by a Social Decision Maker

It is often argued that individuals who make social choices are faced with very rigid constraints. Shame at acting against the interests of others could be one such constraint, moderating individuals' decisions as compared to their selfish interest. We build on this interpretation to explain why a social decision maker may implement policies with relatively opaque consequences. To first illustrate by simple example why such lack of transparency (or obfuscation) might be valuable to DM at all, consider an indivisible good that can be assigned to one individual. All individuals have the same probability of needing it the most. Under obfuscation, this uncertainty never gets resolved, hence all allocations are equally fair and DM can take the good for herself without shame. If, on the other hand, the uncertainty does get resolved before DM chooses an allocation, she can only claim the good without shame in the event that she values it the most.

The literature that studies obfuscation in policy making usually considers redistributive policies. As an example, Tullock (1983) uses the decision of where to locate a new road: Depending on the road's location, some citizens will gain, others might lose. These consequences will not be entirely transparent at the time of decision making.

While building a road in a certain location clearly has a redistributive component, we assert that it may also generate value for the society as a whole. In this section, we therefore consider more general policies, which carry both an uncertain social value and an uncertain distribution of gains among citizens. All citizens (including DM) have identical information with respect to both types of uncertainty at every stage of the process.[24] Evaluating poli-

---

[24]This assumption stands in contrast to the usual asymmetric-information assumption (either among

cies requires some degree of public deliberation, for example, the consultation of experts. Before this deliberation takes place, DM can limit the degree to which deliberation will resolve uncertainty. She does so by choosing the transparency level of the policies that will be considered.[25] We assume that even for the lowest feasible transparency level, deliberation will reveal DM's selfish payoff from each relevant policy. This assumption is intuitive due to DM's arguably exposed role. It is also appropriate when addressing DM, who is constrained by shame. Finally, it is crucial for the results established below, as the assumption introduces an asymmetry between DM and all other citizens despite the information structure assumed above: The probability that DM's preferences will become public is larger than that of any other citizen's. Therefore, when deciding on the transparency of policies, DM has to trade off the benefit of obfuscation, which makes selfish choices seem more fair, and the value of transparency, which reveals efficient choices as such.

The time sequence is as follows: Firstly, DM chooses the transparency level for the policies under consideration. Secondly, public deliberation symmetrically reduces the uncertainty about the consequences of those policies. The higher the transparency level was set, the less uncertainty remains. Lastly, DM chooses one policy. It is important to note that, in slight contrast to our model, stage-one choice does not alter, in terms of expected payoffs, the set of policies that are relevant for stage-two choice. Instead, it alters the expected differential in fairness between the policies in which DM will have a selfish interest, and those that will be perceived as fairest.

Formally, consider a very large population of N individuals indexed by i. Let $\Omega = \mathbb{R}$ be identified both as the set of possible policy choices $a \in \Omega$ and as the type space. When referring to a particular individual i, we denote her type as $x_i \in \Omega$. Individual i's selfish preferences are commonly known to be represented by $u : \mathbb{R}^+ \to \mathbb{R}^+$, which is continuous and strictly decreasing with the (standard Euclidian) distance, $d(a, x_i)$, between the implemented policy a and her type x_i. Types are identically and independently distributed according to a Normal distribution with an unknown mean, $\theta \in \mathbb{R}$, and known variance

$\sigma^2 > 0$. The (conjugate) common prior distribution of θ is Normal, $\theta \sim N\left(\overline{\theta_0}, \overline{\nu_0^2}\right)$, where $\overline{\nu_0^2} := \nu^2$ and, without loss of generality, $\overline{\theta_0} = 0$. Thus, σ captures the uncertainty about the redistributive consequences of policies, while ν relates to the uncertainty about the value generated for society as a whole. Let Ω^n be the set of all possible type profiles of length $n \leq N$, with a typical element $x^n = (x_1, .., x_n)$. Upon observing the realization $x^n \in \Omega^n$, each individual, including DM, updates her beliefs according to Bayes' rule. The resulting common posterior distribution of θ is Normal as well, $\theta \sim N\left(\overline{\theta_n}, \overline{\nu_n^2}\right)$, with $\overline{\theta_n} = \frac{\sigma^2(x_1+...+x_n)}{\nu^2+n\sigma^2}$ and $\overline{\nu_n^2} = \frac{\nu^2\sigma^2}{\nu^2+n\sigma^2}$. Note that $x_1 \in x^n$. Since we identify DM as individual 1, her type is always revealed for $n \geq 1$.

After the entire population observes x^n, DM makes a social choice $a \in \Omega$. We assume that DM's preferences satisfy $P_1 - P_5$ as well as F_1, F_2, F_3^N and F_4^N. As is implied by Theorem 4, DM would like to choose a to maximize

$$u\left(d\left(a, x_1\right)\right) + \beta_N \prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right) - \max_{b\in\Omega}\left[\beta_N \prod_{i=1}^{N} u\left(d\left(b, x_i\right)\right)\right],$$

where, for simplicity, we assume h to be linear ($h(x) = \beta_N x$). After observing x^n, however, $x_{n+1}, ..., x_N$ remain unknown, so $\prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right)$ cannot be evaluated. To accommodate the uncertainty about the distribution of types in the population, we assume instead that the *expected* fairness of an allocation conditional on x^n,

$$E\left[\prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right) | x^n\right]$$

is what determines shame. Thus, the relevant characteristics of a policy are both its proximity $d\left(a, x_1\right)$ to DM's type x_1 and its expected fairness. According to our representation, the action that the public perceives as the fairest is

$$a^* := \arg\max_{a\in\Omega} E\left[\prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right) | x^n\right].$$

DM's choice is then governed by maximizing the term

$$u\left(d\left(a, x_1\right)\right) + \beta_N \left(E\left[\prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right) | x^n\right] - E\left[\prod_{i=1}^{N} u\left(d\left(a^*, x_i\right)\right) | x^n\right]\right).$$

Note that for fixed n and large population size, $N \to \infty$, $a^* \longrightarrow \overline{\theta_n}$.

Before x^n is observed by both DM and the public, DM can pick $n \in [1, \overline{n}]$, $\overline{n} << N$. The number n is interpreted as the transparency level of policies in Ω: The more transparent the policies are, the larger the number of individuals whose type becomes revealed by the public deliberation. Different transparency levels n introduce different distributions over expected fairness, while leaving $d(a, x_i)$ unchanged. Thus, the choice of n is equivalent to stage-one choice of different distributions over menus with the same cardinality that differ in the expected fairness of their elements. In contrast to the policy choice, we assume that the choice of the transparency level, n, is free of shame.[26] This assumption could rest on the fact that the transparency level is chosen before any uncertainty is resolved and, hence, cannot bias the ex post expected fairness of any policy. Alternatively, the public might simply be unaware of transparency as a choice dimension.

We are interested in finding the optimal transparency level, that is, the optimal first-stage choice according to DM's selfish preferences. DM faces the following trade-off: On the one hand, she benefits from high transparency, which reduces the uncertainty about the fairness of policies and allows the public to interpret fair choices as such. On the other hand, she benefits from low transparency, as it gives her selfish payoff more weight in public observation, limiting the public's ability to detect selfish behavior.

To determine DM's optimal transparency choice, n^*, define the ratio of the standard deviations σ and ν as $s := \frac{\sigma}{\nu}$ and let $s = s^*$ solve $2 + 3s^2 - 3s^4 - 6s^6 - 2s^8 = 0$, $s^* \approx 0.84$. Ignoring the integer constraint on n, we state:

Proposition $n^*(s)$ *exists and is unique. For* $s < s^*$, $m = n^*(s)$ *is the solution to* $2 + (2m + 1)s^2 - 3s^4 - 2(2m + 1)s^6 - m(m + 1)s^8 = 0$, *which is decreasing in* s. *For* $s \geq s^*$, $n^*(s) = 1$.

Note that the optimal transparency level does not depend on DM's susceptibility to shame, β_N.[27] This means that while for the case of a standard economic agent with $\beta_N = 0$,

---

[26] Due to this assumption, the model nicely fits our general framework.

[27] The proposition is only concerned with the transparency choice. The allocation DM chooses in the second stage obviously does depend on β_N, as it determines the extent to which DM yields to shame at the

the choice of n is arbitrary, even a small positive β_N implies the same unique amount of obfuscation as an arbitrarily large β_N does. Note as well that absolute uncertainty is irrelevant for the optimal transparency choice, only relative uncertainty $s = \frac{\sigma}{\nu}$ matters. This makes the prediction of the proposition very robust.

The proof of the proposition is in the appendix. It establishes that DM's utility is decreasing in the absolute value of the random variable $X_n := \overline{\theta_n} - x_1$, and that X_n is normally distributed. Then $\frac{p(|X_n|=z)}{p(|X_m|=z)}$ can be shown to satisfy the Monotone Likelihood Ratio Property (MLRP). Since DM's utility is decreasing in z, she strictly prefers m over n if and only if $\frac{p(|X_n|=z)}{p(|X_m|=z)}$ is increasing in z. Assuming $n > m$, we find, with some straightforward algebra, that this is the case if and only if $2 + (m+n) s^2 - 3s^4 - 2(m+n) s^6 - mns^8 < 0$. Thus DM has well-defined preferences over levels of transparency, n, and these preferences depend only on s. We then establish that $n^*(s)$ is unique and for $s < s^*$ is also decreasing. As a result, if DM prefers $n = 1$ to $n = 2$, then $n^*(s) = 1$ is her globally preferred transparency level. If she prefers n to both $n-1$ and $n+1$, then $n^*(s) = n$ is her globally preferred level.

Before interpreting this proposition, consider again the trade-off DM faces: Since her utility is decreasing in the random variable $|X_n| := \left| \overline{\theta_n} - x_1 \right|$, she would like the mean of the public posterior on types to be as close as possible to her type. Since DM's type is always observed (and thus always affects the posterior), it has a higher expected weight in the posterior than any other citizen's, which is only observed with probability $\frac{n-1}{N}$. The lower n is, the greater the advantage DM has over other citizens. Lowering n, however, increases the weight that the common prior gets in the public posterior. Since types are correlated, this is not in DM's interest. Now consider the proposition in the context of this trade-off: DM prefers to consider more opaque policies for the future if the standard deviation of the distribution of benefits across the population, measured by σ, becomes larger compared to the uncertainty about which policy is socially optimal, which is measured by ν. Intuitively, in this case she is concerned about the situation where her selfish preferences conflict with considerations of fairness. Therefore, she would like her own selfish preferences to impact the public posterior as much as possible. Since her preferences always become public, she
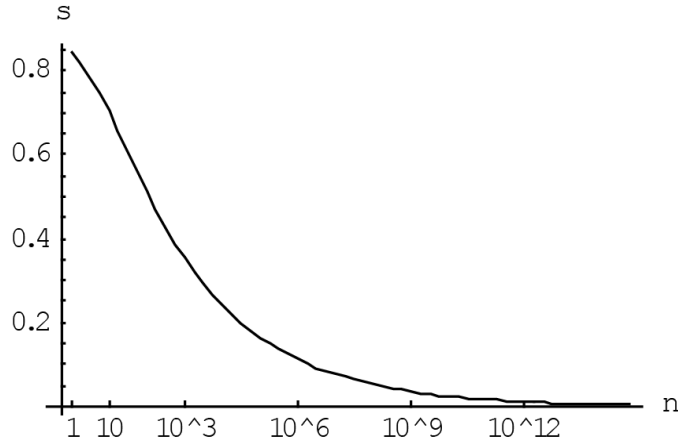
cost of her selfish interest.

Figure 2.2: n as a function of s.

would like other citizens' preferences to remain unobserved. Conversely, she prefers more transparent policies if the uncertainty is more about the socially optimal policy and less about the distribution of gains. In this case, she is mostly concerned about the situation where her selfish preferences are in line with considerations of fairness but an uninformed public would perceive her as selfish if she chose accordingly. Figure 2.2 shows $n^*(s)$.

2.6 Related Literature

Other-regarding preferences have been considered extensively in economic literature. In particular, inequality aversion as studied by Fehr and Schmidt (1999) is based on an objective function with a similar structure to the representation of second-stage choice in Theorem 3.[28] Both works attach a cost to any deviation from choosing the fairest alternative. In Fehr and Schmidt's work, the fairest allocation need not be feasible and is independent of the choice situation. In our work, the fairest allocation is always a feasible choice and it is identified through the axioms. This dependence of the fairest allocation on the choice situation allows us to distinguish observed from unobserved choice.

The idea that there may be a discrepancy between DM's preference to behave "pro-socially" and her desire to be viewed as behaving pro-socially is not new to economic

---

[28] Neilson (2006-a) axiomatizes a reference-dependent preference, that can be interpreted in terms of Fehr and Schmidt's objective function.

literature. For a model thereof, see Benabou and Tirole (2006).

Neilson's (2006-b) work is motivated by the same experimental evidence as ours. He also considers menus of allocations as objects of choice. Neilson does not axiomatize a representation result, but points out how choices among menus should relate to choices from menus, if shame were the relevant motive. He relates the two aspects of shame that also underlie the *Set Betweenness* property in our work; DM might prefer a smaller menu over a larger menu either because avoiding shame compels her to be generous when choosing from the larger menu, or because being selfish when choosing from the larger menu bears the cost of shame.

The structure of our representation resembles the representation of preferences with self-control under temptation, as axiomatized in GP. GP study preferences over sets of lotteries and show that their axioms lead to a representation of the following form:

$$U^{GP}(A) = \max_{a \in A}\left\{u^{GP}(a) + v^{GP}(a)\right\} - \max_{b \in A}\left\{v^{GP}(b)\right\}$$

with u^{GP} and v^{GP} both linear in the probabilities and where A is now a set of lotteries. In their context, u^{GP} represents the "commitment"- and v^{GP} the "temptation"-ranking. While the two works yield representations with a similar structure, their domains - and therefore the axioms - are different. In particular, the objects in GP's work are sets of lotteries. They impose the independence axiom and indifference to the timing of the resolution of uncertainty. This allows them to identify the representation above that consists of two functions that are linear in the probabilities. Each of these functions is an expected utility functional. The objects in our work, in contrast, are sets of allocations and there is no uncertainty. The natural way to introduce uncertainty to our model is to treat our representation as the utility function, which should be used to calculate the expected utility of lotteries over sets. Therefore, DM would typically not be indifferent to the timing of the resolution of uncertainty.[29] However, one of GP's axioms is the *Set Betweenness* axiom, $A \succcurlyeq B \Rightarrow A \succcurlyeq A \cup B \succcurlyeq B$. We show that our axioms *Strong Left Betweenness* (P_2), *Shame* (P_3) and *Fairness Ranking* (F_1) imply *Set Betweenness*. Hence, GP's Lemma 2 can be

---

[29] In section 5.2 we account for uncertainty, which can be translated into uncertainty over sets.

employed, allowing us to confine attention to sets with only two elements.

Our model is positive in nature, but it is interesting to contrast moral or normative elements in its interpretation with those in the context of the temptation literature: In a work related to GP, Dekel, Lipman and Rustichini (2005) write: "*...by 'temptation' we mean that the agent has some view of what is normatively correct, what she should do, but has other, conflicting desires which must be reconciled with the normative view in some fashion.*" According to this interpretation, the commitment ranking is given a normative value. In our work, shame is based on deviating from some fairness norm that tells DM what she should do. This norm conflicts with DM's selfish commitment ranking.

Empirically, the assumption that only two elements of a choice set matter for the magnitude of shame (the fairest available alternative and the chosen alternative) is clearly simplifying: Oberholzer-Gee and Eichenberger (2004) observe that dictators choose to make much smaller transfers when their choice set includes an unattractive lottery. In other words, the availability of an unattractive allocation seems to lessen the incentive to share.

Lastly, it is necessary to qualify our leading example: The experimental evidence on the effect of (anonymous) observation on the level of giving in dictator games is by no means conclusive. Behavior tends to depend crucially on surroundings, like the social proximity of the group of subjects and the phrasing of the instructions, as, for example, Bolton, Katok and Zwick (1994); Burnham (2003); and Haley and Fessler (2005) record. On the one hand, there is more evidence in favor of our interpretation: In a follow-up to the experiment cited in the introduction, Dana et al (2006) verify that dictators do not exit the game if second-stage choice is also unobserved. Similarly, Pillutla and Murningham (1995) find evidence that people's giving behavior under anonymity depends on the information given to the observing recipient. In experiments related to our leading example, Lazear, Malmendier and Weber (2005) as well as Broberg, Ellingsen and Johannesson (2008) even predict and find that the most generous dictators are keenest to avoid an environment where they could share with an observing recipient.[30] Broberg et al further elicit the price subjects are willing to pay in order to exit the dictator game, finding that the mean exit reservation price equals 82% of the dictator game endowment. On the other hand, this is in contrast

---

[30]This nicely underlines our interpretation of "shame" as a motive.

to evidence collected by Koch and Normann (2005), who claim that altruistic behavior persists at an almost unchanged level when observability is credibly reduced. Similarly, Johannesson and Persson (2000) find that incomplete anonymity - not observability - is what keeps people from being selfish. Ultimately, experiments aimed at eliciting a norm share the same problem: Since people use different (and potentially contradictory) norms in different contexts, it is unclear whether the laboratory environment triggers a different set of norms than would other situations: Frohlich, Oppenheimer and Moore (2000) point out that money might become a measure of success rather than a direct asset in the competition-like laboratory environment, such that the norm might be "do well" rather than "do not be selfish."[31] More theoretically, Miller (1999) suggests that the phrasing of instructions might determine which norm is invoked. For example, the reason that Koch and Normann do not find an effect of observability might be that their thorough explanation of anonymity induces a change in the regime of norms, in effect telling people "be rational," which might be interpreted as "be selfish." Then being observed might have no effect on people who, under different circumstances, might have been ashamed to be selfish.

2.7 Conclusion

We study a decision maker who cares about others' well-being only when being observed. We term the motive that distinguishes choice behavior when observed from choice when unobserved "shame." Theorem 1 features a representation that captures the tension between the interest to maximize private payoff and the shame that results from not choosing the fairest alternative within a set. Theorem 2 identifies a set-independent choice criterion with the help of a separability axiom. If there is a set-independent choice criterion, the representation should be more tractable for applications. More importantly, the separability assumption provides a criterion on preferences over sets to decide whether or not the period-two choice of alternatives might look as if it was generated by an altruistic concern. In Theorem 3 we further specify the norm of fairness. We show that the fairest alternative in a

---

[31]Surely the opposite is also conceivable: Subjects might be particularly keen to be selfless when the experimentor observes their behavior. This example is just ment to draw attention to the difficulties faced by experimenters in the context of norms.

set can be interpreted as the Nash Bargaining Solution of an associated game. Because the utility functions used to generate this game are private, so is the norm. The most appealing interpretation relies on the descriptive power of the NBS in many bargaining situations, giving it normative appeal as the solution to a symmetric mechanism. Lastly, Theorem 4 extends Theorem 2 to situations where DM faces multiple other individuals whose welfare depends on her choice. We apply our model to a social decision maker, whose selfish utility is correlated with fairness. She faces a trade-off when choosing the transparency of her policies: Being more transparent makes it easier for the public to perceive fair choices as such, while less transparency makes it harder for society to detect selfish choices. In our setup, the optimal transparency level is unique and is independent of DM's susceptibility to shame.

Let us conclude with another experiment to suggest how to incorporate uncertainty into our model. Dana, Weber and Kuang (2005) make a dictator face a choice between $5 and $6 for herself. An anonymous recipient will receive either $5 or $1. Which recipient payoff is paired with which dictator payoff is determined by a coin flip. The dictator can reveal (without being observed) the outcome of the coin flip prior to her decision. The authors find that many dictators choose not to reveal the outcome. This action seems weakly dominated, because whether or not the dictator is willing to give up $1 in order to give the recipient the extra $4, knowing whether such a trade-off is necessary should not hurt DM. We propose to interpret the revealed and the unrevealed conditions as two different choice situations. If all functions in the combination of Theorem 2 and Theorem 3 are identities, and if DM subscribes to the vNM axioms, the utilities to be compared are

$$U\left(\{(6,3),(5,3)\}\right) = 6$$

versus

$$\frac{1}{2}U\left(\{(6,5),(5,1)\}\right) + \frac{1}{2}U\left(\{(6,1),(5,5)\}\right) = \frac{1}{2}6 + \frac{1}{2}5 = 5.5$$

This could explain the observed behavior. However, since in the experiment the recipient knows the full instructions and does not observe DM's decision to reveal, observability

would require a more involved interpretation. To ease the application of our model, it would be interesting to see how the subject's behavior changes if the recipient is only told the information DM has *after* her decision to reveal or not.

2.8 Appendix

2.8.1 Proof of Theorem 1

Let $U : K \to \mathbb{R}$ be a continuous function that represents $\succ$. Define $u(a_1) \equiv U(\{(a_1, 0)\})$. By P_1, $u(a_1) = U(\{(a_1, a_2)\})$ independent of a_2, with $u(a_1)$ continuous and strictly increasing.

Let $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ be a continuous function that represents $\succ_f$. By F_2, φ is also strictly increasing.

Claim 1.1 (Right Betweenness): $A \succeq B \Rightarrow A \cup B \succeq B$.

Proof: There are two cases to consider:

Case 1) $\forall a \in A$, $\exists b \in B$ such that $b \succ_f a$. Let $A = \{\mathbf{a}^1, \mathbf{a}^2 ..., \mathbf{a}^N\}$ and $C_0 = B$. Define $C_n = C_{n-1} \cup \{\mathbf{a}^n\}$ for $n = 1, 2, .., N$. According to F_1, there exists $b \in B$ such that $a^n \not\succ_f b$. By P_3, $C_{n-1} \not\succ C_n$. By negative transitivity of $\succ$, $C_0 \not\succ C_N$ or $A \cup B \succeq B$.

Case 2) $\exists a \in A$ such that $a \succ_f b$, $\forall b \in B$. Let $B = \{\mathbf{b}^1, \mathbf{b}^2 ..., \mathbf{b}^M\}$. Define $C_0 = A$ and $C_m = C_{m-1} \cup \{\mathbf{b}^m\}$ for $m = 1, 2, .., M$. By P_3, $\forall C$ such that $a \in C$, $C \not\succ C \cup \{\mathbf{b}^m\}$. Hence, $C_{m-1} \not\succ C_m$. By negative transitivity of $\succ$, $C_0 \not\succ C_M$ or $A \cup B \succeq A \succeq B$, hence $A \cup B \succeq B$. $\|$

Combining Claim 1.1 with P_2 guarantees Set Betweenness (SB): $A \succeq B \Rightarrow A \succeq A \cup B \succeq B$. Having established Set Betweenness, we can apply GP Lemma 2, which states that any set is indifferent to a specific two-element subset of it.

Lemma 1.1 (GP Lemma 2): *If $\succ$ satisfies SB, then for any finite set A, there exist $\mathbf{a}, \mathbf{b} \in A$ such that $A \sim \{\mathbf{a}, \mathbf{b}\}$, $(\mathbf{a}, \mathbf{b})$ solves $\max_{\mathbf{a}' \in A} \min_{\mathbf{b}' \in A} U(\{\mathbf{a}', \mathbf{b}'\})$ and $(\mathbf{b}, \mathbf{a})$ solves $\min_{\mathbf{b}' \in A} \max_{\mathbf{a}' \in A} U(\{\mathbf{a}', \mathbf{b}'\})$.*

Define $f : \mathbb{R}_+^2 \times \mathbb{R}_+^2 \to \mathbb{R}$ such that $f(\mathbf{a}, \mathbf{b}) = u(a_1) - \widetilde{U}(\mathbf{a}, \mathbf{b})$, where $\widetilde{U} : \mathbb{R}_+^2 \times \mathbb{R}_+^2 \to \mathbb{R}$ is a function satisfying:

$$U(\{\mathbf{a}, \mathbf{b}\}) = \max_{\mathbf{a}' \in \{\mathbf{a},\mathbf{b}\}} \min_{\mathbf{b}' \in \{\mathbf{a},\mathbf{b}\}} \widetilde{U}(\mathbf{a}', \mathbf{b}') = \min_{\mathbf{b}' \in \{\mathbf{a},\mathbf{b}\}} \max_{\mathbf{a}' \in \{\mathbf{a},\mathbf{b}\}} \widetilde{U}(\mathbf{a}', \mathbf{b}').^{32}$$

By definition we have $f(\mathbf{a}, \mathbf{a}) = 0$ for every $\mathbf{a} \in \mathbb{R}_+^2$. Note as well that

$$\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \Rightarrow f(\mathbf{a}, \mathbf{b}) > 0,$$

as otherwise we would have:

$$U(\{\mathbf{a}, \mathbf{b}\}) = \max \left\{ \begin{matrix} u(a_1) - \max \left\{ \begin{matrix} f(\mathbf{a},\mathbf{a})=0 \\ f(\mathbf{a},\mathbf{b}) \end{matrix} \right\} \\ u(b_1) - \max \left\{ \begin{matrix} f(\mathbf{b},\mathbf{a}) \\ f(\mathbf{b},\mathbf{b})=0 \end{matrix} \right\} \end{matrix} \right\} \geq u(a_1) - \max \left\{ \begin{matrix} f(\mathbf{a}, \mathbf{a}) = 0 \\ f(\mathbf{a}, \mathbf{b}) \end{matrix} \right\} = U(\{\mathbf{a}\}).$$

For a decision maker who is not susceptible to shame, $U(\{\mathbf{a}, \mathbf{b}\}) = \max\{u(a_1), u(b_1)\}$. Hence setting $f(\mathbf{a}, \mathbf{b}) \equiv 0$ is consistent with her preferences. The following claims help us to further identify f for a decision maker who is susceptible to shame.

Claim 1.2: (i) $[\varphi(\mathbf{a}) < \varphi(\mathbf{b})$ and $a_1 > b_1] \Leftrightarrow \{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\}$

(ii) $[\varphi(\mathbf{a}) < \varphi(\mathbf{b})$ and $a_1 \leq b_1] \Rightarrow \{\mathbf{a}\} \sim \{\mathbf{a}, \mathbf{b}\}$

(iii) $[\varphi(\mathbf{a}) = \varphi(\mathbf{b})$ and $a_1 > b_1] \Rightarrow \{\mathbf{a}\} \sim \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$.

Proof: (i) If $\varphi(\mathbf{b}) > \varphi(\mathbf{a})$ then there exists A such that $\mathbf{a} \in A$ and $A \succ A \cup \{\mathbf{b}\}$. As $a_1 > b_1 \Leftrightarrow \{\mathbf{a}\} \succ \{\mathbf{b}\}$, by P_2 $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\}$. Conversely if $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\}$, then $\mathbf{b} \succ_f \mathbf{a}$ and hence $\varphi(\mathbf{a}) < \varphi(\mathbf{b})$. Further from SB and P_1, $a_1 > b_1$.

(ii) If $a_1 \leq b_1$ then by SB $\{\mathbf{b}\} \succeq \{\mathbf{a}, \mathbf{b}\}$. Since $\varphi(\mathbf{b}) > \varphi(\mathbf{a})$, there is no B such that $\mathbf{b} \in B$ and $B \succ B \cup \{\mathbf{a}\}$, hence $\{\mathbf{b}\} \sim \{\mathbf{a}, \mathbf{b}\}$.

(iii) By P_1 $\{\mathbf{a}\} \succ \{\mathbf{b}\}$ and SB $\{\mathbf{a}\} \succeq \{\mathbf{a}, \mathbf{b}\}$. As $\varphi(\mathbf{a}) = \varphi(\mathbf{b})$, using (i) we have $\{\mathbf{a}\} \sim \{\mathbf{a}, \mathbf{b}\}$.‖

---

^{32}Note that $\max_{\mathbf{a} \in A} \min_{\mathbf{b} \in A} U(\{\mathbf{a}, \mathbf{b}\}) = \max_{\mathbf{a} \in A} \min_{\mathbf{b} \in A} \left[\max_{\mathbf{a}' \in \{\mathbf{a},\mathbf{b}\}} \min_{\mathbf{b}' \in \{\mathbf{a},\mathbf{b}\}} \widetilde{U}(\mathbf{a}', \mathbf{b}') \right] = \max_{\mathbf{a} \in A} \min_{\mathbf{b} \in A} \widetilde{U}(\mathbf{a}, \mathbf{b}).$

Let $(\mathbf{a}^* (A), \mathbf{b}^* (A))$ be the solution of

$$\max_{\mathbf{a}' \in A} \min_{\mathbf{b}' \in A} U \left(\{ \mathbf{a}', \mathbf{b}' \} \right)$$

so $(\mathbf{b}^* (A), \mathbf{a}^* (A))$ solves $\min_{\mathbf{b}' \in A} \max_{\mathbf{a}' \in A} U \left(\{ \mathbf{a}', \mathbf{b}' \} \right)$.

Claim 1.3: There exists $\mathbf{b} \in \arg\max_{\mathbf{a}' \in A} \varphi (\mathbf{a}')$ such that $A \sim \{\mathbf{a}', \mathbf{b}\}$ for some $\mathbf{a}' \in A$ and $\mathbf{b}^* (A) = \mathbf{b}$.

Proof: Assume not, then there exist $\mathbf{a}, \mathbf{c}$ such that $A \sim \{\mathbf{a}, \mathbf{c}\}$, $(\mathbf{a}, \mathbf{c}) = (\mathbf{a}^* (A), \mathbf{b}^* (A))$. Therefore,

$$\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}, \mathbf{c}\} \sim \{\mathbf{a}, \mathbf{b}, \mathbf{c}\} \sim A^{33} \forall \mathbf{b} \in \arg\max_{\mathbf{a}' \in A} \varphi \left(\mathbf{a}' \right)$$

and hence $\mathbf{c} \succ_f \mathbf{b}$, which is a contradiction. $\|$

For the remainder of the proof, let $I_f(\varphi) := \{\mathbf{b}' : \varphi (\mathbf{b}') = \varphi\}$. That is, $I_f(\varphi (\mathbf{b}))$ is the $\sim_f$ equivalence class of $\mathbf{b}$. Define

$$Y (\mathbf{a}, \varphi) = \left\{ \mathbf{b}' \in I_f (\varphi) : \{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}'\} \succ \{\mathbf{b}'\} \right\}$$

We make the following four observations:

(1) $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$, $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{c}\}$ and $\mathbf{b} \succ_f \mathbf{c}$ imply $\{\mathbf{a}, \mathbf{c}\} \succeq \{\mathbf{a}, \mathbf{b}\}$.

(2) $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$, $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{c}\} \succ \{\mathbf{c}\}$ and $\mathbf{b} \sim_f \mathbf{c}$ imply $\{\mathbf{a}, \mathbf{c}\} \sim \{\mathbf{a}, \mathbf{b}\}$.

(3) $\mathbf{b} \in Y (\mathbf{a}, \varphi)$, $\mathbf{b}' \sim_f \mathbf{b}$ and $\{\mathbf{b}\} \succ \{\mathbf{b}'\}$ imply $\mathbf{b}' \in Y (\mathbf{a}, \varphi)$.

(4) If $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$, $\{\mathbf{b}'\} \succ \{\mathbf{b}\}$ and $\mathbf{b}' \in I_f(\varphi (\mathbf{b}))$, then either $\{\mathbf{a}, \mathbf{b}'\} \sim \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}'\}$ or $\{\mathbf{a}, \mathbf{b}'\} \sim \{\mathbf{b}'\} \succeq \{\mathbf{a}, \mathbf{b}\}$.

To verify these observations, suppose first that (1) did not hold. Then $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}, \mathbf{c}\}$ and $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$, hence by SB $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ and therefore $\mathbf{c} \succ_f \mathbf{b}$, which is a con-

---

[33] Note that if $(\mathbf{a}, \mathbf{c})$ $((\mathbf{c}, \mathbf{a}))$ solves the maximin- (minimax-) problem over A, it clearly solves this problem over the subset $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ for all $\mathbf{b} \in A \backslash \{\mathbf{a}, \mathbf{c}\}$.

tradiction. If (2) did not hold, we would get a contradiction to $\mathbf{b} \sim_f \mathbf{c}$ immediately. Next suppose that (3) did not hold. Then $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\} \succ \{\mathbf{b}'\} \sim \{\mathbf{a}, \mathbf{b}'\}$. Note that by SB $\{\mathbf{b}\} \succeq \{\mathbf{b}, \mathbf{b}'\}$ and, applying SB again , $\{\mathbf{b}\} \succeq \{\mathbf{a}, \mathbf{b}, \mathbf{b}'\}$. But then $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}, \mathbf{b}, \mathbf{b}'\}$, contradicting $\mathbf{b}' \sim_f \mathbf{b}$. To verify (4), assume $\{\mathbf{a}, \mathbf{b}'\} \succ \{\mathbf{b}'\}$. Then by Claim 1.2 (i) $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}'\} \succ \{\mathbf{b}'\}$ and then by observation (2) $\{\mathbf{a}, \mathbf{b}'\} \sim \{\mathbf{a}, \mathbf{b}\}$. If on the other hand $\{\mathbf{a}, \mathbf{b}'\} \sim \{\mathbf{b}'\}$, then if $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}, \mathbf{b}'\}$, $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\}$ and SB imply $\{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{a}, \mathbf{b}, \mathbf{b}'\}$, a contradiction to $\mathbf{b}' \in I_f(\varphi(\mathbf{b}))$. Note that by Claim 1.3 we cannot have $\{\mathbf{b}'\} \succ \{\mathbf{a}, \mathbf{b}'\}$.∥

Next we claim that $\varphi(\mathbf{b})$ is a sufficient statistic for the impact of $\mathbf{b}$ on a two element set.

Claim 1.4: There exits a function $\widetilde{U}$ satisfying the condition specified above such that $\varphi(\mathbf{b}) > \varphi(\mathbf{a})$ implies $f(\mathbf{a}, \mathbf{b}) = g(\mathbf{a}, \varphi(\mathbf{b}))$, where $g : \mathbb{R}_+^2 \times \mathbb{R} \to \mathbb{R}$ is weakly increasing in its second argument.

Proof: Such $\widetilde{U}$ exists, if and only if $f(\mathbf{a}, \mathbf{b}) = g(\mathbf{a}, \varphi(\mathbf{b}))$ is consistent with $\succ$. Therefore it is enough to consider the constraints $\succ$ puts on f. Given $\mathbf{a}$ and $\mathbf{b}$, look at all $\mathbf{c}$ such that $\varphi(\mathbf{b}) > \varphi(\mathbf{c})$. We should show that $f(\mathbf{a}, \mathbf{b}) \geq f(\mathbf{a}, \mathbf{c})$.

First note that if $\varphi(\mathbf{b}) \geq \varphi(\mathbf{a}) \geq \varphi(\mathbf{c})$, then $f(\mathbf{a}, \mathbf{b}) \geq 0 \geq f(\mathbf{a}, \mathbf{c})$ is consistent with $\succ$. If $\varphi(\mathbf{a}) \geq \varphi(\mathbf{b}) > \varphi(\mathbf{c})$, then $0 \geq f(\mathbf{a}, \mathbf{b}) \geq f(\mathbf{a}, \mathbf{c})$ is consistent with $\succ$. If $a_1 = 0$, then $f(\mathbf{a}, \mathbf{b}) \geq f(\mathbf{a}, \mathbf{c}) \geq 0$ is consistent with $\succ$. Therefore, confine attention to the case where $a_1 > 0$ and $\varphi(\mathbf{b}) > \varphi(\mathbf{c}) > \varphi(\mathbf{a})$.

By Claim 1.2 (i), F_2 and F_3, there exists $\mathbf{b}' \in I_f(\varphi(\mathbf{b}))$ such that $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}'\}$. Thus, there are two cases to consider:

1) $Y(\mathbf{a}, \varphi(\mathbf{b})) \neq \emptyset$.

2) $Y(\mathbf{a}, \varphi(\mathbf{b})) = \emptyset$.

Case 1) Suppose $Y(\mathbf{a}, \varphi(\mathbf{b})) \neq \emptyset$. Define $f(\mathbf{a}, \mathbf{b}) := f(\mathbf{a}, \mathbf{b}')$ for some $\mathbf{b}' \in Y(\mathbf{a}, \varphi(\mathbf{b}))$ (note that by observation (2) $f(\mathbf{a}, \mathbf{b}') = f(\mathbf{a}, \mathbf{b}'') \ \forall \mathbf{b}', \mathbf{b}'' \in Y(\mathbf{a}, \varphi(\mathbf{b}))$ and using observations (3) and (4), this definition is consistent with $\succ$.) If $Y(\mathbf{a}, \varphi(\mathbf{c})) \neq \emptyset$ then by observation (1) $\{\mathbf{a}, \mathbf{c}\} \succeq \{\mathbf{a}, \mathbf{b}\}$ and hence $f(\mathbf{a}, \mathbf{b}) \geq f(\mathbf{a}, \mathbf{c})$. If $Y(\mathbf{a}, \varphi(\mathbf{c})) = \emptyset$ then $\forall \mathbf{c}' \in I_f(\mathbf{c})$, $\{\mathbf{a}, \mathbf{c}'\} \sim \{\mathbf{c}'\}$. By F_2 and continuity of $\succ_f$, there exists $\mathbf{c}' \in I_f(\mathbf{c})$

with $c_1' < b_1'$ for some $\mathbf{b}' \in Y(\mathbf{a}, \varphi(\mathbf{b}))$. Then by Claim 1.1, P_1 and observation (1) $\{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{c}'\} \succeq \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\} \succ \{\mathbf{c}\}$, so $\mathbf{c}' \in Y(\mathbf{a}, \varphi(\mathbf{c}))$. Contradiction.

Case 2) Suppose $Y(\mathbf{a}, \varphi(\mathbf{b})) = \emptyset$. Define $f(\mathbf{a}, \mathbf{b}) := u(a_1) - u(0)$. If $Y(\mathbf{a}, \varphi(\mathbf{c})) \neq \emptyset$, then $f(\mathbf{a}, \mathbf{c}) < u(a_1) = f(\mathbf{a}, \mathbf{b})$. If $Y(\mathbf{a}, \varphi(\mathbf{c})) = \emptyset$ then $f(\mathbf{a}, \mathbf{c}) = u(a_1) = f(\mathbf{a}, \mathbf{b})$. ‖

Let $S := \{(\mathbf{a}, \varphi) : Y(\mathbf{a}, \varphi) \neq \emptyset\}$. Note that S is an open set.

Claim 1.5: There is $g(\mathbf{a}, \varphi)$, which is continuous.

Proof: If $Y(\mathbf{a}, \varphi) \neq \emptyset$, then $g(\mathbf{a}, \varphi) = u(a_1) - U(\{\mathbf{a}, \mathbf{b}\})$ for some $\mathbf{b} \in Y(\mathbf{a}, \varphi)$, and is clearly continuous. If $Y(\mathbf{a}, \varphi) = \emptyset$, then $\varphi \leq \varphi(\mathbf{a})$ implies $g(\mathbf{a}, \varphi) \leq 0$, while $\varphi > \varphi(\mathbf{a})$ implies $g(\mathbf{a}, \varphi) \geq u(a_1) - u(0)$. Define a switch point $(\widehat{\mathbf{a}}, \widehat{\varphi})$ to be a boundary point of S such that there exists $\widehat{\mathbf{b}} \in \mathbb{R}_+^2$ with $\varphi(\widehat{\mathbf{b}}) = \widehat{\varphi}$. For $\widehat{\varphi} = \varphi(\widehat{\mathbf{a}})$ define $g(\widehat{\mathbf{a}}, \widehat{\varphi}) := 0$ and for $\widehat{\varphi} > \varphi(\widehat{\mathbf{a}})$ define $g(\widehat{\mathbf{a}}, \widehat{\varphi}) := u(\widehat{a}_1) - u(0)$.

Consider a sequence $\{(\mathbf{a}^n, \varphi^n)\} \to (\widehat{\mathbf{a}}, \widehat{\varphi})$ in S. Pick a sequence $\{\mathbf{b}^{n'}\}$ with $\mathbf{b}^{n'} \in Y(\mathbf{a}^n, \varphi^n) \, \forall n$. Define $\{b_1^n\} = \left\{ \min\left[\frac{1}{n}, b_1^{n'}, \widehat{b}_1\right] \right\}$. Define b_2^n to be a solution to $\varphi(b_1^n, b_2^n) = \varphi^n$. By F_2 and F_3, b_2^n is well defined. Note that by observation (3) $\mathbf{b}^n = (b_1^n, b_2^n) \in Y(\mathbf{a}^n, \varphi^n)$. Lastly, let $\widehat{b}_1^n \equiv b_1^n$ and $\widehat{b}_2^n$ be the solution to $\varphi\left(\widehat{b}_1^n, \widehat{b}_2^n\right) = \widehat{\varphi}$. We have $U(\{\mathbf{a}^n, \mathbf{b}^n\}) = u(a_1^n) - g(\mathbf{a}^n, \varphi^n)$. If in the switch point $\widehat{\varphi} = \varphi(\widehat{\mathbf{a}})$, then $U\left(\left\{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}^n\right\}\right) = u(\widehat{a}_1)$. By continuity, $U(\{\mathbf{a}^n, \mathbf{b}^n\}) - U\left(\left\{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}^n\right\}\right) \underset{n \to \infty}{\to} 0$, hence

$$\lim_{n \to \infty} g(\mathbf{a}^n, \varphi^n) = \lim_{n \to \infty} [u(a_1^n) - u(\widehat{a}_1)] = u(\widehat{a}_1) - u(\widehat{a}_1) = 0 = g(\widehat{\mathbf{a}}, \widehat{\varphi}).$$

If in the switch point $\widehat{\varphi} > \varphi(\widehat{\mathbf{a}})$, then $U\left(\left\{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}^n\right\}\right) = u\left(\widehat{b}_1^n\right) = u(b_1^n)$. By the same continuity argument

$$\lim_{n \to \infty} g(\mathbf{a}^n, \varphi^n) = \lim_{n \to \infty} [u(a_1^n) - u(b_1^n)] = u(\widehat{a}_1) - u(0) = g(\widehat{\mathbf{a}}, \widehat{\varphi}).$$

For $\varphi < \varphi(\mathbf{a})$ let $g(\mathbf{a}, \varphi) < 0$. This satisfies the constraint on f. So g can be continuous in both arguments and increasing in φ and such that for any sequence $\{(\mathbf{a}^n, \varphi^n)\}$ in S, with

$\{(\mathbf{a}^n, \varphi^n)\} \to (\widehat{\mathbf{a}}, \widehat{\varphi})$, we have $\lim_{n \to \infty} g(\mathbf{a}^n, \varphi^n) = 0.\|$

That the representation satisfies the axioms is easy to verify. This completes the proof of Theorem 1.[34]$\blacksquare$

2.8.2 Proof of Theorem 2

Theorem 2 and Theorem 4 (i) are analogous, where Theorem 2 covers the case $N = 2$, while Theorem 4 (i) covers the case $N \geq 3$. We prove Theorem 4 (i) below by first establishing that the analogous version of Theorem 1 holds. From there on the proof of Theorem 2 is identical to the proof of Theorem 4 (i), with a_2 substituted for $\mathbf{a}_{-1}$.

2.8.3 Proof of Theorem 3

Luce and Tukey [1964] prove the necessity and sufficiency of Solvability (which is implied by Negative Transitivity, Weak Solvability, Pareto and Continuity (apply corollary 1 in the text to the case N=2)) and the Corresponding Trade-offs Condition (the label they use for F_4) to admit an additive representation.[35] To see how a proof works, consider the Lock-Step Procedure,[36] as illustrated by Figure 2.3:

By F_2, $\succ_f$ indifference curves are downward sloping and continuous. Fix $\left(a_1^0, a_2^0\right)$ and $a_2^1 > a_2^0$. Recursively construct a flight of stairs between the indifference curves through $\left(a_1^0, a_2^0\right)$ and $\left(a_1^0, a_2^1\right)$.

In the direction of increasing a_2 (and hence decreasing a_1) :

a_1^n solves $(a_1^n, a_2^n) \sim_f \left(a_1^0, a_2^0\right)$. F_3 guarantees that a solution exists whenever $(0, a_2^n) \preceq_f \left(a_1^0, a_2^0\right)$. If $(0, a_2^n) \succ_f \left(a_1^0, a_2^0\right)$, the flight of stairs terminates.

a_2^{n+1} solves $\left(a_1^n, a_2^{n+1}\right) \sim_f \left(a_1^0, a_2^1\right)$. A solution exists by F_3, as $(a_1^n, 0) \prec_f \left(a_1^0, a_2^1\right)$ by F_2.

In the direction of decreasing a_2 (and increasing a_1):

a_1^{-n} solves $\left(a_1^{-n}, a_2^{-n+1}\right) \sim_f \left(a_1^0, a_2^1\right)$. A solution exists by F_3, as $\left(0, a_2^{-n+1}\right) \prec_f \left(a_1^0, a_2^1\right)$ by F_2.

---

[34]If F_2 and F_3 were only posed on $\mathbb{R}^2_{++}$ as suggested in section 3, we would have to choose $\widehat{b}_1 > 0$ and $b_1^n > 0$ to use these axioms. This is possible for any switch point other than $(\widehat{\mathbf{a}}, \widehat{\varphi}) = (\mathbf{0}, \varphi(0))$, for which continuity can be established easily.

[35]Their theorem is stated in section 5.1 of the text.
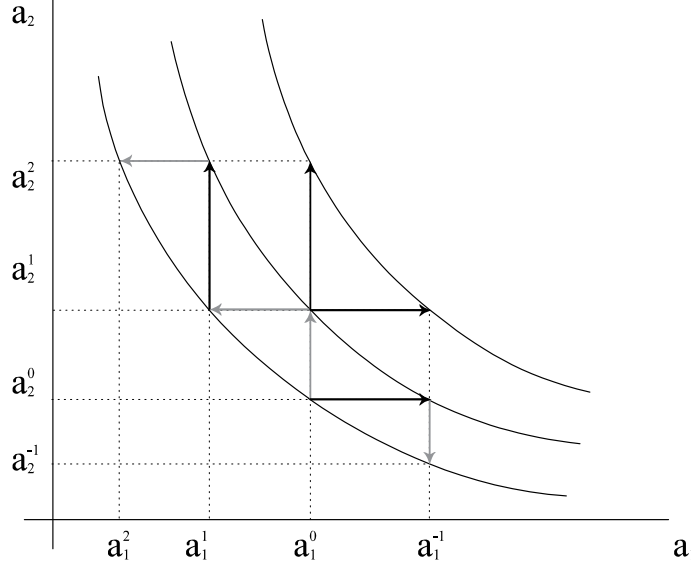
[36]See Keeney and Raiffa (1976).

Figure 2.3: Lock-Step Procedure, Constructing a flight of stairs.

a_2^{-n} solves $\left(a_1^{-n}, a_2^{-n}\right) \sim_f \left(a_1^0, a_2^0\right)$. F_3 guarantees that a solution exists whenever $\left(a_1^{-n}, 0\right) \preceq_f \left(a_1^0, a_2^0\right)$. If $\left(a_1^{-n}, 0\right) \succ_f \left(a_1^0, a_2^0\right)$, the flight of stairs terminates.

By construction $\left(a_1^{n+1}, a_2^{n+2}\right) \sim_f \left(a_1^n, a_2^{n+1}\right)$ and then by F_4, $\left(a_1^n, a_2^{n+2}\right) \sim_f \left(a_1^{n-1}, a_2^{n+1}\right)$. Thus we have constructed a discrete set of points on another indifference curve from the initial two curves. Repeating this procedure we can fill $\mathbb{R}_+^2$ with countable sets of points on countably many indifference curves.

Now consider a particular indifference curve that lies between two members of this set, as illustrated in Figure 2.4: Define $\left(a_1^{\frac{1}{2}}, a_2^{\frac{1}{2}}\right)$ implicitly by $\left(a_1^{\frac{1}{2}}, a_2^1\right) \sim_f \left(a_1^0, a_2^{\frac{1}{2}}\right)$ and $\left(a_1^{\frac{1}{2}}, a_2^{\frac{1}{2}}\right) \sim_f \left(a_1^0, a_2^0\right)$. Construct a flight of stairs between the indifference curves through $\left(a_1^0, a_2^{\frac{1}{2}}\right)$ and through $\left(a_1^0, a_2^0\right)$ as described above. Then we have in direction of decreasing a_2: $\left(a_1^{\frac{n+1}{2}}, a_2^{\frac{n+1}{2}}\right) \sim_f \left(a_1^{\frac{n}{2}}, a_2^{\frac{n}{2}}\right)$ and $\left(a_1^{\frac{n-1}{2}}, a_2^{\frac{n}{2}}\right) \sim_f \left(a_1^{\frac{n+1}{2}}, a_2^{\frac{n+2}{2}}\right)$. Therefore, by construction $\left(a_1^{\frac{n}{2}}, a_2^{\frac{n+1}{2}}\right) \sim_f \left(a_1^{\frac{n+1}{2}}, a_2^{\frac{n+2}{2}}\right)$ and then by F_4, $\left(a_1^{\frac{n-1}{2}}, a_2^{\frac{n+1}{2}}\right) \sim_f \left(a_1^{\frac{n}{2}}, a_2^{\frac{n+2}{2}}\right)$.

Proceed analogously in the direction of increasing a_2.

This demonstrates that if the vertical distance, measured in second component's units, between the indifference curves through $\left(a_1^0, a_2^0\right)$ and $\left(a_1^0, a_2^1\right)$ in a_1^n is the same as between those through $\left(a_1^0, a_2^1\right)$ and $\left(a_1^0, a_2^2\right)$ in a_1^{n-1}, then it is also the same between those through
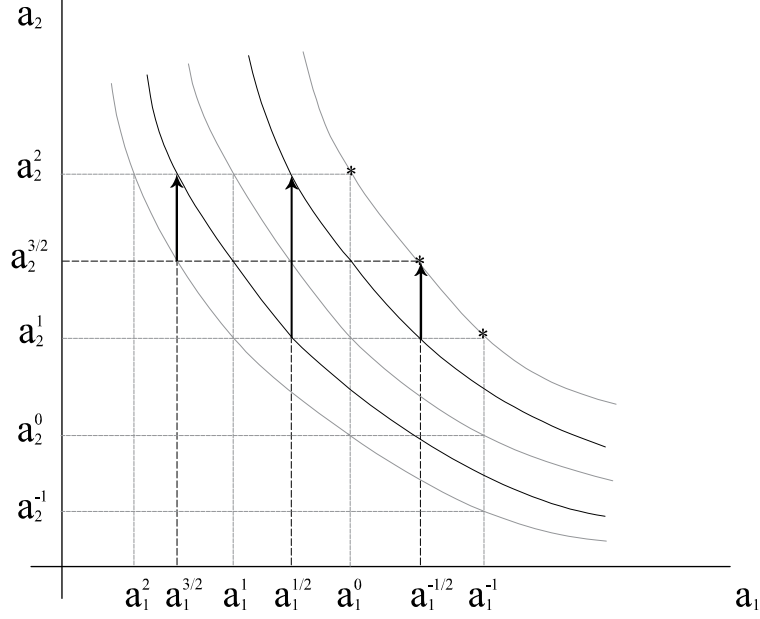
Figure 2.4: Lock-Step Procedure, Completing the indifference map.

$\left(a_1^0, a_2^0\right)$ and $\left(a_1^0, a_2^{\frac{1}{2}}\right)$ in $a_1^{\frac{n}{2}}$ and between those through $\left(a_1^0, a_2^{\frac{2}{2}}\right)$ and $\left(a_1^0, a_2^1\right)$ in $a_1^{\frac{n-1}{2}}$.
Repeating this procedure we can generate a dense set of points on indifference curves that are dense in $\mathbb{R}_+^2$. Then continuity of $\succ_f$ allows us to complete the entire map. Hence, if $(a_1, a_2) \sim_f (a_1', a_2')$ and $(a_1, \widetilde{a}_2) \sim_f (a_1', \widetilde{a}_2')$, then $(\widetilde{a}_1, a_2) \sim_f (\widetilde{a}_1', a_2') \Leftrightarrow (\widetilde{a}_1, \widetilde{a}_2) \sim_f (\widetilde{a}_1', \widetilde{a}_2')$.

As a result, we can create a mapping $a_2 \to \gamma(a_2)$ that transforms the original indifference map to be quasi-linear (vertically parallel indifference curves). The algorithm, which is formally described below, involves proceeding in infinitesimal steps and equalizing the step heights .

Set $\gamma(1) := 0$. To determine $\gamma(a_2)$ for $a_2 > 1$, pick an arbitrary a_1 and let a_1^0 solve $(a_1, a_2) \sim_f \left(a_1^0, 1 + \Delta\right)$, where Δ will be infinitesimal for the integration.[37] This solution exists by F_3. Then for every $a_2^* \in (1, a_2]$:[38]

Let a_1^* solve $(a_1^*, a_2^*) \sim_f \left(a_1^0, 1 + \Delta\right)$.

Let a_1^{**} solve $(a_1^{**}, a_2^* + \Delta) \sim_f \left(a_1^0, 1 + \Delta\right)$.

Let a_2' solve $(a_1^*, a_2') \sim_f \left(a_1^0, 1\right)$.

---

[37] As established above, the result of this mapping will be independent of the choice of a_1.

[38] The existence of solutions in the two cases below is guaranteed by the same reasoning as in the above discussion.

Let da_2' solve $(a_1^{**}, a_2' + da_2') \sim_f (a_1^0, 1)$.

Note that by F_2, $a_2' < a_2^*$ and $a_2' + da_2' < a_2^* + \Delta$.

Define implicitly $d\gamma(a_2^*) := \widetilde{\gamma}(a_2' + da_2') - \gamma(a_2')$, where

$$\widetilde{\gamma}(a) := \begin{cases} \gamma(a) & \text{for } a \leq a_2^* \\ \gamma(a_2^*) + a - a_2^* & \text{for } a > a_2^* \end{cases}$$

and then

$$\gamma(a_2) := \gamma(1) + \int_1^{a_2} d\gamma(a_2^*) = \int_1^{a_2} d\gamma(a_2^*).$$

Analogously determine $\gamma(a_2)$ for $a_2 < 1$: Pick an arbitrary a_1^0 and let a_1 solve $(a_1, a_2) \sim_f (a_1^0, 1)$. Then for every $a_2^* \in [a_2, 1)$:

Let a_1^* solve $(a_1^*, a_2^*) \sim_f (a_1^0, 1)$.

Let a_1^{**} solve $(a_1^{**}, a_2^* - \Delta) \sim_f (a_1^0, 1)$.

Let a_2' solve $(a_1^*, a_2') \sim_f (a_1^0, 1 + \Delta)$.

Let da_2' solve $(a_1^{**}, a_2' - da_2') \sim_f (a_1^0, 1 + \Delta)$.

Note that $a_2' < a_2^*$ and $a_2' + da_2' < a_2^* + \Delta$ by F_2.

Define implicitly $d\gamma(a_2^*) := \gamma(a_2') - \widetilde{\gamma}(a_2' - da_2')$, where

$$\widetilde{\gamma}(a) := \begin{cases} \gamma(a) & \text{for } a \geq a_2^* \\ \gamma(a_2^*) - a + a_2^* & \text{for } a < a_2^* \end{cases}$$

and

$$\gamma(a_2) := \gamma(1) + \int_1^{a_2} d\gamma(a_2^*) = -\int_{a_2}^1 d\gamma(a_2^*) < 0.$$

Then $\gamma : \mathbb{R}_+ \to \mathbb{R}$, is a continuous and increasing function. The $\succ_f$ indifference curves are quasi-linear with respect to $\gamma(a_2)$, so there is an increasing continuous function $\xi : \mathbb{R}_+ \to \mathbb{R}$, such that $\xi(a_1) + \gamma(a_2)$ generates the same indifference map. Hence re-defining

$$\varphi(a) := \xi(a_1) + \gamma(a_2)$$

represents $\succ_f$. Define

$$v_1(a_1) := \exp(\xi(a_1)) \text{ and } v_2(a_2) := \exp(\gamma(a_2)).$$

Then $v_1, v_2 : \mathbb{R}_+ \to \mathbb{R}_{++}$ are increasing and continuous and if we re-define, yet again, $\varphi(\mathbf{a}) := v_1(a_1) v_2(a_2)$, it represents $\succ_f$. By F_3, the functions v_1, v_2 must be unbounded.

That the representation satisfies the axioms is easy to verify. ■

2.8.4 Proof of Theorem 4

(i) The analogue of Theorem 1 can be established by substituting $\mathbf{a}_{-1}$ for a_2 in the theorem and in the proof, where now $\varphi : \mathbb{R}_+^N \to \mathbb{R}$.

Let φ be a representation of $\succ_f$. Let $\overline{\varphi} := \sup_{\mathbf{a} \in \mathbb{R}_+^N} \varphi(\mathbf{a})$ and $\underline{\varphi} := \inf_{\mathbf{a} \in \mathbb{R}_+^N} \varphi(\mathbf{a})$, if they are well defined. Otherwise, take $\overline{\varphi} = \infty$ and $\underline{\varphi} = -\infty$.

As before, let $S := \{(\mathbf{a}', \varphi') : Y(\mathbf{a}', \varphi') \neq \emptyset\}$. By F_3^N and the representation analogous to Theorem 1, $u(a_1) - u(0) > g(\mathbf{a}, \varphi)$ for $(\mathbf{a}, \varphi) \in S$.

Let $\succ_S$ be a binary relation on S defined by $(\mathbf{a}, \varphi) \succ_S (\widetilde{\mathbf{a}}, \widetilde{\varphi}) \Leftrightarrow \{\mathbf{a}, \mathbf{b}\} \succ \left\{\widetilde{\mathbf{a}}, \widetilde{\mathbf{b}}\right\}$ $\forall \mathbf{b} \in Y(\mathbf{a}, \varphi)$ and $\forall \widetilde{\mathbf{b}} \in Y\left(\mathbf{a}, \widetilde{\mathbf{b}}\right)$.

Define $U_S : \mathbb{R}_+^N \times \left(\underline{\varphi}, \overline{\varphi}\right) \to \mathbb{R}$ such that $U_S(\mathbf{a}, \varphi) := U(\mathbf{a}, \mathbf{b})$ for some $\mathbf{b} \in Y(\mathbf{a}, \varphi)$. By Theorem 1, $\succ_S$ is a weak order that can be represented by U_S. Note that the Consistency axiom (P_5) is relevant precisely on this domain. For $(\mathbf{a}, \varphi) \notin S$ define

$$U_S(\mathbf{a}, \varphi) := \begin{cases} 0 & \text{for } \varphi(\mathbf{a}) < \varphi \\ u(a_1) & \text{for } \varphi(\mathbf{a}) \geq \varphi \end{cases}.$$

Claim 4.1: U_S is continuous in all arguments.

Proof: Since the utility function is continuous on S, and because outside of S the function was chosen to be either a constant (hence continuous) or a continuous function, the only candidates for discontinuity are points on the boundary of S. There are two cases:

Case 1) $\varphi(\mathbf{a}) \geqslant \varphi$: Take $(\mathbf{a}, \varphi) \in bdr(S)$. Since $(\mathbf{a}, \varphi)$ is a boundary point, it must be

that $\varphi(\mathbf{a}) = \varphi$. Now let $\{\mathbf{a}^n, \varphi^n\}$ be a sequence in S which converges to $(\mathbf{a}, \varphi)$. By the definition of S, $U_s\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right) = u(a_1^n) - g\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right)$. Because preferences are continuous and using the properties of g from Theorem 1, we have $\lim_{n \to \infty} u(a_1^n) - g\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right) = u(a_1)$ as required.

Case 2) $\varphi(\mathbf{a}) < \varphi$: Take $(\mathbf{a}, \varphi) \in bdr(S)$. Again, let $\{\mathbf{a}^n, \varphi^n\}$ be an arbitrary sequence in S which converges to $(\mathbf{a}, \varphi)$. By the definition of S,

$$U_s\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right) = u(a_1^n) - g\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right) > \inf_{\mathbf{b}} \{u(b_1) : \varphi(\mathbf{b}) = \varphi^n \text{ and } b_1 < a_1^n\}.$$

Since $\succ$ is continuous, we have

$$
\begin{aligned}
\lim_{n \to \infty} u(a_1^n) - g\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right) &= u(a_1) - g((a_1, \mathbf{a}_{-1}), \varphi) \geq \\
\inf_{\mathbf{b}} \{u(b_1) : \varphi(\mathbf{b}) = \varphi \text{ and } b_1 < a_1\} &= u(0).
\end{aligned}
$$

where the last equality is implied by F_3^N. As $(\mathbf{a}, \varphi) \notin S$, we claim that $u(a_1) - g((a_1, \mathbf{a}_{-1}), \varphi) \leq \inf_{\mathbf{b}} \{u(b_1) : \varphi(\mathbf{b}) = \varphi \text{ and } \{\mathbf{b}\} \sim \{\mathbf{a}, \mathbf{b}\}\} = u(0)$. If not, then $u(a_1) - g((a_1, \mathbf{a}_{-1}), \varphi) = u(c_1) > u(0)$. But for any $\mathbf{c}$ with $c_1 > 0$, using F_3^N, we could find $\mathbf{c}'$ with $c_1' < c_1$ and $\varphi(\mathbf{c}') = \varphi(\mathbf{c})$. Using Theorem 1, this would imply that $(\mathbf{a}, \varphi) \in S$, which is a contradiction. Combining we have $\lim_{n \to \infty} u(a_1^n) - g\left((a_1^n, \mathbf{a}_{-1}^n), \varphi^n\right) = u(0)$, as required.‖

Definition: For $(\mathbf{a}, \varphi) \in S$, define $I_S(\mathbf{a}, \varphi) := \{(\mathbf{a}', \varphi') : (\mathbf{a}', \varphi') \sim_S (\mathbf{a}, \varphi)\} \subseteq S$. That is, $I_S(\mathbf{a}, \varphi)$ is the $\succ_S$ equivalence class of $(\mathbf{a}, \varphi)$.

Let $a_1^* : \mathbb{R}_+^2 \times (\underline{\varphi}, \overline{\varphi}) \to \mathbb{R}_+$ be the solution to

$$u(a_1^*(\mathbf{a}, \varphi)) = u(a_1) - g(\mathbf{a}, \varphi) = U_S(\mathbf{a}, \varphi).$$

a_1^* is the "first component equivalent" functional on S.[39] Since $u(a_1) > u(a_1) - g(\mathbf{a}, \varphi) > u(0)$ and $\succ_S$ is continuous, a_1^* is well defined and we have $(\mathbf{a}, \varphi) \succ_S (\widetilde{\mathbf{a}}, \widetilde{\varphi}) \Leftrightarrow a_1^*(\mathbf{a}, \varphi) >$

---

[39] Formally, $\forall \mathbf{x} \in \mathbb{R}_+^{N-1}$, $\{(a_1^*(\mathbf{a}, \varphi), \mathbf{x})\} \sim \{\mathbf{a}, \mathbf{b}\}, \forall \mathbf{b} \in Y(\mathbf{a}, \varphi)$

$a_1^* \left(\widetilde{\mathbf{a}}, \widetilde{\varphi} \right).$

Claim 4.2: The shame $g \left(\mathbf{a}, \varphi \right)$ is strictly increasing in φ.

Proof: Assume to the contrary that there is $\varphi' > \varphi$ and $\left(\mathbf{a}, \varphi' \right) \sim_S \left(\mathbf{a}, \varphi \right)$ for some $\mathbf{a}$. Then for $\varphi' > \varphi'' > \varphi''' > \varphi$ we must have $\left(\mathbf{a}, \varphi'' \right) \sim_S \left(\mathbf{a}, \varphi''' \right)$ as shame is weakly increasing in φ. Now pick $\mathbf{a}'$ such that $\left(\mathbf{a}', \varphi \right) \succ_S \left(\mathbf{a}', \varphi' \right)$ and $\left(\mathbf{a}', \varphi \right), \left(\mathbf{a}', \varphi' \right) \in S$. This is possible by continuity of U_S, since for $\mathbf{a}''$ such that $\varphi \left(\mathbf{a}'' \right) = \varphi$ the definition of U_S yields $U_S \left(\mathbf{a}'', \varphi \right) > U_S \left(\mathbf{a}'', \varphi' \right)$. Then by P_5, $\left(\mathbf{a}', \varphi''' \right) \succ_S \left(\mathbf{a}', \varphi'' \right)$, a contradiction to shame being weakly increasing in φ. $\|$

Claim 4.3: For all $\left(\mathbf{a}, \varphi \right)$ and $\widetilde{\varphi} \in \left(\varphi \left(a_1, \mathbf{0} \right), \overline{\varphi} \right)$ there exists $\widetilde{\mathbf{a}}$ such that $\left(\widetilde{\mathbf{a}}, \widetilde{\varphi} \right) \in I_S \left(\mathbf{a}, \varphi \right)$.

Proof: Define φ^* implicitly by $U_s \left(\left(a_1, \mathbf{0} \right), \varphi^* \right) = U_s \left(\mathbf{a}, \varphi \right)$. This is possible by the Intermediate Value Theorem, as $U_s \left(\left(a_1, \mathbf{0} \right), \varphi \left(a_1, \mathbf{0} \right) \right) = u \left(a_1 \right) > U_s \left(\mathbf{a}, \varphi \right) > U_s \left(\left(a_1, 0 \right), \varphi \right)$, where the last inequality is due to P_4 and Claim 4.2. There are two cases to consider:

Case 1) $\widetilde{\varphi} \geq \varphi^*$: Then $U_s \left(\left(a_1, \mathbf{0} \right), \widetilde{\varphi} \right) \leq U_s \left(\mathbf{a}, \varphi \right)$ according to the monotonicity of shame. By F_3^N there is $\overline{a}_2 \left(\widetilde{\varphi} \right)$ that solves $\varphi \left(a_1, \overline{a}_2 \left(\widetilde{\varphi} \right), \mathbf{0} \right) = \widetilde{\varphi}$. Then $U_s \left(\left(a_1, \overline{a}_2 \left(\widetilde{\varphi} \right), \mathbf{0} \right), \widetilde{\varphi} \right) \geq U_s \left(\mathbf{a}, \varphi \right)$ and by the Intermediate Value Theorem there is $\widetilde{a}_2 \left(\widetilde{\varphi} \right) \in \left[0, \overline{a}_2 \left(\widetilde{\varphi} \right) \right)$ such that

$$U_s \left(\left(a_1, \widetilde{a}_2 \left(\widetilde{\varphi} \right), \mathbf{0} \right), \widetilde{\varphi} \right) = U_s \left(\mathbf{a}, \varphi \right).$$

Case 2) $\widetilde{\varphi} < \varphi^*$: Then

$$U_s \left(\left(a_1^* \left(\mathbf{a}, \varphi \right), \mathbf{0} \right), \widetilde{\varphi} \right) \leq U_s \left(\mathbf{a}, \varphi \right) \leq U_s \left(\left(a_1, \mathbf{0} \right), \widetilde{\varphi} \right).$$

By the Intermediate Value Theorem there is $\widetilde{a}_1 \left(\widetilde{\varphi} \right) \in \left[a_1^* \left(\mathbf{a}, \varphi \right), a_1 \right]$ such that

$$U_s \left(\left(\widetilde{a}_1 \left(\widetilde{\varphi} \right), \mathbf{0} \right), \widetilde{\varphi} \right) = U_s \left(\mathbf{a}, \varphi \right). \|$$

Combining the two cases we see that $\widetilde{\varphi}$ parametrizes a path

$$\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right) := \begin{cases} \left(\widetilde{a}_1\left(\widetilde{\varphi}\right),\mathbf{0}\right) \ \text{for} \ \widetilde{\varphi} < \varphi^* \\ \left(a_1,\widetilde{a}_2\left(\widetilde{\varphi}\right),\mathbf{0}\right) \ \text{for} \ \widetilde{\varphi} \geq \varphi^* \end{cases}$$

of allocations. According to Claim 4.2 $\varphi\left(\mathbf{a}\right)$ must be strictly increasing along this path. This implies $\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)$ is strictly increasing in its first component for $\widetilde{\varphi} < \varphi^*$ and in its second component for $\widetilde{\varphi} \geq \varphi^*$.

Now we construct a $\succ_S$ indifference class close to the original one:

Claim 4.4: For $\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)$ as defined above, $\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)$ that solves

$$\left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right),\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)\right) \in I_S\left(\mathbf{a},\varphi + d\varphi\right)$$

is increasing in $\widetilde{\varphi}$.

Proof: Assume $\widetilde{\varphi}' > \widetilde{\varphi}$. There are two cases to consider:

Case 1) $\widetilde{\varphi}' > \varphi^*$: Then $\widetilde{a}_{1(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right) = a_1$, $\widetilde{a}_{1(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right) \leq a_1$ and $\widetilde{a}_{2(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right) > \widetilde{a}_{2(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)$. P_4 implies

$$\left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right),\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)\right) \prec_S \left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right),\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)\right).$$

Case 2) $\widetilde{\varphi}' \leq \varphi^*$: Then $\widetilde{a}_{2(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right) = \widetilde{a}_{2(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right) = 0$ and $\widetilde{a}_{1(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right) > \widetilde{a}_{1(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)$. As $\succ_S$ is increasing in a_1,

$$\left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right),\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)\right) \prec_S \left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right),\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)\right).$$

As shame increases in φ, we must have $\widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}'\right) > \widetilde{\varphi + d\varphi}_{(\mathbf{a},\varphi)}\left(\widetilde{\varphi}\right)$ in both cases.$\|$

Now we define a re-scaling $\varphi \mapsto \gamma\left(\varphi\right)$ in order to transform the original indifference map of $U_S\left(\mathbf{a},\varphi\right)$ to be quasi-linear. We proceed similarly to the proof of Theorem 3. Choose $\varphi^0 \in \left(\underline{\varphi},\overline{\varphi}\right)$ and define $\gamma\left(\varphi^0\right) := 1$. Further set $\gamma\left(\varphi^0 + d\varphi\right) := 1 + d\gamma$, where $d\varphi$ is

infinitesimal. To define $\gamma(\varphi)$ for $\varphi \neq \varphi^0$, pick $\mathbf{a}$ such that $\varphi^*_{(\mathbf{a},\varphi)} < \varphi^0$. As $\varphi^*_{(\mathbf{a},\varphi)} < \varphi$, this implies $\varphi^*_{(\mathbf{a},\varphi)} < \min\left[\varphi, \varphi^0\right]$. Choose $\mathbf{a}^0$ such that $\left(\mathbf{a}^0, \varphi^0\right) \in I_S(\mathbf{a}, \varphi)$. We will look at the increasing graphs $\widetilde{\varphi}$ and $\widetilde{\varphi + d\varphi_{(\mathbf{a},\varphi)}}(\widetilde{\varphi})$ as defined above. Consider two cases for applying the Lock-Step Procedure:

Case 1) $\varphi > \varphi^0$: Define a climbing flight of stairs between the graphs $\widetilde{\varphi}$ and $\widetilde{\varphi} + d\widetilde{\varphi}_{(\mathbf{a},\varphi)}(\widetilde{\varphi})$ recursively: Let φ^{n+1} solve $\left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}(\varphi^n), \varphi^{n+1}\right) \sim_S \left(\mathbf{a}^0, \varphi^0 + d\varphi\right)$. The solution exists by the construction of $\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}(\varphi^n)$.

Case 2) $\varphi < \varphi^0$: Define a descending flight of stairs between the graphs $\widetilde{\varphi}$ and $\widetilde{\varphi} + d\widetilde{\varphi}_{(\mathbf{a},\varphi)}(\widetilde{\varphi})$ recursively: Let φ^{-n-1} solve $\left(\widetilde{\mathbf{a}}_{(\mathbf{a},\varphi)}(\varphi^{-n-1}), \varphi^{-n}\right) \sim_S \left(\mathbf{a}^0, \varphi^0 + d\varphi\right)$.

Then $\gamma(\widetilde{\varphi})$ can be determined analogously to the proof of Theorem 2 by equalizing all step-heights to $d\varphi$ and integrating. Due to P_5 this definition is independent of the choice of $\mathbf{a}^0$.

Now the indifference map of $U_S(\mathbf{a}, \varphi)$ is quasi linear in $\gamma(\varphi)$, where $\gamma: \mathbb{R}_{++} \to \mathbb{R}$ is strictly increasing and continuous. Further remember that $U_S(\mathbf{a}, \varphi)$ is strictly decreasing in φ. Therefore, there exists $H: \mathbb{R}^N_+ \to \mathbb{R}$, such that $H(\mathbf{a}) - \gamma(\varphi)$ represents $\succ_S$ on S.

Define $u_S(a_1) := H(\mathbf{a}) - \lim_{\varphi \to \varphi(\mathbf{a})} \gamma(\varphi)$. Because of P_1,

$$
U(\{\mathbf{a}, \mathbf{b}\}) := \begin{cases} u_S(a_1) \text{ if } \{\mathbf{a}\} \sim \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\} \\ H(\mathbf{a}) - \gamma(\varphi(\mathbf{b})) \text{ if } \{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \succ \{\mathbf{b}\} \\ u_S(b_1) \text{ if } \{\mathbf{a}\} \succ \{\mathbf{a}, \mathbf{b}\} \sim \{\mathbf{b}\} \end{cases}
$$

represents $\succ$ confined to the collection of all two element sets. Therefore, $H(\mathbf{a}) \equiv u_S(a_1) + \gamma(\varphi(\mathbf{a}))$ must hold. Hence

$$
U(A) = \max_{\mathbf{a} \in A}\left[u_s(a_1) + \gamma(\varphi(\mathbf{a}))\right] - \max_{\mathbf{b} \in A}\left[\gamma(\varphi(\mathbf{b}))\right]
$$

represents $\succ$ on K, where φ represents $\succ_f$, and u_s and γ are strictly increasing. Since φ represents $\succ_f$, so does $\gamma(\varphi)$. Hence, there is a representation φ of $\succ_f$, such that γ is the identity and

$$
U(A) = \max_{\mathbf{a} \in A}\left[u_s(a_1) + \varphi(\mathbf{a})\right] - \max_{\mathbf{b} \in A}\left[\varphi(\mathbf{b})\right]
$$

represents $\succ$ on K.

(ii) To establish the analogue of Theorem 3, namely that there are N increasing unbounded functions $v_1, .., v_N$, such that the fairness ranking $\succ_f$ can be represented by $\varphi(\mathbf{a}) = v_1(a_1) \cdot ... \cdot v_N(a_n)$, if and only if it satisfies F_1, F_2, F_3^N and F_4^N we apply the Theorem of Luce and Tukey, just as in the proof of Theorem 3. It establishes the existence of an additive representation $\xi_1(a_1) + ... + \xi_N(a_N)$ of $\succ_f$. Define $v_n(a_n) := \exp(\xi_n(a_n))$ for all $n \in \{1, .., N\}$. Then $v_1, .., v_N : \mathbb{R}_+ \to \mathbb{R}_{++}$ are increasing and continuous and if we re-define $\varphi(\mathbf{a}) := v_1(a_1) \cdot ... \cdot v_N(a_N)$, it represents $\succ_f$. By F_3^N, the functions $v_1, .., v_N$ must be unbounded.

That the representations satisfy the axioms is easy to verify.■

2.8.5 Proof of Proposition

Define the random variable $X_n := \overline{\theta_n} - x_1$.

Claim 5.1: DM's utility is a decreasing function of $|X_n|$

Proof: DM chooses a to maximize

$$u(d(a, x_1)) + \beta_N \left(E\left[\prod_{i=1}^{N} u(d(a, x_i)) | x^n \right] - E\left[\prod_{i=1}^{N} u(d(a^*, x_i)) | x^n \right] \right).$$

Since θ is a sufficient statistic for x^N, we can write:

$$E\left[\prod_{i=1}^{N} u(d(a, x_i)) | x^n \right] = E\left[E\left[\prod_{i=1}^{N} u(d(a, x_i)) | \theta \right] | x^n \right].$$

This expression is single peaked as a function of a:

$$E\left[\prod_{i=1}^{N} u(d(a, x_i)) | \theta \right] = E\left[\prod_{i=1}^{N} u(d(a - \theta, x_i)) | 0 \right] =: f(a - \theta),$$

where f is symmetric and single peaked, with a peak in 0 and .[40] Write $\pi_{x^n}(\theta)$ for the density function corresponding to $\theta \sim N\left(\overline{\theta_n}, \overline{\nu_n^2}\right)$. It is single peaked with peak in $\overline{\theta_n}$.

---

[40] The first equality is justified, since only the distance, which is symmetric, enters the utility functions.

Thus,

$$E\left[\prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right) | x^n\right] = \int_{\Omega} f\left(a - \theta\right) \pi_{x^n}\left(\theta\right) d\theta$$

is the convolution of two symmetric single peaked functions, with peak in 0 and $\overline{\theta_n}$, respectively. Then $E\left[\prod_{i=1}^{N} u\left(d\left(a, x_i\right)\right) | x^n\right]$ is single peaked with peak in $a = \overline{\theta_n}$. This means that fairness is maximized at $a^* = \overline{\theta_n}$ and, therefore, shame is increasing with $\left|\overline{\theta_n} - a\right|$. By assumption, DM's selfish utility is decreasing with $|a - x_1|$. Therefore, DM in effect chooses $\left|\overline{\theta_n} - a\right| \in [0, |X_n|]$. Fix X_n and denote by $l\left(|X_n|\right)$ DM's optimal choice of $\left|\overline{\theta_n} - a\right|$ and $V\left(l\left(|X_n|\right), |X_n|\right)$ the associated (total) utility. Then for $|X_n'| < |X_n|$,

$$\max\left\{0, \left|X_n'\right| - l\left(|X_n|\right)\right\} \leq |X_n| - l\left(|X_n|\right)$$

and

$$\min\left\{\left|X_n'\right|, l\left(|X_n|\right)\right\} \leq l\left(|X_n|\right)$$

with at least one inequality strict. Therefore,

$$V\left(l\left(|X_n|\right), |X_n|\right) < V\left(\min\left\{\left|X_n'\right|, l\left(|X_n|\right)\right\}, \left|X_n'\right|\right)$$

and by definition

$$V\left(\min\left\{\left|X_n'\right|, l\left(|X_n|\right)\right\}, \left|X_n'\right|\right) \leq V\left(l\left(\left|X_n'\right|\right), \left|X_n'\right|\right).$$

Combining the two inequalities establishes the result.$\|$

For given θ, note that

$$X_n | \theta \sim N\left(-\frac{\theta}{\nu^2 + n\sigma^2}, \frac{\sigma^2\left(\sigma^2\left(n - 1\right)\left(2v^2 + n\sigma^2\right) + v^4\right)}{\left(n\sigma^2 + \nu^2\right)^2}\right).$$

Define $H_n(\theta) := -\frac{\nu^2}{\nu^2 + n\sigma^2}\theta$, so $H_n(\theta) \sim N\left(0, \frac{\nu^6}{(\nu^2 + n\sigma^2)^2}\right)$. Then

$$[X_n - H_n(\theta)] \,|\, H_n(\theta) \sim N\left(0, \frac{\sigma^2\left(\sigma^2(n-1)\left(2\nu^2 + n\sigma^2\right) + \nu^4\right)}{(n\sigma^2 + \nu^2)^2}\right).$$

Let $h_n(H_n(\theta))$ and $g_n(X_n - H_n(\theta))$ denote the associated density functions. The convolution

$$\int h_n(H_n(\theta))\, g_n(X_n - H_n(\theta))\, dH_n(\theta)$$

yields

$$X_n \sim N\left(0, \frac{\sigma^2\left(\sigma^2(m-1)\left(2\nu^2 + n\sigma^2\right) + \nu^4\right) + \nu^6}{(n\sigma^2 + \nu^2)^2}\right).$$

Hence for every n, [41]

$$p(|X_n| = z) \propto \exp\left[-\frac{z^2}{2}\frac{\left(\nu^2 + n\sigma^2\right)^2}{\left(\sigma^2\left(\sigma^2(n-1)\left(2\nu^2 + n\sigma^2\right) + \nu^4\right) + \nu^6\right)}\right].$$

Consequently

$$\frac{p(|X_n| = z)}{p(|X_m| = z)}$$

$$\propto \exp\left[-\frac{z^2}{2}\left(\frac{\left(\nu^2 + n\sigma^2\right)^2}{(\sigma^2(\sigma^2(n-1)(2\nu^2 + n\sigma^2)) + \nu^4) + \nu^6)} - \frac{\left(\nu^2 + m\sigma^2\right)^2}{(\sigma^2(\sigma^2(m-1)(2\nu^2 + m\sigma^2)) + \nu^4) + \nu^6)}\right)\right]$$

$$=\ : \exp\left[-\frac{z^2}{2}c_{\nu,\sigma}(m,n)\right].$$

Thus, the ratio $\frac{p(|X_n|=z)}{p(|X_m|=z)}$ satisfies the Monotone Likelihood Ratio Property (MLRP): If $c_{\nu,\sigma}(m,n) < (>)\,0$, then $\frac{p(|X_n|=z)}{p(|X_m|=z)}$ increases (decreases) with z. Since DM's utility is decreasing in z, she will strictly prefer m over n (n over m) if and only if $c_{\nu,\sigma}(m,n) < (>)\,0$. In the text we define the ratio of the standard deviations σ and ν as $s := \frac{\sigma}{\nu}$. s is a sufficient statistic for (ν,σ). Assuming $n > m$, we find, with some straightforward algebra, that $c_s(m,n) < (>)\,0$, if and only if $2 + (m+n)\,s^2 - 3s^4 - 2(m+n)\,s^6 - mns^8 < (>)\,0$.

---

[41] We use the fact that for a Normal distribution with mean 0, $p(X_n = z) = p(X_n = -z)$. With $\propto$ we denote "proportional to".

Claim 5.2: For $n > m$, $c_s(m, m+1) < 0$ implies $c_s(n, n+1) < 0$.

Proof:

$$2 + (m + (m+1)) s^2 - 3s^4 - 2(m + (m+1)) s^6 - m(m+1) s^8 < 0$$

is equivalent to

$$\frac{2 - 3s^4}{2m + 1} < 2s^6 + \frac{m^2 + m}{2m + 1} s^8 - s^2.$$

Let $lhs := \frac{2 - 3s^4}{2m+1}$ and $rhs := 2s^6 + \frac{m^2 + m}{2m+1} s^8 - s^2$. Consider two cases:

i) $2 - 3s^4 > 0$: $\frac{\partial}{\partial m}(lhs) < 0$ and $\frac{\partial}{\partial m}(rhs) > 0$ implies $c_s(n, n+1) < 0$ for $n > m$.

ii) $2 - 3s^4 \leq 0$: Then for all m, $lhs \leq 0$ and $rhs > 0$, which implies $c_s(n, n+1) < 0$ for

$n > m$. $\|$

s^* as defined in the text solves $c_s(1, 2) = 0$.

Claim 5.3: For $s \leq s^*$, $c_s(m, m+1) = 0$ has a unique solution $n^*(s) \in \mathbb{R}_+$. For $s > s^*$,

no positive solution exists.

Proof: Assume first $s \leq s^*$: To show existence, note that due to the quadratic term

in m, $m \to \infty$ implies $c_s(m, m+1) \to -\infty$ for all s. Since $c_s(m, m+1)$ is continuous in

m, it is sufficient to show that $c_s(1, 2) > 0$ for all $s \leq s^*$. $c_s(1, 2) > 0$ is equivalent to

$$2 + 3s^2 - 3s^4 - 6s^6 - 2s^8 > 0$$

or

$$2/s^2 + 3 > 3s^2 + 6s^4 + 2s^6 \ \forall s \leq s^*.$$

For this last inequality $\frac{\partial}{\partial s}(lhs) < 0$ and $\frac{\partial}{\partial s}(rhs) > 0$. Since $2/s^{*2} + 3s^* = 3s^{*2} + 6s^{*4} + 2s^{*6}$,

$c_s(1, 2) > 0$ must hold for all $s \leq s^*$. Hence a solution $m = n^*(s)$ exists. Its uniqueness

follows directly from claim 5.1.

106

Consider now the case $s > s^*$: In that case $c_s(1,2) < 0$. Claim 1 implies that no solution exists to the equation $c_s(m, m+1) = 0$ for any $m \in \mathbb{R}_+$.$\|$

Claim 5.4: $n^*(s)$ is decreasing in s.

Proof:

$$2 + (n^* + (n^* + 1)) s^2 - 3s^4 - 2(n^* + (n^* + 1)) s^6 - n^*(n^* + 1) s^8 = 0$$

is equivalent to

$$\frac{2/s^2 - 3s^2}{2n^* + 1} = 2s^4 + \frac{n^{*2} + n^*}{2n^* + 1} s^6 - 1.$$

For this equality $\frac{\partial}{\partial n^*}(lhs) < 0$ and $\frac{\partial}{\partial n^*}(rhs) > 0$ for all s, while $\frac{\partial}{\partial s}(lhs) < 0$ and $\frac{\partial}{\partial s}(rhs) > 0$ for all n^*.$\|$

Thus, if DM prefers $n = 1$ to $n = 2$, then $n^*(s) = 1$ is her globally preferred value. If she prefers m to both $m - 1$ and $m + 1$, then $n^*(s) = m$ is her globally preferred value. Neglecting the integer constraint, $m = n^*(s)$ is the unique positive solution to $2 + (m + (m+1)) s^2 - 3s^4 - 2(m + (m+1)) s^6 - m(m+1) s^8 = 0$, if a solution exists. Furthermore, $n^*(s)$ is a decreasing function.∎

Bibliography

[1] Anderoni, James and John H. Miller (2002) "Giving according to GARP: An experimental test of the consistency of preference for altruism." *Econometrica*, 70, 737-753.

[2] Benabou, Roland and Jean Tirole (2006) "Incentives and Prosocial Behavior." *American Economic Review*, 96(5), 1652-1678.

[3] Bolton, Gary E., Elena Katok and Rami Zwick (1998) "Dictator game giving: Rules of fairness versus acts of kindness." *International Journal of Game Theory,* 27: 269-299.

[4] Broberg, Thomas, Tore Ellingsen and Magnus Johannesson (2007) "Is Generosity Involuntary?" *Economics Letters,* 94, 32-37.

[5] Burnham, Terence C. (2003) "Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving." *Journal of Economic Behavior and Organization,* 50, 133-144.

[6] Buss, Arnold H. (1980) "Self-Consciousness and Social Anxiety." San Francisco, W. H. Freeman.

[7] Camerer, Colin (2003) "Behavioral Game Theory: Experiments in Strategic Interaction". Princeton University Press.

[8] Charnes, Gary and Mathew Rabin (2002) "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117 (3), 817-870.

[9] Coate, Stephen and Stephen Morris (1995) "On the Form of Transfers to Special Interests." *Journal of Political Economy,* Vol. 103, No. 6, 1210-1235.

[10] Dana, Jason D., Dalian M. Cain and Robin M. Dawes (2006) "What you don't Know Won't Hurt me: Costly (but quiet) Exit in a Dictator Game." *Organizational Behavior and Human Decision Processes*, 100(2), 193-201.

[11] Dana, Jason D., Roberto A. Weber and Jason Xi Kuang (2005) "Exploiting moral wriggle room: Behavior inconsistent with a preference for fair outcomes." mimeo.

[12] Davis, Douglas D. and Charles A. Holt (1993) " Experimental economics." Princeton University Press.

[13] Dekel, Eddie, Barton L. Lipman and Aldo Rustichini (2005) "Temptation Driven Preferences." mimeo

[14] Fehr Ernst, Klaus M. Schimdt (1999) "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114, 817-868.

[15] Frohlich, Norman, Joe Oppenheimer and J. Bernard Moore (2001) "Some doubts measuring self-interst using dictator experiments: The cost of anonymity." *Journal of Economic Behavior and Organization,* 46, 271-290.

[16] Gauthier, David and Robert Sugden (editors) (June 1993) "Rationality, Justice and Social Contract: Themes from Morals by Agreement." University of Michigan Press.

[17] Gul, Faruk and Wolfgang Pesendorfer (2005) "The Simple Theory of Temptation and Self-Control." mimeo

[18] —— (2001) "Temptation and Self Control." *Econometrica*, Vol. 69, No. 6, 1403-1435

[19] Haley, Kevin J. and Daniel M.T. Fessler (2005) "Nobody's watching? Subtle cues affect generosity in an anonymous economic game." *Evolution and Human Behavior,* 26, 245-256.

[20] Hammond, Peter J.(1991), "Interpersonal Comparisons of Utility: Why and How They Are and Should Be Made", in Elster and Roemer, "Interpersonal Comparisons of Well Being". Cambridge: Cambridge University Press, pp 200-254.

[21] Johannesson, Magnus and Bjoran Persson (2000) "Non-reciprocal altruism in dictator games." *Economics Letters,* 69, 137-142.

[22] Karni, Edi and Zvi Safra (1998) "The Hexagon Condition and Additive Representation for Two Dimensions: An Algebraic Approach." *Journal of Mathematical Psychology,* 42, 393-399.

[23] Keeney, Ralph L. and Howard Raiffa, with a contribution by Richard F. Meyer (1976) "Decisions with multiple objectives : preferences and value trade-offs". New York : Wiley.

[24] Koch, K. Alexander and Hans-Theo Norman (2005) "Giving in Dictator Games: Regard *for* Others or Regard *by* others?" mimeo.

[25] Krantz, David H., R. Duncan Luce, Patrick C. Suppes and Amos Tversky (1971) " Foundations of Measurments, Vol 1." Academic Press, New York.

[26] Lazear, Edward P., Ulrike Malmendier and Roberto A. Weber (2005) "Sorting in Experiments with Application to Social Preferences." mimeo.

[27] Luce, R. Duncan and John W. Tukey (1964) "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement." *Journal of Mathematical Psychology,* 1,1-27.

[28] Mariotti, Marco (1999) "Fair Bargains: Distributive Justice and Nash Bargaining Theory." *Review of Economic Studies*, Vol.66, 733-41.

[29] Miller, Dale T. (1999) "The Norm of Self-Interest." *American Psychologist*, Vol. 54, No. 12, 1053-1060.

[30] Nash, John F. (1953) "Two-Person Cooperative Games." *Econometrica,* Vol. 21, No. 1, 128-140.

[31] —— (1950) "The Bargaining Problem." *Econometrica*, Vol. 18, No. 2, 155-162.

[32] Neilson, William S. (2006-a) "Axiomatic Reference Dependence in Behavior towards Others and Toward Risk." *Economic Theory* 28, 681-692.

[33] —— (2006-b) "A Theory of Kindness, Reluctance, and Shame in Dictator Games." mimeo.

[34] Oberholzer-Gee, Felix and Reiner Eichenberger (2004) "Fairness in Extended Dictator Game Experiments". Working paper

[35] Osborne, Martin J. and Ariel Rubinstein (1994) "A course in Game Theory." MIT press, ISBN 0-262-65040-1.

[36] Pillutla, Madan M. and J. Keith Murningham (1995) "Being fair or appearing fair: Strategic behavior in ultimatum bargaining." *Academy of Managment Journal,* 38,1408-1426.

[37] Rawls, John (1971) "A Theory of Justice." *The Belknap Press of Harvard University Press.*

[38] Tullock, Gordon (1983) "Economics of Income Redistribution." Boston: Kluwer-Nijhoff.

[39] Wittman, Donald (1989) "Why Democracies Produce Efficient Results." *The Journal of Political Economy*, Vol. 97, No. 6, 1395-1424.